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.

Synergistic Enhancement: A Randomized Controlled Trial of AI-Virtual Simulation Combined with Traditional Model Training to Improve Thoracentesis Skill Transfer Efficacy in Resident Physicians

Haonan Zhang
Department of Hepatopancreatobiliary Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

Heng Yang
Department of Cardiothoracic and Vascular Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

Zhengwen Lei
Department of Cardiothoracic and Vascular Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

Jia Liu
Department of Cardiothoracic and Vascular Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

Juan Luo
Department of Cardiothoracic and Vascular Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

Jieyu Cao
Department of Cardiothoracic and Vascular Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

Kejiang Long
Department of Cardiothoracic and Vascular Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

Juan Wu
15886406878@163.com

Department of Cardiothoracic and Vascular Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

---

Research Article

Keywords: Blended Learning, Thoracentesis, Standardized Residency Training, Artificial Intelligence, Clinical Skills Training, Multidimensional Evaluation System

Posted Date: April 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9224471/v1

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

Additional Declarations: No competing interests reported.

Synergistic Enhancement: A Randomized Controlled Trial of AI-Virtual Simulation Combined with Traditional Model Training to Improve Thoracentesis Skill Transfer Efficacy in Resident Physicians

Haonan Zhang[2]+, Heng Yang[1]+, Zhengwen Lei[1], Jia Liu[1], Juan Luo[1], Jieyu Cao[1], Kejiang Long[1], Juan Wu[1]*

[1] Department of Cardiothoracic and Vascular Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China
[2] Department of Hepatopancreatobiliary Surgery, The First Affiliated Hospital of Hengyang Medical School, University of South China, Hengyang, Hunan 421001, China

+ Haonan Zhang and Heng Yang contributed equally to this work and share first authorship.

* Corresponding author: Juan Wu，Email: 15886406878@163.com

Background: While artificial intelligence virtual simulation (AI-VS) offers standardization advantages in clinical skills training, its inherent "digital divide" impedes the transfer of skills to real-world procedures. Conversely, traditional model training (TMT) provides essential tactile feedback but often lacks efficiency and standardization in delivering that feedback.
Methods: This study employed a double-blind, randomized controlled trial design. Thirty-six resident physicians were randomly allocated into either the experimental group (AI-VS + TMT, n=18) or the control group (AI-VS only, n=18). A time-matched control protocol ensured equal total training duration. Upon achieving proficiency in AI-VS, the experimental group progressed to the TMT phase, while the control group engaged in additional, duration-matched AI-VS modules. Procedural skills (primary outcome), theoretical knowledge, and teaching satisfaction were assessed before and after the training.
Results: All measured outcomes demonstrated significant improvement from baseline in both groups (all *P* < 0.001). However, the experimental group achieved significantly higher post-training scores in core technical skills (H) compared to the control group (91.72 ± 2.78 vs. 83.28 ± 4.74, *P* < 0.001; Cohen's d = 2.16). Furthermore, the experimental group exhibited comprehensive superiority across multiple other dimensions, including guideline mastery (G), theoretical knowledge (A), and teaching satisfaction (F).
Conclusion: The "AI-VS + TMT" hybrid model, through a synergistic "digital cognitive construction—physical tactile calibration" mechanism, effectively bridges the skill transfer gap. It reorients the instructor's role toward cultivating higher-order clinical thinking, providing an empirically supported novel paradigm for clinical skills education in the post-AI era.
Keywords: Blended Learning, Thoracentesis, Standardized Residency Training, Artificial Intelligence, Clinical Skills Training, Multidimensional Evaluation System.

Thoracentesis is a critical diagnostic and therapeutic procedure for pleural effusion[1]. Proficiency in this procedure directly impacts patient safety and outcomes[2], making it an essential core clinical skill that resident physicians must master.However, the traditional training paradigm relying on

apprenticeship-based instruction and model-based practice has long been plagued by persistent limitations, including an uneven distribution of qualified instructors, feedback that is often delayed and subjective, and a lack of standardized assessment systems.In the training of clinical procedural skills, the introduction of simulation trainingand assessment tools has opened new avenues for enhancing physicians' operational competence[3]. Research has developed a novel assessment tool specifically designed to evaluate physician performance in thoracentesis. Concurrently, simulation-based training ensures that physicians attain sufficient competency before engaging in actual procedures. This simulation-based approach effectively addresses the shortcomings inherent in the traditional apprenticeship model by delivering more standardized and timely feedback.In recent years, artificial intelligence virtual simulation (AI-VS) technology has offered a novel pathway to address these bottlenecks.Exemplifying this trend, virtual surgical assistant systems utilize AI to provide explanatory feedback, assisting trainees in performing brain tumor resection tasks within virtual reality environments. These systems deliver automated, proficiency-tailored educational feedback based on expert performance benchmarks, which has been shown to significantly enhance trainees' skill acquisition.[4]Furthermore, virtual reality simulation technology has played a significant role in the training of neurosurgery residents. By simulating the management of common neurosurgical conditions, it effectively enhances trainees' anatomical knowledge and procedural skills.[5] AI-VS, through high-fidelity virtual environments, enables risk-free, repeatable practice and delivers real-time, quantitative, precise feedback, thereby efficiently constructing standardized cognitive schemas of procedural workflows. This demonstrates significant potential for improving the accessibility and efficiency of training.However, as the application of AI-VS deepens, its inherent limitation—the "digital divide"—has become increasingly apparent. The purely virtual environment has an intrinsic deficiency in providing critical tactile feedback (e.g., the popping sensation of needle penetration), force perception, and handling real-world physical uncertainties. This leads to a prominent contradiction: trainees may perform excellently within the virtual system, yet their skills struggle to transfer completely and reliably to the bedside of real patients, creating a disconnect between "virtual competence" and "real-world performance."

In contrast, traditional model training (TMT) provides indispensable real tactile sensation and physical interaction, but its feedback is often less immediate, objective, and difficult to scale.Consequently, the core challenge facing medical educators today is no longer a simple choice between AI and traditional methods, but rather how to systematically and structurally integrate the strengths of both to generate a synergistic "1+1>2" effect. Addressing this challenge urgently requires guidance from educational theory.Clinical trial number: Not applicable. This study was not a registered clinical trial as it did not involve any investigational drug or device intervention.

