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.

Online-Offline Integrated Multimodal Teaching Combining PBL, CBL, and CTTM to Enhance Clinical Self-Efficacy in Endocrine and Metabolic Critical Illness Among Internal Medicine Residents: A Convergent Parallel Mixed- Methods Randomized Controlled Trial

Jianyong Zhao
The Fourth Affiliated Hospital of Anhui Medical University
Li Zhu
The Fourth Affiliated Hospital of Anhui Medical University
Na Li
The Fourth Affiliated Hospital of Anhui Medical University
Shiqi Li
The Fourth Affiliated Hospital of Anhui Medical University
Meiyan Chi
The Fourth Affiliated Hospital of Anhui Medical University
Jiaqing Dou
djqch1@163.com

The Fourth Affiliated Hospital of Anhui Medical University

Research Article

Keywords: online-offline blended teaching, PBL, CBL, CTTM, endocrine and metabolic critical illness, clinical self-efficacy, internal medicine residents, convergent parallel mixed methods

Posted Date: September 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10815117/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 Endocrine and metabolic emergencies frequently trigger clinical anxiety and reduce clinical self-efficacy among internal medicine residents. Traditional lecture-based learning (LBL) relies on one-way knowledge transmission and lacks hands-on training in critical care scenarios. According to Bandura's self-efficacy theory, insufficient clinical self-efficacy directly leads to diagnostic hesitation and decision avoidance. Although problem-based learning (PBL), case-based learning (CBL), and case three-dimensional teaching method (CTTM) each offer distinct advantages, no study has yet integrated all three into a unified online-offline teaching system or employed mixed-methods approaches to elucidate the underlying mechanisms across Bandura's four dimensions of self-efficacy. The present study addresses this gap.

Methods This quantitatively driven, single-center, assessor-blinded, convergent parallel mixed-methods randomized controlled trial was conducted at the Fourth Affiliated Hospital of Anhui Medical University from September 2025 to August 2026 (Ethics Approval No.: KYXM202508021). Ninety internal medicine residents on their first endocrine rotation were enrolled and randomly assigned in a 1:1:1 ratio to a control group (LBL alone), experimental group 1 (PBL+LBL), or experimental group 2 (integrated PBL-CBL-CTTM closed-loop teaching). All three groups received a standardized 12-hour curriculum on endocrine and metabolic critical illness. Outcome measures included blinded theoretical examinations, clinical skills assessments, the General Self-Efficacy Scale (GSES), a self-developed Endocrine High-Risk Management Self-Efficacy Source Questionnaire (Cronbach's α=0.81, CVI=0.92), and a satisfaction scale. Statistical methods included one-way ANOVA with Bonferroni-corrected post-hoc pairwise comparisons (adjusted α'=0.017), chi-square tests, and receiver operating characteristic (ROC) curve analysis. For the qualitative component, 24 residents were selected via stratified purposive sampling to participate in eight 30–40 minute semi-structured focus group interviews, analyzed using the Braun & Clarke six-step thematic analysis method with NVivo 12. Quantitative and qualitative data were integrated at the interpretation stage using joint display tables to identify convergence and divergence.

Results All baseline characteristics were comparable across the three groups (P>0.05). Post-intervention, experimental group 2 achieved significantly higher theoretical scores (45.10±3.12), clinical skills scores (46.20±2.41), and composite total scores (91.30±4.20) compared with the other two groups (P<0.017), with 97% of experimental group 2 participants endorsing improved critical care management ability and superior scores across all four self-efficacy dimensions. Complete convergence was observed for mastery experience, vicarious learning, and positive affect; only the verbal persuasion dimension showed partial divergence—single bedside immediate guidance had limitations, while multi-stage systematic feedback effectively compensated for this shortcoming.

Conclusions The integrated PBL-CBL-CTTM teaching approach simultaneously enhances theoretical knowledge, hands-on proficiency in critical cases, and clinical self-efficacy among residents, while alleviating fear of endocrine and metabolic emergency management. The entire teaching program can be implemented at different tiers using routine hospital resources. Future studies should disentangle the

independent effects of each module and conduct multi-center longitudinal follow-up to verify long-term training outcomes.

Trial registration: ClinicalTrials.gov NCT07801170; date of registration: August 28, 2026; Retrospectively registered.

Introduction

Diabetic ketoacidosis (DKA), hyperosmolar hyperglycemic syndrome (HHS), thyroid storm, and refractory hypoglycemia are all classified as endocrine and metabolic emergencies[1–3]. Their clinical management requires rapid assessment of internal environment disturbances, dynamic adjustment of glucose-lowering and antithyroid medications, and continuous monitoring of vital signs[4, 5]. According to Bandura's self-efficacy theory, insufficient clinical self-efficacy directly leads to diagnostic hesitation, decision avoidance, and reduced willingness to perform hands-on procedures, substantially increasing the risk of medication adjustment errors in critically ill patients and seriously threatening patient safety[3, 6]. Currently, endocrine residency training in most tertiary hospitals in China remains dominated by traditional lecture-based learning (LBL), characterized by one-way didactic instruction with inadequate supervised hands-on training in critical cases and limited pathophysiological visualization aids[7, 8]. Hospital-based clinical surveys indicate that over 60% of residents on their first endocrine rotation experience significant anxiety when independently managing critical metabolic diseases[9–11]. Furthermore, complex disease pathways such as metabolic-associated fatty liver disease (MAFLD) combined with type 2 diabetes are abstract in nature, making it difficult for text-only lectures to facilitate the translation of theoretical knowledge into clinical decision-making competence.

Problem-based learning (PBL) centers on critical clinical problems[12], guiding residents to independently search the literature and participate in online group discussions before class, thereby fostering self-directed learning and logical reasoning skills[13]. However, PBL lacks authentic ward-based hands-on scenarios. Case-based learning (CBL) leverages hospitalized patients for bedside case discussions, bridging the gap between theory and clinical practice, yet lacks visualization tools for abstract hormonal and glucolipid metabolic pathways[12, 14, 15]. The case three-dimensional teaching method (CTTM) employs 3D reconstructed imaging, pathological animations, and standardized emergency procedure videos to visualize complex metabolic pathways, but lacks pre-class preparation and sustained hands-on closed-loop training[16, 17]. Previous single-center teaching studies in China suggest that any single modality alone can only address one specific teaching deficiency and cannot simultaneously improve theoretical knowledge, hands-on skills, and clinical confidence[18, 19]. Online-offline multimodal integration represents the mainstream direction of medical education reform in China[20, 21].