Methods

Study Design and Ethics

This study was a single-center, prospective, double-blind, randomized controlled trial. The protocol was approved by the Ethics Committee of the First Affiliated Hospital of University of South China (IRB approval number: 2023LL0927001). All research procedures strictly adhered to the principles of the Declaration of Helsinki. All participants provided written informed consent after being fully informed of the study details prior to enrollment.

Participants

From December 2023 to October 2024, 36 first-year clinical resident physicians undergoing

standardized training at the First Affiliated Hospital of University of South China were consecutively enrolled.Inclusion criteria were as follows: (1) New entry into the standardized residency training system after obtaining an undergraduate degree in clinical medicine; (2) Age between 22 and 24 years; (3) No prior independent experience or systematic training in thoracentesis; (4) Voluntary participation and commitment to comply with the study arrangements.Exclusion criteria included: (1) Any previous virtual simulation training specific to thoracentesis; (2) Concurrent participation in other clinical skills educational interventions within the 3 months prior to the study initiation that could potentially confound the results; (3) Diagnosed visual, auditory, or upper limb motor dysfunction that would impair procedural performance; (4) Anticipated attendance rate during the study period deemed insufficient based on the rotation schedule.Termination criteria: Participants were withdrawn from the study if any of the following occurred: (1) A serious adverse medical event related to the study procedure; (2) Unexcused absence from two or more scheduled training sessions; (3) Voluntary withdrawal requested by the participant.

Instructional Materials

The National Health Commission Planning Textbook for Standardized Residency Training (Surgery Volume) and the Bedside Teaching Guide for Clinical Procedural Skills in Standardized Residency Training (2022 Edition) served as the unified curriculum and primary reference materials. The procedural guide for the operation was specifically based on the Bedside Teaching Guide for Clinical Procedural Skills in Standardized Residency Training (2022 Edition). Theoretical knowledge assessment content for thoracentesis was selected from the National Health Commission Planning Textbook for Standardized Residency Training: General Surgery Volume (2nd Edition) (People's Medical Publishing House) and Surgery (10th Edition) (People's Medical Publishing House).

Teaching Team

The supervising experts and instructional team were all affiliated with the First Affiliated Hospital of University of South China. Supervising experts were required to hold a senior professional title (Chief or Associate Chief Physician) and possess over ten years of clinical experience in thoracentesis. Peer instructors and tutors were required to hold an intermediate or higher professional title (Attending Physician or above) with a minimum of five years of practical experience in the relevant field. All personnel held provincial-level residency teaching qualifications in Hunan Province. Furthermore, all team members underwent standardized training to ensure consistency in teaching protocols and assessment criteria, thereby guaranteeing the professionalism and uniformity of both instructional delivery and evaluation.

Randomization and Allocation Concealment

Sequence Generation:A block randomization sequence (block size = 4) was generated using computer software to allocate the 36 participants in a 1:1 ratio to either the experimental group (AI-VS + TMT) or the control group (AI-VS only).Allocation Concealment: The generated random sequence was kept in a sealed envelope by an independent research administrator who was not involved in any subsequent stages of the study. When a participant completed the baseline assessment and was formally enrolled, the study coordinator contacted this administrator. The administrator then provided the corresponding group assignment according to the sequence. This procedure ensured that neither the research team nor the participants could foresee the allocation prior to enrollment.

Implementation of Blinding Procedures

Blinding of Participants: Throughout recruitment and the entire study process, all participants were informed: "This study aims to compare two different, complete clinical skills training programs. You

will be randomly assigned to one of them. Both programs include practice on simulators and models, but they differ in focus and sequence." This statement was technically truthful but skillfully concealed the core hypothesis that the experimental group received the full 'AI + physical' training while the control group received AI-only training. This led participants to believe they were receiving a reasonable "combined program," not a reduced one, thereby effectively mitigating potential disappointment or negative placebo-like effects from knowing they were in the "control" group. Procedural Masking: Trainees from both groups were scheduled for sessions in the physical model training room (the experimental group for TMT, the control group for an unrelated skill like wound suturing), making their external activities appear similar and further strengthening the blinding.Blinding of Outcome Assessors: The supervising experts and peer instructors responsible for the final unified procedural skills assessment were not involved in any training sessions and were completely unaware of the study's grouping hypothesis. Participant Anonymization: During assessment, all participants were identified only by a random study ID; their names and group allocation were hidden from the assessors. Environmental Standardization: The assessment environment was uniform, utilizing only traditional physical models and devoid of any items (such as AI simulators) that could potentially indicate group allocation.

Intervention Protocol

All training sessions were conducted at the Clinical Skills Center over a period of two weeks. Both the experimental and control groups utilized equipment from the same batch and of identical specifications: the AI virtual simulator (the Touch Surgery™ application provided by Medtronic) and the physical models (available at the Clinical Skills Center of the First Affiliated Hospital of University of South China). The core procedural workflow was standardized and identical for both groups.Experimental Group (AI-VS + TMT, n=18):This group received sequential, hybrid training.Phase 1 (AI-VS):Trainees performed a standardized case on the AI simulator, which provided real-time quantitative feedback. A stringent proficiency gate was implemented: advancement to the next phase required achieving a simulator score of ≥90 points (out of 100) in two consecutive attempts. This phase aimed to establish standardized cognitive schemata and muscle memory in a risk-free environment, with an average duration of approximately 3.5 hours.The standardized procedure of AI virtual simulation training is shown in Figure 2.Phase 2 (TMT):Within three days of meeting the AI-VS proficiency criterion, trainees participated in a single 2-hour, small-group, instructor-led session. The instructional focus was threefold: (1)Tactile Calibration:Guiding trainees to experience and cognitively associate the characteristic "loss-of-resistance" sensation upon needle penetration of the pleura; (2) Force Feedback:Developing perception of real tissue resistance and the tactile feedback during fluid aspiration; (3)Managing Uncertainty:The instructor introduced simulated complications (e.g., verbal cue for "patient cough") during practice to train clinical adaptability and decision-making skills. The objective was to facilitate the transition from performing a "correct procedure" to executing a "robust procedure.The corresponding steps performed on the physical model are illustrated in Figure 3.Control Group (AI-VS only, n=18):This group received AI-VS training exclusively.They completed the identical AI-VS core module as the experimental group and were held to the same proficiency standard (two consecutive scores ≥90).Time-Matching Control:To rigorously control for the potential confounding variable of "total training duration," participants in the control group who achieved the proficiency standard early were required to complete an additional cumulative 2 hours of supplementary AI-VS module training(e.g., modules involving thoracentesis in different patient positions). This ensured their total training time (5.5 hours) was strictly matched to that of the

experimental group. Critically, the skill sets emphasized in these supplementary modules were deliberately distinct from those evaluated in the final assessment (standard supine thoracentesis). This design prevented conferring a targeted, practice-related advantage to the control group for the primary outcome, thereby strengthening the inference that any observed between-group differences were attributable to the fundamental difference in training content (the presence or absence of the TMT component).