A review of Chinese and English literature on endocrine residency training over the past five years reveals three progressive gaps in existing research. First, no study has established an integrated PBL-CBL-CTTM endocrine teaching system covering the full workflow of pre-class online preparation, in-person lectures,

and ward-based hands-on training. Second, few studies have employed convergent parallel mixed-methods designs to comprehensively dissect the intrinsic psychological pathways through which teaching interventions affect residents across Bandura's four self-efficacy dimensions. Third, there is a lack of targeted teaching interventions for endocrine-specific critical care anxiety, with no qualitative data elucidating the mechanisms by which 'fear of critical metabolic emergencies' is alleviated.

Building on these research gaps, the present study developed a three-stage closed-loop multimodal teaching program and conducted triangulated validation using objective assessments, standardized scales, and stratified focus group interviews, aiming to provide a standardized, low-cost, and implementable teaching reform plan for endocrine residency training at hospitals of all tiers. The study employed a quantitatively driven convergent parallel mixed-methods design, in which quantitative and qualitative data were collected in parallel and integrated at the interpretation stage using joint display tables[22].

This study proposed three objectives: (1) to quantitatively compare inter-group differences in theoretical knowledge, critical care procedural skills, and learning satisfaction across three teaching modalities; (2) to quantitatively analyze differences in four self-efficacy dimensions when managing critical illnesses; and (3) to qualitatively extract core teaching themes through interviews and elucidate the intrinsic pathways by which multimodal teaching alleviates clinical fear and fosters self-efficacy, with integrated joint-display analysis identifying convergence and divergence across the four self-efficacy dimensions.

Methods

Study Design

This study was a quantitatively driven, single-center, prospective, assessor-blinded, convergent parallel mixed-methods randomized controlled trial[23, 24]. Quantitative indicators objectively evaluated teaching effectiveness, while qualitative interviews explained the intrinsic psychological mechanisms underlying changes in self-efficacy. Quantitative and qualitative data were independently collected and analyzed in parallel, with integrated analysis performed at the interpretation stage to identify convergence and divergence using joint display tables, following the convergent parallel mixed-methods design framework and joint display integration guidelines for health professions education[22]. The complete participant screening, randomization, and mixed-methods flowchart is presented in Fig. 1. The study period spanned September 2025 to August 2026, conducted at the Department of Endocrinology and Metabolism, Fourth Affiliated Hospital of Anhui Medical University. All procedures followed the Declaration of Helsinki, and participants provided written informed consent. This randomized controlled trial is reported in accordance with the CONSORT 2010 statement (http://www.consort-statement.org/), and the completed CONSORT checklist is provided as an additional file.

Participants

Inclusion criteria: Full-time internal medicine residents on their first endocrine rotation; no prior endocrine remedial training or advanced study experience; no independent management experience with critical cases; ability to complete all courses with balanced baseline clerkship duration; voluntary completion of assessments, scales, and recorded interviews. Exclusion criteria: Mid-study withdrawal or cumulative absences exceeding 2 class hours; inability to complete the full evaluation battery; refusal of interview recording; long-term endocrine specialty training or independent management of critically ill patients. Of the 112 residents assessed for eligibility, 22 were excluded before randomization for the following reasons: not first endocrine rotation (n = 8), prior endocrine remedial training or advanced study experience (n = 4), independent critical case management experience (n = 3), inability to complete all courses due to unbalanced clerkship duration (n = 5), and declined participation or failure to provide written informed consent (n = 2).

Sample size was calculated based on the primary outcome of clinical skills scores using a two-independent-samples t-test formula (conservative estimate for the three-group design), with $\alpha = 0.05$, power = 0.8, an expected inter-group skills difference of 4 points, and a standard deviation of 3.3 (derived from a pilot study conducted at our institution), yielding a minimum of 27 participants per group. Accounting for an estimated 10% attrition rate, 30 participants were enrolled per group, for a total sample of 90. Simple randomization in a 1:1:1 ratio was performed using a computer-generated random number sequence (SPSS 26.0) by an independent statistician not involved in participant enrollment, teaching, or assessment. Allocation concealment was achieved using sequentially numbered, sealed, opaque envelopes; the enrolling researcher opened each envelope only after confirming participant eligibility and written informed consent. Teaching content was uniformly based on the endocrine critical illness chapters of the 9th edition of Internal Medicine (People's Medical Publishing House). After all evaluations were completed, the full set of experimental group 2 teaching resources was made available to the control and experimental group 1 to ensure ethical equity.

Three Standardized Teaching Protocols

Control group (LBL alone): Received only in-person one-way didactic lectures with self-directed review after class; no online discussions, ward-based training, CTTM resources, or post-class debriefing.

Experimental group 1 (PBL + LBL): One week before class, a study group was established via DingTalk, where open-ended questions on critical cases were distributed. Residents independently searched the literature and participated in online discussions[12, 14], with faculty providing scheduled Q&A. In-person lecture content and duration were identical to the control group; no CBL bedside hands-on training or CTTM visualization materials were provided.

Experimental group 2 (integrated PBL-CBL-CTTM closed-loop teaching): Pre-class PBL procedures were identical to experimental group 1; in-person lecture content was standardized. After class, residents worked in groups of 3–4 in the ward, with attending physicians providing continuous supervision while completing history-taking, physical examination, and individualized treatment plan discussions. Case-

related thought questions were distributed via DingTalk, accompanied by full-cycle CTTM resources including 3D reconstructions of pancreatic islets and liver[16, 25], pathological animations, and emergency micro-lectures[17]. A weekly collective debriefing session for critical cases was additionally established[26]. A comparison of the three teaching configurations is presented in Table 1.

Table 1
Comparison of Complete Teaching Intervention Components Across Three Groups

Teaching Component	Control (LBL alone)	Exp. Group 1 (PBL + LBL)	Exp. Group 2 (Full PBL-CBL-CTTM)
Pre-class online PBL autonomous discussion	No	Yes	Yes
In-person standardized LBL classroom lecture	Yes	Yes	Yes
Physician-supervised CBL bedside training	No	No	Yes
Full-cycle CTTM visualization resources	No	No	Yes
Routine DingTalk online Q&A	No	Yes	Yes
Post-class critical case debriefing	No	No	Yes

Treatment fidelity: All teaching faculty completed a standardized 4-hour training session on the three teaching protocols before study initiation, including role-play demonstrations and checklist-based calibration. A structured 15-item fidelity checklist covering core teaching components (PBL discussion quality, CBL bedside supervision, CTTM resource utilization, lecture standardization, and debriefing completeness) was used by two senior educators to independently review 20% of randomly selected audio-recorded teaching sessions. Inter-rater agreement was 92.3%, and overall fidelity adherence exceeded 90% across all three groups, confirming consistent intervention delivery.

Outcome Measures

The primary outcome was the clinical skills assessment score (50 points). Secondary outcomes included theoretical examination scores, General Self-Efficacy Scale (GSES) scores, the four self-efficacy dimension scores from the self-developed questionnaire, and satisfaction scores. All assessments and scales were administered within 3 working days after course completion by an independent evaluation team with no teaching responsibilities, who were blinded to group allocation (assessor-blinded). To maintain assessor blinding, participants were explicitly instructed not to disclose their group assignment or teaching modality during any assessment; assessors were separated from teaching materials and a research assistant monitored each session for potential unblinding. No instances of unblinding were recorded. Objective assessment (total 100 points): Theoretical examination (50 points), including single-

choice, multiple-choice, and short-answer questions with unified blinded grading; clinical skills assessment (50 points), evaluating critical case analysis, insulin tiered regimens, and fatty liver risk stratification, with standardized scoring rubrics for examiners.

Standardized scales (completed anonymously throughout to reduce social desirability bias): The GSES comprises 10 items (Cronbach's α = 0.89). The self-developed questionnaire was based on Bandura's four-dimensional framework, initially containing 22 items. After two rounds of Delphi screening by 5 endocrine experts (items with mean importance < 3.5 or coefficient of variation > 0.25 were eliminated; expert authority coefficient Cr = 0.88), and a pilot study of 30 participants, 6 items were retained (Cronbach's α = 0.81, CVI = 0.92, EFA variance explained = 78.6%, CFA fit indices: $\chi^2/df = 1.73$, RMSEA = 0.046, CFI = 0.952, TLI = 0.947; complete items in Appendix Table 1). The satisfaction questionnaire included 5 dimensions: willingness for long-term adoption, enhanced learning interest, improved learning efficiency, deepened pathophysiological understanding, and improved critical care management ability, with both recognition rates and mean scores calculated simultaneously (Cronbach's α = 0.804).

Focus group interviews: Stratified purposive sampling was employed, with 6 participants from the control group, 6 from experimental group 1, and 12 from experimental group 2, totaling 8 sessions of 30–40 minutes each (2 sessions for the control group, 2 for experimental group 1, and 4 for experimental group 2). Interviews were conducted sequentially by group, and qualitative saturation was confirmed by no new themes emerging across the final 3 consecutive sessions (the final 2 sessions of experimental group 2 and the final session of experimental group 1). Due to time and funding constraints, random sampling was not employed; this will be optimized in future multi-center studies. Interviews were conducted in a quiet independent office, with a moderator having no teaching or evaluation authority. Recordings were transcribed verbatim within 24 hours. Data analysis followed the Braun & Clarke six-step thematic analysis method, using NVivo 12 software for data management[23, 27]. Trustworthiness was ensured through dual independent coding (two researchers coded independently, with discrepancies resolved through group discussion), member checking (participants reviewed transcribed themes for accuracy), and data triangulation (cross-referencing quantitative scale results with qualitative interview findings). The interview guide is provided in Appendix Table 2. Core questions addressed critical care hands-on opportunities, observation channels, teaching feedback, and changes in clinical emotions.

Table 2
Baseline Characteristics Across Three Groups (n = 30 per group, $\bar{x}\pm s$)

Indicator	Control	Exp. Group 1	Exp. Group 2	F/χ^2	*P*
Age (years)	26.10 ± 1.80	25.80 ± 1.60	26.30 ± 1.50	1.23	> 0.05
Male/Female (n)	11/19	10/20	12/18	0.76	> 0.05
Prior final exam score	74.20 ± 9.40	73.50 ± 10.10	74.80 ± 9.70	0.89	> 0.05
Baseline GSES	26.80 ± 3.10	27.10 ± 2.90	26.50 ± 3.30	1.04	> 0.05
Endocrine clerkship duration (h)	12.60 ± 3.50	13.10 ± 3.20	12.80 ± 3.40	0.97	> 0.05

Data Analysis

Quantitative: Analyzed using SPSS 26.0. Measurement data are presented as mean ± standard deviation ($\bar{x}\pm s$) with two decimal places. One-way ANOVA was employed for inter-group comparisons of continuous outcomes, with partial eta-squared (η^2p) reported as a measure of effect size (theoretical scores, clinical skills scores, self-efficacy dimension scores, and satisfaction dimension mean scores), with Bonferroni-corrected post-hoc pairwise comparisons (adjusted $\alpha'=0.017$). Categorical data (satisfaction recognition rates) were analyzed using chi-square tests, with a two-sided overall $P < 0.05$ considered statistically significant. Receiver operating characteristic (ROC) curves were plotted to assess the predictive performance of each teaching mode for excellent clinical-skill achievement (defined as clinical skills score ≥ 45/50), with area under the curve (AUC) compared across groups using the DeLong test with Bonferroni correction for multiple pairwise comparisons (adjusted $\alpha'=0.017$). Qualitative: Braun & Clarke six-step thematic analysis method, with dual independent coding and consensus resolution for discrepancies through group discussion. Mixed-methods integration: Following established joint display integration methodology, convergence and divergence were operationalized for this study as follows: Complete convergence = quantitative trends highly matching interview themes; Partial divergence = scales capturing only instantaneous indicators while interviews reflecting long-term systemic effects. Integrated findings are presented in Table 6.

Table 6
Joint Display Integration Analysis of Quantitative and Qualitative Findings Across Four Self-Efficacy Dimensions (Quantitative percentages derived from item-level analysis of the self-developed self-efficacy source questionnaire; see Appendix Table 5.)

Efficacy Dimension	Quantitative Inter-Group Characteristics	Core Content of Stratified Interviews	Convergence/Mechanism Interpretation
Mastery experience	72% of Exp. Group 2 could independently complete supervised full critical care management	Control/Exp. Group 1 had virtually no autonomous hands-on experience; Exp. Group 2 received routine comprehensive CBL training	Complete convergence; supervised hands-on experience is Bandura's strongest source of self-efficacy, continuously accumulating positive clinical experiences
Vicarious learning	76% of Exp. Group 2 could fully observe standardized critical care management procedures	Control/Exp. Group 1 lacked observation channels; Exp. Group 2 received stratified multidisciplinary demonstrations	Complete convergence; senior faculty demonstrations establish indirect clinical reasoning learning pathways
Verbal persuasion	Only 53% of Exp. Group 1 endorsed bedside immediate feedback, compared with 90% in Exp. Group 2	Exp. Group 1 had only pre-class online Q&A; Exp. Group 2 received four-stage continuous feedback	Partial divergence; single bedside feedback is limited by workload constraints, while full-cycle systematic feedback is effective long-term
Positive affect	81% of Exp. Group 2 could calmly manage critical emergencies	Control/Exp. Group 1 experienced severe clinical anxiety; Exp. Group 2 used simulation + graduated hands-on practice to alleviate fear	Complete convergence; visualization pre-training buffers negative emotions in critical scenarios, preventing decision paralysis

Results

All 90 participants completed every evaluation with no attrition. No significant differences were observed across the three groups in age, sex, prior academic performance, baseline scale scores, or clerkship duration (all $P > 0.05$), confirming baseline comparability (Table 2).