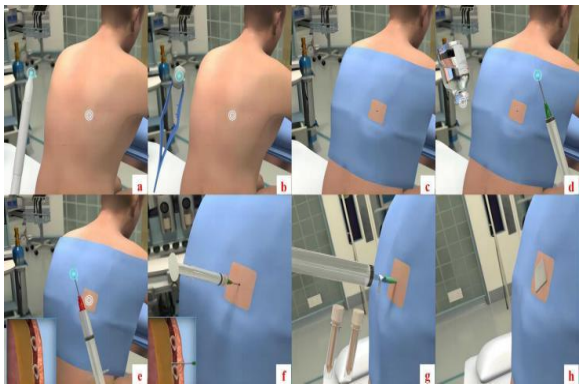


Figure 2 AI Virtual Thoracentesis Training Procedure (Touch Surgery™ application provided by Medtronic, USA): (A) Positioning → (B) Disinfection → (C) Draping → (D) Anesthesia → (E) Fine needle aspiration → (F) Large-bore needle drainage → (G) Effusion culture → (H) Dressing application.



Figure 3 Physical Model Thoracentesis Training Procedure (Clinical Skills Center, the First Affiliated Hospital of University of South China): (A) Positioning → (B) Disinfection → (C) Draping → (D) Anesthesia → (E) Fine needle aspiration → (F) Large-bore needle drainage → (G) Effusion culture → (H) Dressing application.

Statistical Analysis

Data were processed using SPSS version 26.0. Quantitative data conforming to a normal distribution are presented as mean ± standard deviation. Between-group comparisons were performed using independent samples t-tests, while within-group pre-post comparisons employed paired samples t-tests. Categorical data are presented as frequency (percentage), with between-group comparisons conducted using the chi-square (χ^2) test. Non-normally distributed quantitative data are presented as median (interquartile range), and between-group comparisons were made using the Mann-Whitney U test (with Z-values reported). All tests were two-tailed, and a P-value of less than 0.05 was considered statistically significant.

Results

1.Baseline Characteristics

A total of 36 resident physicians were enrolled and randomly allocated to either the experimental or control group, with 18 participants in each. The baseline demographic characteristics were well-balanced between the two groups, as detailed in Table 1. Gender distribution was identical, with 9 males (50.0%) and 9 females (50.0%) in each group (χ^2 test, P = 0.632). The median age was 25.00 years (IQR: 24.00, 25.00) in the experimental group and 24.75 years (IQR: 25.00, 26.00) in the control group (Mann-Whitney U test, P = 0.163).

Table 1. Comparison of Baseline Characteristics Between the Two Groups of Resident Physicians

Characteristic	Experimental (n=18)	Control(n=18)	Z/χ^2	*P*-value

Gender	Male,n(%)	9(50.00%)	9(50.00%)	0	0.632
	Female,n(%)	9(50.00%)	9(50.00%)		
Age（years）		25.00(24.00,25.00)	24.75(25.00,26.00)	-1.396	0.163

2. Comparison of Outcomes Before and After Training in the Experimental Group

As presented in Table 2, the experimental group, which received the "AI-VS + TMT" hybrid training, demonstrated highly significant improvement across all evaluation metrics following the intervention ($P < 0.001$). Specifically, the median score for technical skills increased from 61.50 to 91.50 points; guideline mastery improved from 61.00 to 90.00 points; and the overall SEEQ satisfaction score rose from 150.00 to 200.50 points.

Table 2 Before-/After-Training Outcomes (Experimental Group)

Metric	Before-Training	After-Training	*Z*	*P*
Theoretical Knowledge	62.50(60.00,67.00)	90.00(89,91.75.00)	-3.734	< 0.001
Supervisory Experts	65.00(60.75,70.00)	90.50(85.00,96.25)	-3.727	< 0.001
Peer Instructors	63.50(60.00,69.75)	91.00(88.75,95.25)	-3.726	< 0.001
Instructor evaluate Resident	45.50(42,50.25.00)	63.00(56.00,66.25)	-3.574	< 0.001
Resident evaluate Instructor	31.00(30.00,35.50)	45.50(42.75,48.00)	-3.726	< 0.001
SEEQ Satisfaction	150.00(141.00,158.25)	200.50(191.00,207.25)	-3.681	< 0.001
Guideline Mastery	61.00(58.75,65.25)	90.00(87.25,92.00)	-3.728	< 0.001
Technical Skills	61.50(59.75,64.00)	91.50(89.00,94.00)	-3.728	< 0.001

3. Comparison of Outcomes Before and After Training in the Control Group

As shown in Table 3, the control group, which received the single-modality "AI-VS" training, also demonstrated highly significant improvement across all assessment metrics following the intervention ($P < 0.001$). Specifically, the median score for technical skills increased from 60.00 to 83.50 points; guideline mastery improved from 61.50 to 81.50 points; and the overall SEEQ satisfaction score rose from 149.00 to 191.00 points.

Table 3 Before-/After-Training Outcomes (Control Group)

Metric	Before-Training	After-Training	*Z/t*	*P*
Theoretical Knowledge	60.50(59.00,64.25)	88.00(85.00,89.25)	-3.732	< 0.001

Supervisory Experts	66.78±7.915	84.22±5.683	-8.178	< 0.001
Peer Instructors	67.50(64.00,74.50)	89.00(82.00,94.25)	-3.204	< 0.001
Instructor evaluate Resident	46.00(42.75,51.00)	59.00(56.50,60.00)	-3.579	< 0.001
Resident evaluate Instructor	32.00(30.00,35.00)	43.50(40.00,46.25)	-3.625	< 0.001
SEEQ Satisfaction	149.00(145.25,162.25)	191.00(185.75197.00)	-3.724	< 0.001
Guideline Mastery	61.50(59.50,63.50)	81.50(79.75,85.00)	-3.727	< 0.001
Technical Skills	60.00(58.00,64.00)	83.50(79.75,85.25)	-3.742	< 0.001

4. Comparison of Multidimensional Scores Between Groups Before Training

Before the commencement of training, a between-group comparison of all assessment metrics was performed. The results are presented in Table 4. No statistically significant differences were observed between the two groups in theoretical knowledge (P = 0.078), baseline supervisory expert evaluation (P = 0.578), technical skills (P = 0.288), guideline mastery (P = 1.000), or any other assessment dimensions (P> 0.05). This confirms that both groups were comparable across all measured parameters at baseline.

Table 4 Before-Training Multidimensional Score Comparison Between Groups

Metric	Experimental Group	Control Group	*Z/t*	*P*
Theoretical Knowledge	62.50(60.00,67.00)	60.50(59.00,64.25)	-1.765	0.078
Supervisory Experts	65.00(60.75,70.00)	66.50(62.50,70.00)	-0.556	0.578
Peer Instructors	63.50(60.00,69.75)	67.50(64.00,74.50)	-1.840	0.066
Instructor evaluate Resident	45.50(42.00,50.25)	46.00(42.75,51.00)	-0.558	0.577
Resident evaluate Instructor	31.00(30.00,35.50)	32.00(30.00,35.00)	-0.176	0.860
SEEQ Satisfaction	149.50±18.841	152.50±15.455	-0.522	0.605
Guideline Mastery	61.83±4.204	61.83±4.004	0	1
Technical Skills	61.61±2.638	60.50±3.485	1.079	0.288

5. Comparison of Multidimensional Scores Between Groups After Training

The between-group comparison results after training are presented in Table 5. The hybrid training

model demonstrated comprehensive superiority. For the primary outcome of technical operational skills, the score in the experimental group (91.72 ± 2.78) was significantly higher than that in the control group (83.28 ± 4.74) ($P < 0.001$). Regarding operational performance evaluations, the experimental group received significantly higher ratings from both supervisory experts ($P = 0.002$) and peer instructors ($P = 0.047$). In terms of theoretical knowledge depth, the experimental group also scored significantly better than the control group in both theoretical knowledge ($P = 0.011$) and guideline mastery ($P < 0.001$).

Table 5 After-Training Multidimensional Score Comparison Between Groups

Metric	Experimental Group	Control Group	*Z/t*	*P*
Theoretical Knowledge	90.28±2.718	87.78±2.819	2.708	0.011
Supervisory Experts	90.72±5.665	84.22±5.683	3.437	0.002
Peer Instructors	91.67±4.393	87.89±6.425	2.522	0.047
Instructor evaluate Resident	61.94±5.795	58.00±3.029	2.559	0.017
Resident evaluate Instructor	45.22±3.457	42.78±4.11	3.921	0.047
SEEQ Satisfaction	200.28±10.758	190.17±10.037	2.916	0.006
Guideline Mastery	90.00(87.25,92.00)	81.50(79.75,85.00)	5.014	< 0.001
Technical Skills	91.72±2.782	83.28±4.738	5.400	< 0.001

6. Analysis of Multidimensional Competency Comparison via Radar Chart Post-Training

As illustrated in Figure 4, the post-training multidimensional competency radar chart clearly demonstrates the performance disparities between the two groups across the eight assessment domains. The experimental group (AI-VS + TMT) achieved higher scores than the control group (AI-VS only) in every dimension. The advantages for the experimental group were most pronounced in the areas of technical skills, guideline mastery, and supervisory expert evaluation, where their scores were markedly higher. The experimental group also maintained stable advantages in theoretical knowledge, peer instructor evaluation, and SEEQ satisfaction. In contrast, although the experimental group still outperformed the control group in student-instructor mutual evaluations, the gap between the two groups was relatively narrower in these dimensions.

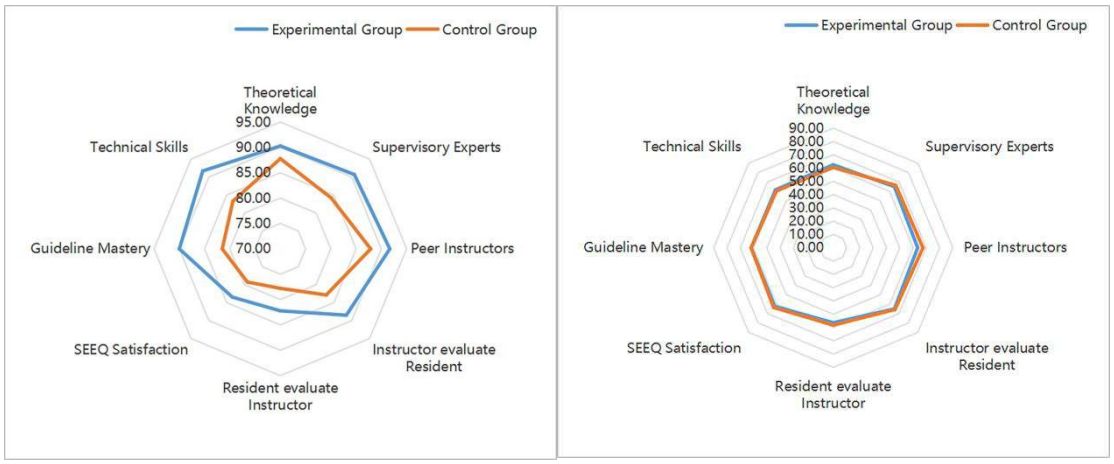


Figure 4 Multidimensional competency comparison before and after training. Left panel shows pre-training competency profiles; right panel shows post-training profiles. The experimental group (AI-VS + TMT) demonstrates superior performance across all eight dimensions compared to the control group (AI-VS only), with the most pronounced advantages in technical skills, guideline mastery, and supervisory expert evaluation.

7. Comparative Analysis of Skill Level Distribution Before and After Training

The changes in skill level distribution for both groups before and after training are depicted in Figure 5, vividly illustrating skill improvement at the individual level. Prior to training, the skill distributions of the two groups were similar,predominantly concentrated within the score range of 60-70 points, indicating a relatively centralized, moderate proficiency level. This finding aligns with and corroborates the baseline equilibrium results presented in Table 4. Following the training, both groups exhibited significant skill enhancement, yet the distribution patterns differed markedly. In the experimental group, skill scores were concentrated in the high-score segment above 90 points, with a more compact distribution. This suggests that the majority of trainees achieved a high and consistent skill level through the hybrid training. Although the control group also showed significant improvement, its score distribution was relatively dispersed, primarily clustered in the 80-85 point range, with some trainees scoring below 85 points. This indicates a degree of heterogeneity in skill mastery within the control group.