Significant overall differences were found across the three groups in theoretical scores, skills scores, and total scores (all $P < 0.05$). Post-hoc pairwise comparisons with Bonferroni correction revealed that both experimental groups 1 and 2 achieved higher theoretical scores than the control group, with no significant difference between the two experimental groups. Only experimental group 2 demonstrated significantly superior clinical skills compared with the other two groups. For composite total scores, the ranking was experimental group 2 > experimental group 1 > control, with all pairwise comparisons $P < 0.017$ (Table 3).

Table 3
Comparison of Theoretical, Clinical Skills, and Composite Total Scores Across Three Groups

Group	n	Theory (max 50)	Clinical Skills (max 50)	Composite Total (max 100)
Control	30	39.70 ± 2.51	35.90 ± 3.30	75.60 ± 4.42
Exp. Group 1	30	43.60 ± 2.31▲	36.90 ± 3.42	80.50 ± 4.31▲
Exp. Group 2	30	45.10 ± 3.12▲	46.20 ± 2.41▲■	91.30 ± 4.20▲■
F value	-	8.15	9.72	8.93
Overall P	-	< 0.05	< 0.05	< 0.05

Note: Pairwise inter-group comparisons employed Bonferroni-corrected q-tests with adjusted α'=0.017; ▲ indicates P < 0.017 compared with the control group, ■ indicates P < 0.017 compared with experimental group 1. Partial eta-squared (η^2p) effect sizes, calculated as F×df_between/(F×df_between+df_within) with df_between = 2 and df_within = 87, were 0.158 for theoretical scores, 0.183 for clinical skills scores, and 0.170 for composite total scores, all indicating large effect sizes ($\eta^2p > 0.14$).

All five satisfaction dimensions followed the same ranking: experimental group 2 > experimental group 1 > control. For recognition rates, all chi-square tests were statistically significant (P < 0.05); for mean scores, one-way ANOVA with Bonferroni correction showed experimental group 2 significantly higher than both other groups across all five dimensions (P < 0.017). The recognition rate for improved critical care management ability reached 97% in experimental group 2 (Table 4).

Table 4
Satisfaction Recognition Rates and Dimension Mean Scores (%/$\bar{x}$±SD)

Evaluation Dimension	Control	Exp. Group 1	Exp. Group 2	χ^2	*P*
Willingness for long-term adoption	46/2.11 ± 0.43	82/3.25 ± 0.47	93/4.52 ± 0.35	11.24	< 0.05
Enhanced learning interest	43/2.08 ± 0.41	79/3.18 ± 0.45	90/4.46 ± 0.33	13.15	< 0.05
Improved learning efficiency	41/2.02 ± 0.39	81/3.22 ± 0.46	93/4.50 ± 0.34	16.02	< 0.05
Deepened pathophysiological understanding	72/2.86 ± 0.52	86/3.61 ± 0.48	94/4.58 ± 0.30	7.41	< 0.05
Improved critical care management ability	66/2.63 ± 0.48	82/3.34 ± 0.44	97/4.65 ± 0.32	10.13	< 0.05

Note: Data are presented as recognition rate (%)/mean ± SD. Recognition rates were compared using chi-square tests; mean scores were compared using one-way ANOVA with Bonferroni correction. All P < 0.05.

Experimental group 2 achieved the highest scores across all four self-efficacy dimensions (Table 5). On the self-developed self-efficacy source questionnaire, item-level analysis revealed that only 53% of experimental group 1 participants considered bedside feedback timely, compared with 90% in experimental group 2 (χ^2=9.32, P < 0.01); complete pairwise comparison data are presented in Appendix Table 5. Item-level analysis of the self-developed questionnaire further revealed that 72% of experimental group 2 participants could independently complete supervised full critical care management, 76% could fully observe standardized critical care management procedures, and 81% could calmly manage critical emergencies (complete item-level data in Appendix Table 5).

Table 5
Inter-Group Comparison of Four Self-Efficacy Dimensions

Dimension	**Control**	**Exp. Group 1**	**Exp. Group 2**	***P***
Mastery experience	4.12 ± 0.51	4.35 ± 0.48	5.76 ± 0.43	< 0.017
Vicarious learning	4.08 ± 0.49	4.41 ± 0.52	5.68 ± 0.40	< 0.017
Verbal persuasion	4.20 ± 0.47	4.32 ± 0.50	5.53 ± 0.45	< 0.017
Positive affect	4.01 ± 0.53	4.28 ± 0.46	5.82 ± 0.39	< 0.017

Note: Data are presented as mean ± SD. P values shown in the table reflect Bonferroni-corrected post-hoc pairwise comparisons (adjusted α' = 0.017), indicating that experimental group 2 was significantly higher than both the control and experimental group 1 for all four dimensions. Overall one-way ANOVA across the three groups was statistically significant for all four dimensions (all P < 0.001).

Five key themes were extracted from the interviews, organized into negative experiential themes (predominant in the control and experimental group 1) and positive experiential themes (predominant in experimental group 2). Two negative themes were identified: (1) Fear of critical metabolic emergencies—attributable to the absence of routine supervised hands-on training and abstract disease pathways, leading to avoidance behavior; (2) Inadequate vicarious learning channels—residents reported few opportunities to observe standardized critical care management by senior faculty. In contrast, experimental group 2 interviews yielded three positive themes: (1) Safety-tiered mentorship—including risk-free hands-on environments and multidisciplinary demonstrations, providing vicarious learning channels; (2) Full-cycle practice—pre-class preparation, graduated hands-on training, and weekly debriefing cumulatively building positive clinical confidence; (3) Multi-stage systematic feedback—continuous correction across online Q&A, bedside guidance, and collective debriefing, addressing the limitation of single-time-point bedside feedback. These five themes were mapped onto Bandura's four self-efficacy dimensions for joint display integration.