Discussion

This study, through a rigorous double-blind randomized controlled trial, provides the first systematic evaluation of the comprehensive efficacy of a hybrid model integrating "Artificial Intelligence Virtual Simulation (AI-VS) with Traditional Model Training (TMT)" in thoracentesis skill development for resident physicians. The core data reveal that the hybrid model (experimental group) not only achieved a significantly higher post-training score in technical operational skills compared to the AI-VS-only model (control group) (91.72 ± 2.78 vs. 83.28 ± 4.74, $P < 0.001$) but also demonstrated superiority across seven other key dimensions: theoretical knowledge, supervisory expert evaluation, peer instructor evaluation, student-instructor mutual evaluation, overall satisfaction, and guideline mastery. This constellation of findings robustly confirms that the hybrid model transcends the limitations of singular approaches, achieving a synergistic "1+1>2" effect.The classic Miller's pyramid framework emphasizes that the development of competence requires a progressive transition from "knows" to "shows how."[6]Our study reveals that the purely AI-based pathway faces a bottleneck at this critical transition. By breaking down complex skill acquisition into two focused phases—"cognitive

construction" and "physical calibration"—the hybrid model may more effectively manage learners' intrinsic and extraneous cognitive loads, thereby promoting schema automation and deeper knowledge integration. Thus, rather than merely reiterating the advantages in outcomes, the following discussion aims to delve deeply into the synergistic mechanisms behind the integrated "AI-VS + TMT" paradigm. First, drawing on cognitive science theory, we will deconstruct how this model achieves skill enhancement through dual-channel "digital simulation–physical embodiment" learning. Next, we will provide a balanced assessment of the core value of AI virtual training and the efficacy-enhancing role of the hybrid model. Finally, we will explore how this model reshapes the teaching and interaction ecosystem and advances educational equity.

1. The Cognitive-Somesthetic Dual-Channel Synergistic Mechanism

The core mechanism underlying the superior efficacy of the hybrid model lies in its sequential design, which systematically manages cognitive load and accomplishes critical somesthetic calibration.[7]This mechanism is most directly evidenced in the primary outcome measure of technical operational skills. While both groups showed highly significant pre-to-post improvement within themselves (Experimental: median 61.50 to 91.50, $P < 0.001$; Control: median 60.00 to 83.50, $P < 0.001$), the post-training between-group comparison revealed a significantly higher score for the experimental group (91.72 ± 2.78 vs. 83.28 ± 4.74, $P < 0.001$). Notably, the smaller standard deviation in the experimental group indicated more uniform skill mastery. This 8.44-point gap with statistical significance quantitatively represents the bridging of the "digital divide."The AI-VS phase efficiently optimized extraneous cognitive load, enabling trainees to establish solid, standardized cognitive schemata in both theoretical knowledge (Experimental: 62.50 to 90.00; Control: 60.50 to 88.00) and guideline mastery (Experimental: 61.00 to 90.00; Control: 61.50 to 81.50). However, the subsequent lack of authentic tactile feedback in the control group led to degradation of these schemata during transfer to physical operation.For the experimental group, the TMT phase acted as a "physical intelligence calibrator." Trainees associated and refined the visually acquired visuomotor mappings from the virtual environment with real somatosensory inputs—specifically, the "loss of resistance" upon pleural puncture and tissue feedback. This calibration process is supported by the supervisory expert evaluation: the experimental group's post-training score (90.72 ± 5.67) was significantly higher than the control group's (84.22 ± 5.68, $P = 0.002$), indicating experts could discern more robust and clinically concordant performance. The between-group difference in peer instructor evaluation (91.67 ± 4.39 vs. 87.89 ± 6.43, $P = 0.047$) further corroborates this, showing greater peer recognition for fluid and stable performance.Concurrently, despite significant within-group improvements, the experimental group maintained a between-group advantage in student-instructor mutual evaluations (Instructor evaluation of Resident: $P = 0.017$; Resident evaluation of Instructor: $P = 0.047$) and overall SEEQ satisfaction ($P = 0.006$). This suggests the "calibration" process itself was pedagogically interactive, enhancing the learning experience.In essence, the synergistic mechanism is this: AI-VS efficiently constructs a standardized "cognitive blueprint," and TMT accomplishes "physical anchoring" through somatosensory feedback. Their sequential interaction facilitates the critical transition from "knowing how" to "performing consistently and correctly in real-world clinical scenarios."

2. The Value Proposition of AI and the Efficacy-Enhancing Role of the Hybrid Model

The independent pedagogical value of standalone AI-VS training must first be acknowledged, as it forms the basis for understanding the hybrid model's role as an "efficacy enhancer" rather than a "replacement."Evidence suggests that the hybrid training strategy enhances pedagogical efficacy without necessitating additional data inputs.[8]By integrating data-driven and mechanism-driven

approaches, this strategy enables enhanced generalization capability in complex learning tasks.The application of hybrid training schedules has also demonstrated superiority over purely distributed or centralized training methods in certain tasks, particularly in sensorimotor tasks[9]The control group data provide robust evidence for this: after two weeks of pure AI training, participants showed highly significant within-group improvements across all eight assessment metrics (all *P* < 0.001). Particularly in building a standardized knowledge base and operational framework, theoretical knowledge scores increased from 60.50 to 88.00, guideline mastery improved from 61.50 to 81.50, and technical operational skills rose from 60.00 to 83.50. This unequivocally demonstrates the exceptional efficiency and scalability of AI-VS in providing risk-free repetitive practice, delivering immediate quantitative feedback, and ensuring procedural standardization—qualifying it as a powerful "standardized instructional engine."However, the hybrid model is positioned as a more comprehensive "instructional system." Post-training between-group comparisons across all eight metrics consistently showed that the experimental group's performance was superior or significantly superior to that of the control group. Among these, the between-group effect size was largest for guideline mastery (90.00 vs. 81.50, $P < 0.001$).This finding is of critical importance, as it elucidates a "reverse transfer effect" facilitated by the TMT phase within the hybrid model on theoretical consolidation: when encountering the uncertainties inherent in the physical model and instructor-simulated complications, trainees are compelled to actively and profoundly retrieve and integrate the abstract knowledge acquired during the AI-VS phase. This process effectively promotes the advancement of theoretical knowledge from simple "memorization" toward "contextualized understanding and practical application."[10] The observed intergroup advantages in assessments by both supervising experts and peers suggest that the skills cultivated through the hybrid model exhibit greater clinical authenticity and robustness, thereby receiving stronger endorsement from professional colleagues. This delineates the strategic roles within the model: AI-VS functions as an efficient "cognitive foundational layer" and "guardian of standardization," responsible for the scalable construction of core skill competencies. Conversely, TMT acts as the critical "competency enhancer" and "transfer bridge," dedicated to the contextual anchoring of skills and the stimulation of higher-order clinical thinking. These components are not competitive but rather assume complementary functions within a cohesive instructional sequence, jointly advancing toward the overarching objective of cultivating comprehensive clinical competence.

3. Reconceptualizing the Instructor’s Role: From Skill Coach to Cognitive Mentor

The hybrid model brings about not merely an enhancement in trainees' skills but, more fundamentally, a reconstruction of the quality of pedagogical interactions. At its core lies the strategic elevation of the instructor's role.[11] In terms of mutual student-instructor evaluations, both groups showed highly significant within-group improvement from baseline (both $P < 0.001$). However, the post-training between-group comparison revealed that the experimental group scored significantly higher than the control group in both Instructor evaluation of Resident (61.94 ± 5.80 vs. 58.00 ± 3.03, $P = 0.017$) and Resident evaluation of Instructor (45.22 ± 3.46 vs. 42.78 ± 4.11, $P = 0.047$). This pattern of data differences indicates a fundamental transformation in the content and quality of student-instructor interaction under the hybrid model.In the AI-VS-only pathway, the instructor's primary function is often constrained by the technological platform to that of a "system administrator" or "basic operation corrector," with interactions remaining at the level of technical details and limited in depth. Within the hybrid model, AI assumes the repetitive task of foundational skills training, liberating the instructor from tedious procedural monitoring. During the TMT phase, the instructor's expert cognitive resources are refocused on a higher-order pedagogical mission: the dialogue shifts from asking "Is your puncture

angle correct?" to guiding discussions on "How should you respond if the patient suddenly coughs? What is the anatomical and physiological rationale?" This form of interaction concentrates on cultivating critical thinking abilities related to complication management, clinical decision-making, and patient safety.This profound evolution in the pedagogical relationship directly drove the significant between-group difference in overall satisfaction (200.28 ± 10.76 vs. 190.17 ± 10.04, P = 0.006).This, in turn, led to an enhanced sense of gain and satisfaction among trainees, who benefited from more challenging and thought-provoking guidance.[12] Concurrently, the between-group advantage in peer instructor evaluations (P = 0.047) indirectly corroborates that trainees who received such in-depth guidance were able to demonstrate clearer clinical logic and more effective communication skills during procedural demonstrations and team-based collaboration.Thus, the hybrid model facilitates a strategic transformation of the instructor's role from a "skill coach" to a "clinical reasoning mentor and educational designer," reshaping the pedagogical ecosystem into a new, mutually-enhancing dynamic centered on cultivating higher-order thinking.[13] This pedagogical shift inherently emphasizes student agency in the learning process, positioning them as active participants rather than passive recipients, which in turn enhances overall learning effectiveness.[14] This transformation underscores the instructor's role as an educational designer, who not only imparts skills but, more importantly, guides students toward deep thinking and self-directed learning. Thereby, it achieves the goal of reciprocal enhancement between teaching and learning within the educational process.[15]

4. Promoting Skill Homogenization and Educational Equity

Another significant finding of this study is the "skill homogenization" effect inherent in the hybrid model. Specifically, the integrated instructional approach, which combines online and offline educational resources, has proven effective in enhancing both the theoretical knowledge and practical skills of resident physicians.The strength of this pedagogical approach lies in its capacity to offer more equitable access to learning opportunities for trainees of varied backgrounds and competency levels, thereby facilitating skill homogenization. This holds profound implications for ensuring both equity and quality consistency in large-scale residency training programs.[16] This effect is most directly reflected in the distribution characteristics of the technical operational skills scores. Post-training, the experimental group not only achieved a higher mean score but also exhibited a significantly smaller standard deviation (2.78 vs. 4.74 in the control group), with scores densely clustered in the high-performance bracket above 90 points. This indicates that the hybrid model not only raised the "ceiling" of skill levels but also effectively elevated the "floor," thereby reducing performance disparities attributable to individual differences in prior experience, spatial imagination, or cognitive styles.Data from other dimensions provide multi-faceted support for this "homogenization" effect. Although both groups showed substantial within-group improvement in theoretical knowledge and guideline mastery, the experimental group demonstrated consistently lower post-training intra-group dispersion and maintained significant between-group superiority. This suggests the hybrid pathway fostered a more uniform level of theoretical internalization. The between-group difference in peer evaluations (P = 0.047) further indicates that trainees in the hybrid group exhibited more homogeneous performance during demonstrations, earning consistently higher peer recognition.This effect stems from the inclusive design of the dual-channel learning architecture within the hybrid model. It provides a pathway for learners adept at abstract thinking to rapidly establish competence during the AI-VS phase, while simultaneously offering learners who rely on hands-on practice and somatosensory feedback a critical breakthrough opportunity during the TMT phase. Through the complementary processes of "virtual cognitive construction" and "physical embodied calibration," the model creates

diversified pathways to success for learners with different styles. Consequently, it achieves more balanced and predictable high-quality skill output at the cohort level. This holds significant practical value for advancing educational equity and ensuring quality control in standardized training programs.

Limitations and Future Directions

The single-center design and limited sample size (n=36), while achieving sufficient statistical power for the present study ($1-\beta > 0.99$), may constrain the generalizability of the findings to other institutions with different cultural or educational contexts. Additionally, the durability of the advantages demonstrated by the hybrid model remains unknown—whether the enhanced skills can be maintained over weeks or months (i.e., skill retention rate) warrants investigation.