Integrated quantitative-qualitative analysis: Complete convergence was observed for three dimensions—mastery experience, vicarious learning, and positive affect. Divergence was found for verbal persuasion: 83% of experimental group 1 interviewees (5 of 6) reported only pre-class online Q&A without bedside guidance, consistent with the 53% timely feedback rate on the quantitative scale. Experimental group 2

established a four-tier systematic feedback mechanism (Table 6). As shown in Fig. 2, receiver operating characteristic curves were plotted to compare the predictive performance of three teaching modes for excellent clinical-skill achievement. The integrated PBL-CBL-CTTM teaching (experimental group 2) yielded the highest area under the curve (AUC = 0.873), which was significantly higher than both experimental group 1 and the control group (DeLong test with Bonferroni correction, both $P < 0.017$). Figure 3 (heatmap) demonstrated that experimental group 2 scores were uniformly higher across hands-on experience and emotional dimensions, while the three groups showed greater overlap in the verbal persuasion dimension.

Discussion

The theoretical findings of this study are consistent with the residency teaching study by Zhu et al. (2025), confirming that multimodal integration is superior to LBL alone[12]. PBL + LBL only enhances theoretical knowledge, and the absence of supervised hands-on training and visualization tools precludes improvement in critical care management capabilities[18]. This pattern suggests that PBL alone primarily improves theoretical knowledge, while the combination of CBL and CTTM constitutes the core module for enhancing hands-on competence[19]. The three-stage closed-loop system developed in this study forms a complete translational chain: PBL facilitates pre-class autonomous internalization, CTTM simplifies abstract metabolic pathways, CBL provides comprehensive supervised hands-on experience, and post-class debriefing addresses common error points[2]. Due to grouping limitations, the independent effects of CBL and CTTM could not be separately quantified, but the existing data confirm that their combined use is central to hands-on skills improvement[4].

Building on Bandura's theory, a complete teaching-efficacy pathway was constructed: (1) Mastery experience (strongest source): Supervised complete critical case hands-on training continuously accumulates positive experiences. (2) Vicarious learning: Multidisciplinary case discussions and standardized demonstrations by senior faculty refine clinical reasoning. (3) Multi-layered verbal persuasion: Continuous correction of clinical biases across online platforms, classrooms, ward rounds, and debriefing sessions. (4) Positive affect buffering: CTTM simulation combined with graduated hands-on practice alleviates fear of DKA and thyroid storm management[28]. This study identifies endocrine-specific critical care fear as a distinct clinical education construct, complementing prior reports of procedure-specific anxiety in ICU training settings[7]. Single bedside feedback is limited by workload constraints, and the addition of weekly debriefing can optimize guidance efficiency[10, 29].

The dissemination value is stratified: At the resident level, it addresses the pain point of 'strong theoretical knowledge but reluctance to perform hands-on procedures.' At the preceptor level, it standardizes key and difficult teaching points, reducing preparation burden. At the educational administration level, it can be integrated into the unified residency curriculum. At the patient safety level, it reduces medication adjustment errors. A community hospital adaptation plan: In settings without 3D equipment, CTTM reconstruction content can be omitted while retaining animated micro-lectures; the core PBL + CBL closed-loop should be preserved, requiring no additional equipment.

Study limitations: (1) Single-center sample homogeneity limits generalizability; future multi-center studies should stratify participants across hospitals of different tiers. (2) Self-reported scales are subject to social desirability bias, and the self-developed scale was validated only at our institution; future studies should incorporate objective hands-on assessments and multi-center validity verification. (3) The absence of single-module control groups precludes disentangling independent effects. (4) Only short-term immediate assessments were conducted, without 6- or 12-month longitudinal follow-up. (5) Interviews employed non-random sampling, limiting representativeness. Based on the above-mentioned limitations, future multicenter studies should add single-module subgroups to dissect the contribution of each teaching component. Investigators should enroll hospitals at different tiers for external validation, incorporate objective assessments by supervising physicians and real-world clinical hard endpoints, conduct 6-12-month longitudinal follow-up to observe the long-term benefits of the training program, deliver standardized training for teaching faculty and implement quality control for treatment fidelity, and include interviews with preceptors in the qualitative component to further strengthen the body of evidence.

Conclusions

The integrated online-offline closed-loop multimodal PBL-CBL-CTTM teaching approach significantly improves theoretical knowledge, clinical hands-on proficiency, and learning satisfaction in endocrine and metabolic critical illness among internal medicine residents on rotation. Leveraging pre-class PBL autonomous preparation, ward-supervised CBL training, and full-cycle CTTM visualization assistance, the program comprehensively activates Bandura's four sources of self-efficacy, effectively alleviates fear of high-risk clinical scenarios such as DKA and thyroid storm, and establishes a low-cost, replicable tiered mentorship system that can be differentially disseminated based on the resources of tertiary and community hospitals.

Declarations

Ethics approval and consent to participate

This single-center, quantitatively driven, assessor-blinded, convergent parallel mixed-methods randomized controlled trial was reviewed and approved by the Ethics Committee of the Fourth Affiliated Hospital of Anhui Medical University (Ethics approval number: KYXM202508021). All procedures performed in this study were in accordance with the Declaration of Helsinki. Written informed consent was obtained from every resident prior to enrollment. Participants were fully informed of the study purpose, teaching intervention scheme, assessment contents, focus-group interview recording, data anonymization principle as well as the right to withdraw from the research at any time without penalty.

Consent for publication

Publication agreement has been secured from all contributing authors.

Competing interests

The authors declare no financial or non-financial competing interests.

Funding

This research is supported by the Anhui Provincial Health Science and Technology Project (No. AHWJ2025BAp20022).

Author Contribution

JZ and JD conceived and designed the study, supervised the data analysis, and drafted the original manuscript. JZ and LZ were responsible for the critical interpretation of research findings. JD revised the manuscript extensively for important intellectual content. All authors contributed to the collection and analysis of relevant literature data, and approved the final version of the manuscript for publication.

Acknowledgement

The authors thank all participating internal medicine residents and clinical preceptors of the Department of Endocrinology and Metabolism,The Fourth Affiliated Hospital of Anhui Medical University.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to institutional data governance policies and the presence of potentially identifying resident information, but are available from the corresponding author on reasonable request.