Given these limitations, future research should focus on: 1) conducting multi-center, large-sample randomized controlled trials to verify the robustness and generalizability of this hybrid model across diverse educational settings; 2) implementing long-term longitudinal follow-up studies to map skill decay curves associated with different training modalities; and 3) providing an empirical basis for establishing evidence-based intervals for refresher training.

Reference

1.Deng, Mingming et al. “Ultrasonic Elastography-guided Pleural Biopsy for the Diagnosis of Pleural Effusion: A Multicenter Prospective Study of Diagnostic Test Performance.” *Annals of the American Thoracic Society* vol. 20,9 (2023): 1242-1249.

2.Schildhouse, Richard et al. “Safe and Effective Bedside Thoracentesis: A Review of the Evidence for Practicing Clinicians.” *Journal of hospital medicine* vol. 12,4 (2017): 266-276.

3.Rasmussen, Kasper Møller Boje et al. “Ensuring Basic Competence in Thoracentesis.” *Respiration; international review of thoracic diseases* vol. 97,5 (2019): 463-471.

4.Mirchi, Nykan et al. “The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine.” *PloS one* vol. 15,2 e0229596. 27 Feb. 2020,

5.Schirmer, Clemens M et al. “Virtual reality-based simulation training for ventriculostomy: an evidence-based approach.” *Neurosurgery* vol. 73 Suppl 1 (2013): 66-73.

6.Williams, Betsy White et al. “Miller's Pyramid and Core Competency Assessment: A Study in Relationship Construct Validity.” *The Journal of continuing education in the health professions* vol. 36,4 (2016): 295-299.

7.Bueno-Crespo, Andrés et al. “Bioinspired Architecture Selection for Multitask Learning.” *Frontiers in neuroinformatics* vol. 11 39. 22 Jun. 2017,

8.E Samadi, Moein et al. “A training strategy for hybrid models to break the curse of dimensionality.” *PloS one* vol. 17,9 e0274569. 15 Sep. 2022,

9.Paik, Jaehyon, and Frank E Ritter. “Evaluating a range of learning schedules: hybrid training schedules may be as good as or better than distributed practice for some tasks.” *Ergonomics* vol. 59,2 (2016): 276-90.

10.Liu, Feng et al. “Hybrid knowledge transfer for MARL based on action advising and experience sharing.” *Frontiers in neurorobotics* vol. 18 1364587. 7 May. 2024,

11.Lentz, Alison et al. “AI-ssessment: Towards Assessment As a Sociotechnical System for Learning.” *Academic medicine : journal of the Association of American Medical Colleges* vol. 96,7S (2021): S87-S88.

12.Daniel, Manju, and Rachel A Joseph. “Horizontal Integration of Concepts: An Innovative Teaching

Strategy." *Nurse educator* vol. 44,5 (2019): 284-287.
13.Arjanto, Paul, and Mint H Raya Aditama. "Integrating problem-based and team-based learning: A hybrid approach for advancing medical education pedagogy." *Journal of Taibah University Medical Sciences* vol. 20,1 107-108. 19 Feb. 2025,
14.Bezerra de Lima Danielle, Ana Carolina et al. "Ensino híbrido na formação em saúde: uma revisão sistemática" [Hybrid education in healthcare education : a systematic reviewEducación híbrida en la formación en salud: revisión sistemática]. *Revista Cuidarte* vol. 13,1 e8. 8 Aug. 2022,
15.Jessee, Mary Ann. "Pursuing Improvement in Clinical Reasoning: The Integrated Clinical Education Theory." *The Journal of nursing education* vol. 57,1 (2018): 7-13.
16.Yu, Rili et al. "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." *Frontiers in medicine* vol. 12 1684116. 9 Jan. 2026,