References

1. Veauthier B, Levy-Grau B. Diabetic Ketoacidosis: Evaluation and Treatment. Am Fam Physician. 2024;110(5):476–86.
2. Cohen S, Vempati A, Lei C, Moss H, Moadel T, Bentley S, Stapleton S, Roszczynialski K. Clinical Decision-Making Case: Thyroid Storm. J Educ Teach Emerg Med. 2025;10(5):Ce260–91.
3. Yuen B, Asmar A, Kibble J. Trainee Identification of the Self-efficacy Domains Needed to Succeed in Undergraduate and Graduate Medical Education. Med Sci Educ. 2025;35(3):1587–95.
4. Farooqi S, Raj S, Koyfman A, Long B. High risk and low prevalence diseases: Thyroid storm. Am J Emerg Med. 2023;69:127–35.
5. Abisad DA, Glenn Lecea EM, Ballesteros AM, Alarcon G, Diaz A, Pagan-Banchs P. Thyroid storm in pediatrics: a systematic review. J Pediatr Endocrinol Metab. 2023;36(3):225–33.

6. Jain PN, Hametz P, Banker SL, Escalante E, Gutierrez F, Dodoo C, Dwivedi AK, Beck J, Fromme HB. Flying Solo: Effectiveness of Attendingless Family-Centered Rounds on Improving Resident Autonomy. Hosp Pediatr. 2021;11(10):e218–30.
7. Belhomme N, Lescoat A, Launey Y, Robin F, Pottier P. With Great Responsibility Comes Great Uncertainty. J Gen Intern Med. 2025;40(1):54–62.
8. Kopstick AJ, Sanders BW, Yarris LM, Kelly SP. Preparing for the PICU: A Qualitative Study of Residents as They Prepare for Their First Pediatric Critical Care Rotation. J Pediatr Intensive Care. 2023;12(3):210–8.
9. Du Y, Qiao L, Dong L, Wan C, Yang X, Liu H. The relationship between self-efficacy, resilience, and job burnout in pediatric residents: a cross-sectional study in Western China. BMC Med Educ. 2024;24(1):787.
10. Wang XJ, Kumbamu A, Dunleavy KA, White JD, Dyrbye LN, Leep Hunderfund AN. It's such a shock at the beginning: a multispecialty qualitative interview study exploring first-year resident perspectives on the transition to residency. Acad Med. 2026;101(2):198–213.
11. Bernstein SA, Hanson JL, Slat EA, Duncan JG. Medical Resident Perceptions of Distress during Training. South Med J. 2024;117(12):719–24.
12. Zhu Y, Zhang J, Fei J, Fang H, Zhang Z. Problem-Based Learning and Case-Based Learning in Clinical Practical Teaching for Gynecology Residents: A Narrative Review. Adv Med Educ Pract. 2025;16:1269–79.
13. Ardoin TW, Hamer D, Stumpf M, Miles L. Integrating Problem-Based Learning Into an Internal Medicine Residency Curriculum. Ochsner J. 2022;22(4):324–43.
14. Duan Y, Li Z, Wang X, Gao Z, Zhang H. Application of online case-based learning in the teaching of clinical anesthesia for residents during the COVID-19 epidemic. BMC Med Educ. 2021;21(1):609.
15. Freeman W, Kanouse A, Frank GR. Online case-based learning to enhance pediatric endocrinology resident education. BMC Med Educ. 2026;26(1):206.
16. Wang J, Li W, Dun A, Zhong N, Ye Z. 3D visualization technology for Learning human anatomy among medical students and residents: a meta- and regression analysis. BMC Med Educ. 2024;24(1):461.
17. Gonzalez-Vargas JM, Sinz E, Moore JZ, Miller SR. Stepping into the Virtual Realm: Examining the Utility of Non-Immersive VR on Medical Residents' Self-Efficacy, Performance, and Satisfaction. Hum Factors Health 2024, 5.
18. Yu R, Cheng C, Zhang F. Enhancing competency and self-directed learning in anesthesiology residency: an outcome-based education model integrating online-offline hybrid teaching and mind mapping: a randomized controlled trial. Front Med (Lausanne). 2025;12:1684116.
19. Meunier PY, Schlatter S, Macabrey J, Zorzi F, Colleony T, Boussageon R, Maisonneuve H, Lamort-Bouché M. Blended Learning Compared With Face-to-Face Learning Among Family Medicine Residents: Randomized Controlled Trial. JMIR Med Educ. 2026;12:e86387.

20. Xu Y, Lin L, Qiu P, Luo S, Yan Y, Zhong L. Application of Online-Offline Teaching Combined with SimMan 3G Simulation Teaching Model in Critical Illness Teaching in a Department of Cardiology. Altern Ther Health Med 2024.
21. Zhao YY, Zhang TT, Li LH, Liu Q, Peng LJ, Wang Q, Wang W, Yu WY. Application of online to offline teaching mode in the training of non-anesthesiology residents in the department of anesthesiology: a randomized, controlled trial. Front Med (Lausanne). 2024;11:1329538.
22. Guetterman TC, Fetters MD, Creswell JW. Integrating Quantitative and Qualitative Results in Health Science Mixed Methods Research Through Joint Displays. Ann Fam Med. 2015;13(6):554–61.
23. Jacobs KG, Kugler J, Chi J, Stuart E, Bereknyei Merrell S, Rassbach C. A Mixed Methods Approach to Understanding Curricular Impact of a Capstone Course on the Self-Efficacy of Fourth-Year Medical Students. Cureus. 2020;12(8):e9537.
24. Durling P, Henni J, Mrozowich D, Rankin J, Barlow A, Grimminck R. A Mixed-Methods Realist Analysis of an Interdisciplinary Simulation Intervention for Psychiatry Residents. Acad Psychiatry. 2022;46(1):95–105.
25. Monzani A, Corti E, Scalogna A, Savastio S, Pozzi E, Sainaghi PP, Corte FD, Rabbone I. Effectiveness of a standardized scenario in teaching the management of pediatric diabetic ketoacidosis (DKA) to residents: a simulation cross-sectional study. BMC Med Educ. 2024;24(1):345.
26. Benjamin J, Roy K, Paul G, Kumar S, Charles E, Miller E, Narsi-Prasla H, Mahan JD, Thammasitboon S. Improving Resident Self-Efficacy in Tracheostomy Management Using a Novel Curriculum. *MedEdPORTAL* 2020, 16:11010.
27. Bowers P, Tomlin D. Assessing the Influence of Virtual Patients on Self-Efficacy: Perspectives From Audiology Students. Nurs Health Sci. 2025;27(3):e70198.
28. Xin C, Song X, Wang S, Cui X, Ding N, Wen D. Assessing professional behaviors: a self-administered scale for medical students during clerkships. BMC Med Educ. 2024;24(1):692.
29. Li F, Zhao H, Huang Q, Zhang C, Chen G. Development and validation of a competency-based evaluation framework for clinical medical trainees in the era of artificial intelligence: a mixed-methods study in China. BMC Med Educ 2026, 26(1).