Figure 1 Technical Roadmap

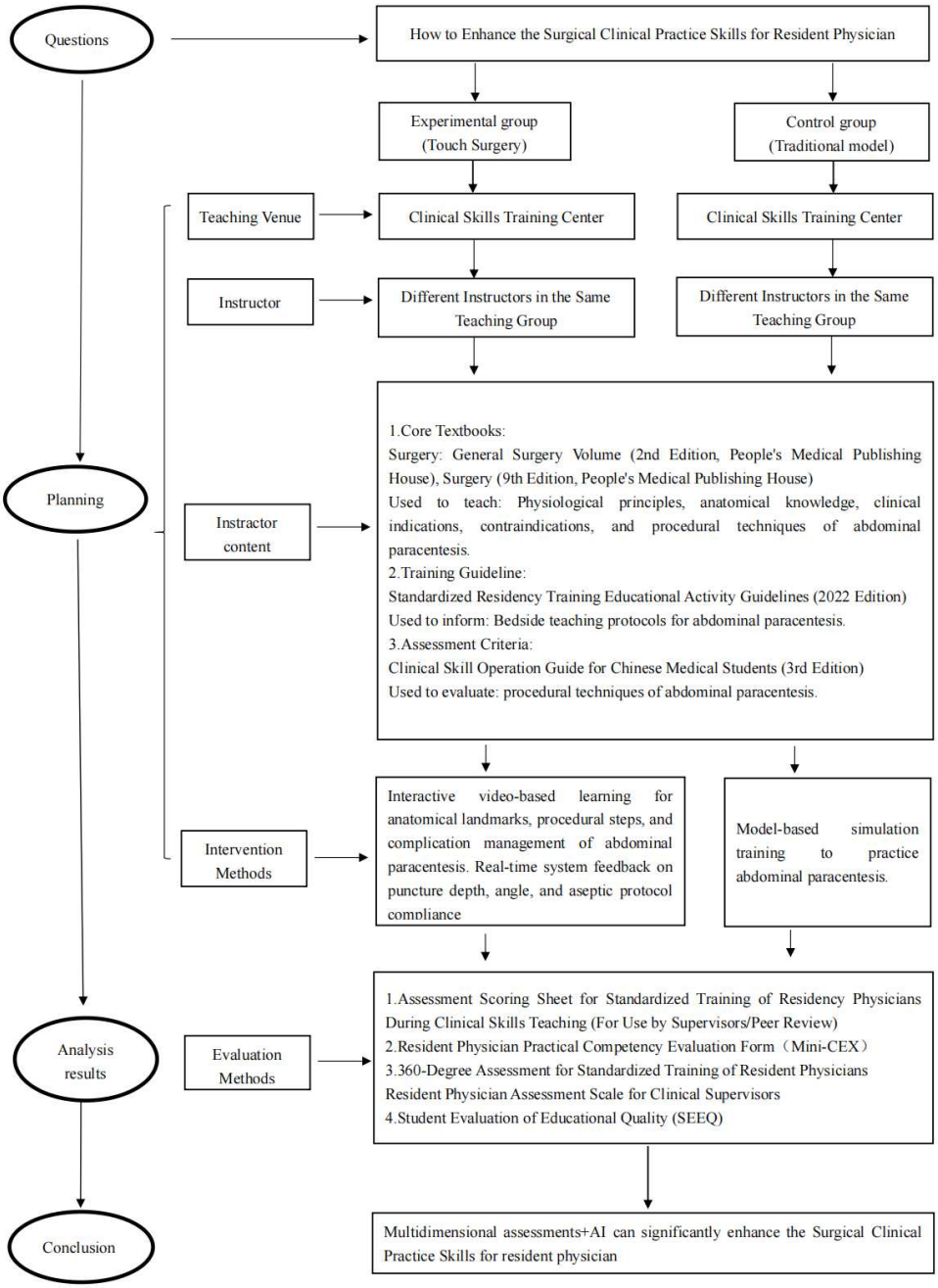


Figure 5 Distribution of technical skill scores before and after training in the experimental and control groups.

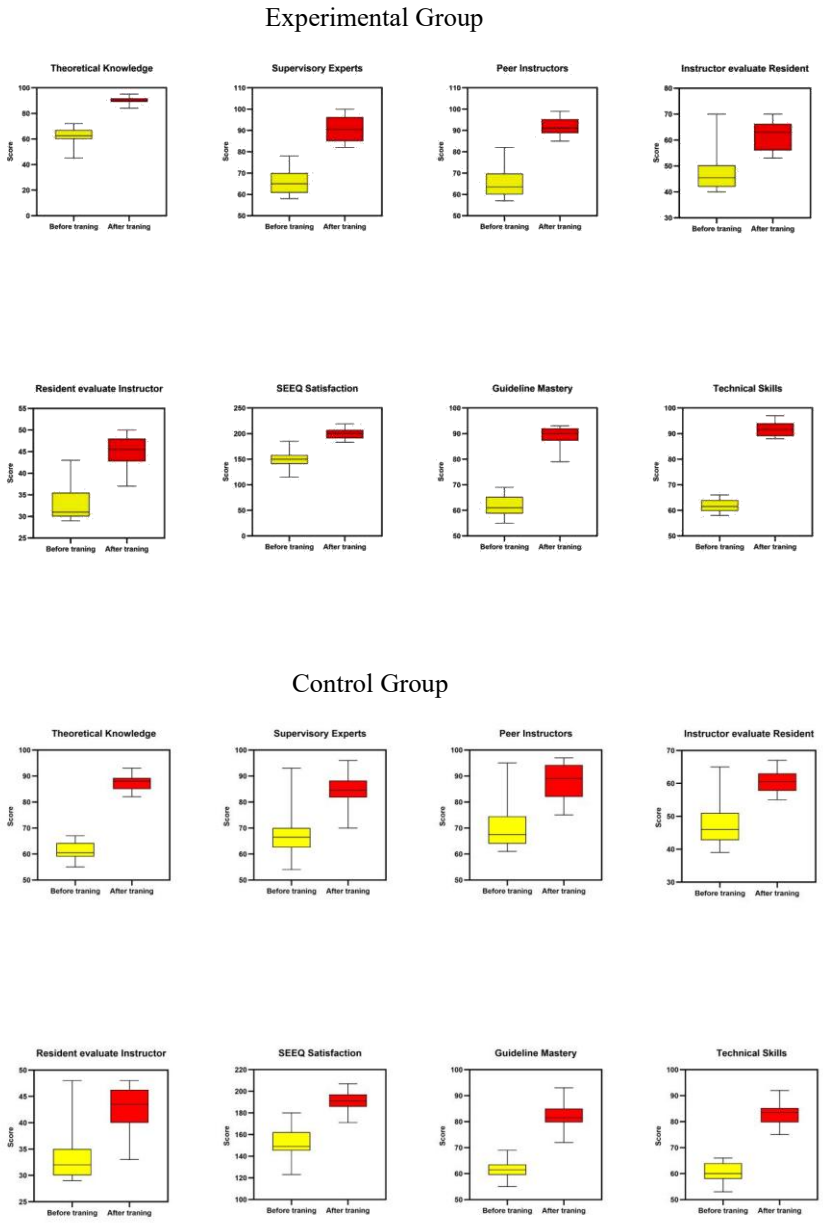


Figure 5 Distribution of technical skill scores before and after training in the experimental and control groups.Before training, both groups exhibited similar distributions predominantly concentrated in the 60–70 point range. After training, the experimental group (AI-VS + TMT) showed a compact distribution in the high-score range (above 90 points), whereas the control group (AI-VS only) displayed a broader distribution primarily clustered between 80–85 points. The interquartile ranges and medians are indicated by box boundaries and central lines, respectively.

Acknowledgments

Supervisory experts, peer educators, and mentors from the Department of Cardiothoracic and Vascular Surgery, the First Affiliated Hospital of University of South China

Clinical Skills Center, the First Affiliated Hospital of University of South China

Touch Surgery™ application provided by Medtronic, USA

Resident physicians who participated in this study

Author Contributions

Juan Wu (Corresponding Author): Conceptualization and design of the study, implementation and evaluation of teaching activities, manuscript revision, and final approval of the version to be published.
Haonan Zhang and Heng Yang (Co-first authors): Conducted the research, collected the data, and drafted the manuscript.
Zhengwen Lei, Jia Liu, and Juan Luo: Performed the project, conducted data analysis, and interpreted the results.
Jieyu Cao, and Kejiang Long: Translated the manuscript into English.

Funding Information

This work was supported by the General Funding Project of the Health Commission of Hunan Province (Grant No. 202214024221) and the Self-funded Project of the First Affiliated Hospital of University of South China (Grant No. 20240057, Project Title: Research on Psychological Intervention in Digitalization in Public Service Systems).

Competing Interests Statement

All authors declare that there are no competing interests related to any products or concepts mentioned in this article.

Ethical Approval and Informed Consent

This study was approved by the Ethics Committee of the First Affiliated Hospital of University of South China (Approval No.: 2023LL0927001).

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.