Figures

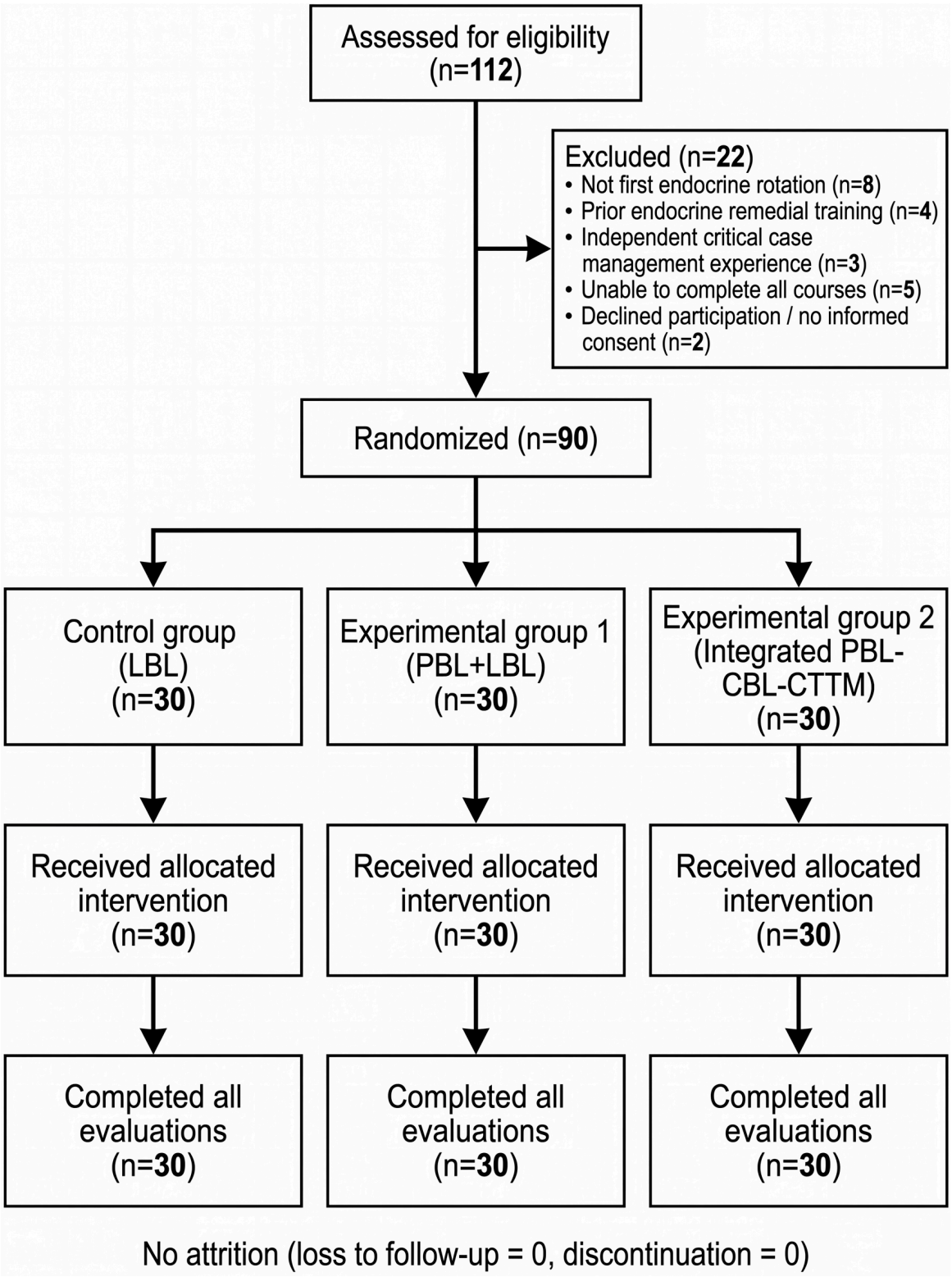


Figure 1

Flow of participants through the trial. A total of 112 internal medicine residents were assessed for eligibility, 22 were excluded, and 90 were randomized 1:1:1 to the LBL control group (n = 30), the PBL+LBL group (n = 30), or the integrated PBL-CBL-CTTM group (n = 30). All participants completed the trial with no attrition; intention-to-treat analysis included all 90 randomized residents.

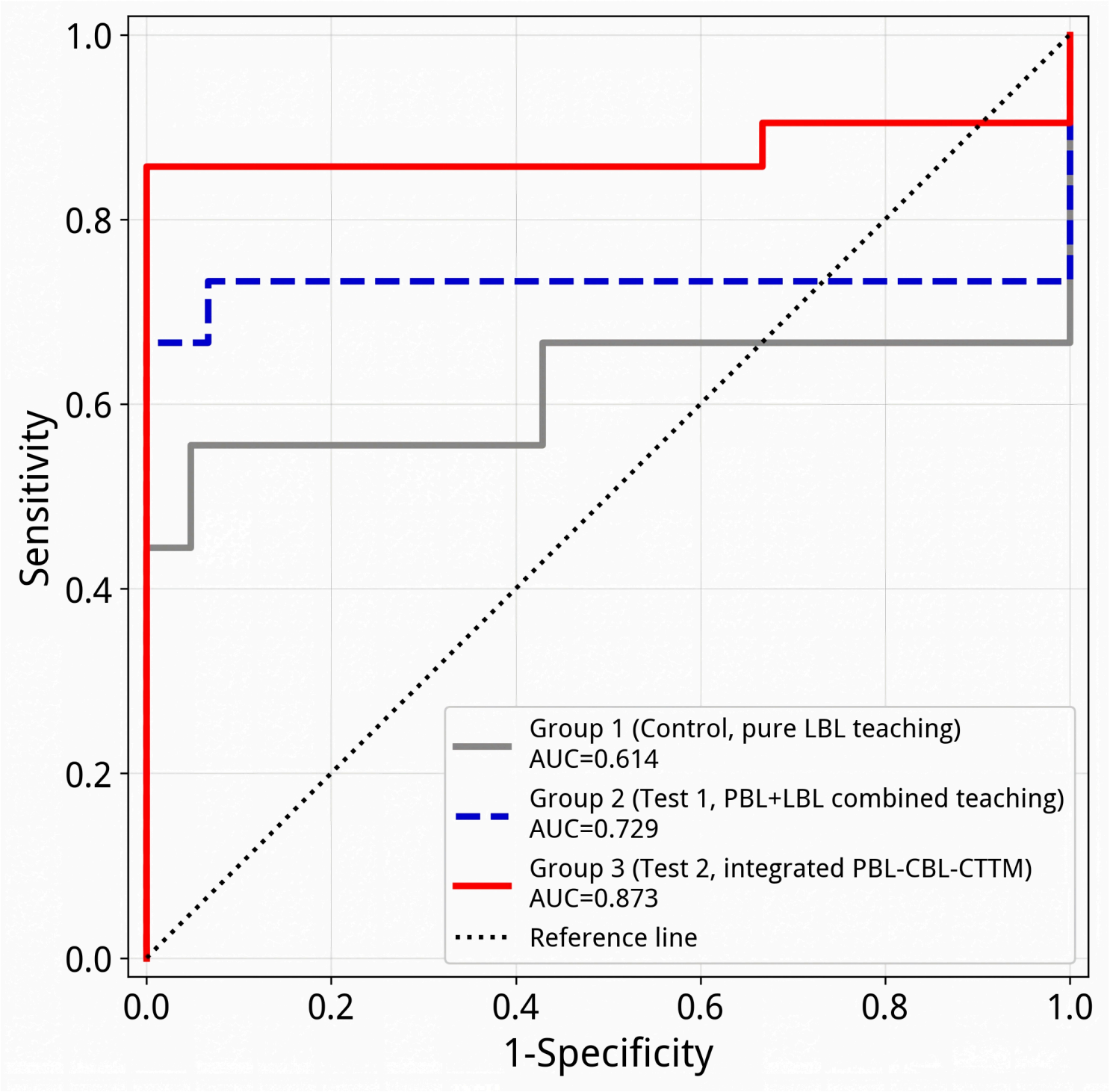


Figure 2

Receiver operating characteristic (ROC) curves for predicting excellent clinical-skill performance (defined as clinical skills score ≥45/50) under three teaching modes. The integrated PBL-CBL-CTTM teaching (experimental group 2) yielded the highest area under the curve (AUC = 0.873), which was significantly higher than both the control and experimental group 1 (DeLong test with Bonferroni correction, both $P < 0.017$).

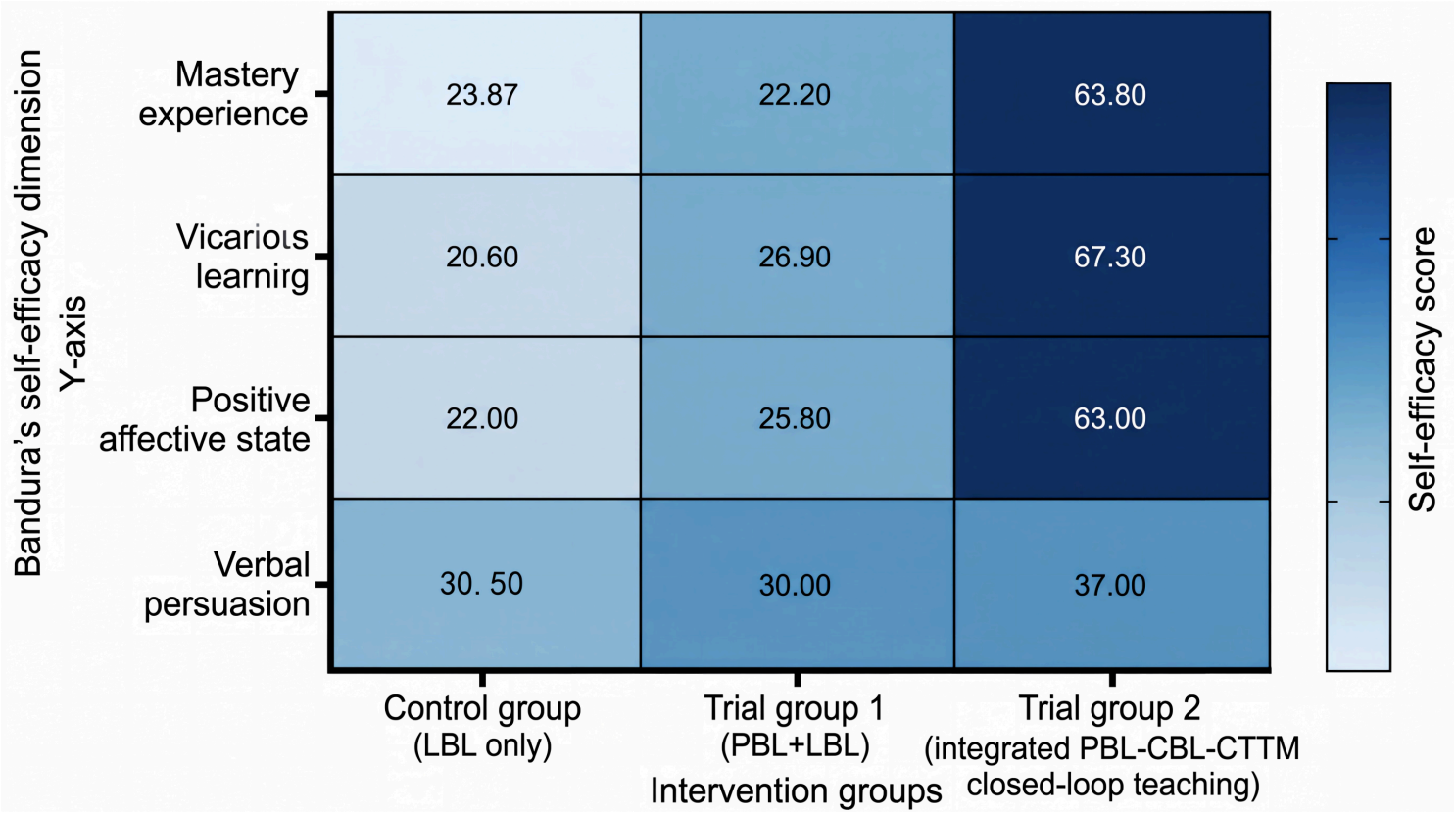


Figure 3

Heatmap of scores across the four sources of self-efficacy (mastery experience, vicarious learning, verbal persuasion, positive affect) among the three groups. Experimental group 2 scores were uniformly higher across hands-on experience and emotional dimensions, while the three groups showed greater overlap in the verbal persuasion dimension.

Supplementary Files

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

- Supplementarymaterial.docx