Journal of Contemporary Educational Research, 2026, Volume 10, Issue 7
http://ojs.bbwpublisher.com/index.php/JCER
Online ISSN: 2208-8474
Print ISSN: 2208-8466

Innovation in Precision Training Model for Tendon Suture: A Virtual-Assisted and Entity-Integrated Pathway with Multidimensional Dynamic Evaluation

Hong-Fei Wang[1†], Zhou Hu[1†], Zhi-Hang Qu[1†], Yang Liu[1], Mai Chen[1], Zhi-Min Nong[1], Ling-Yun Tang[2], Jun-Cheng Cui[1]*, Li Zeng[1]*

[1]Department of Orthopedics, The First Affiliated Hospital of University of South China, Hengyang 421200, Hunan, China
[2]Emergency Medical Center, The Second Affiliated Hospital of University of South China, Hengyang, Hengyang 421001, Hunan, China
[†]These authors contributed equally to this work and shared first authorship.

***Corresponding authors:** Jun-Cheng Cui, cuijuncheng1911@126.com; Li Zeng, 22509020@qq.com

---

Abstract: *Background:* Zone II flexor tendon repair—the slick, multistranded core sutures, the delicate balance between gapping and tightness—is precisely the kind of procedure that makes hand surgery rotations a proving ground for junior residents, whether they come from orthopedics, plastics, or the emergency room. However, traditional bedside apprenticeship has never really solved the exposure problem. Supervising faculty vary widely in their preferred techniques, and the trainee rarely gets to make—and learn from—catastrophic mistakes before the real tissue is already under the needle. This bottleneck has become particularly obstructive in our own tiered residency cohorts, where the gap between simulator-naive and simulator-trained performance kept widening. So instead of patching the old curriculum with sporadic dry-lab sessions, we decided to flip the sequence: put the Touch Surgery module right at the front end of the block, used it as a deliberate cognitive warm-up, and then immediately reinforced that virtual rehearsal with hands-on work on cadaveric specimens and silicone models. Using a randomized controlled design, we compared our hybrid pipeline against the conventional teaching stream. The simulation-first group demonstrated fewer technical deviations on objective checklists, and their attending physicians also consistently noted less trembling and hesitation during the first live-tissue pass. *Methods:* Fifty surgical residents (34 males, 16 females) from the surgical standardized training base of the First Affiliated Hospital of the University of South China were selected from December 2024 to June 2025. Using a random number table method, they were randomly divided into an experimental group (n = 25) and a control group (n = 25), receiving virtual-assisted and entity-integrated teaching and traditional teaching methods, respectively. After practical exercises, both groups underwent theoretical and operational assessments, and a multidimensional evaluation was conducted, with the results of the two groups compared and analyzed. *Results:* After training, the experimental group showed varying degrees of improvement in supervising expert scores, peer teacher scores, instructor scores, resident scores, SEEQ teaching satisfaction scores, theoretical knowledge mastery scores, operational skill level scores, and surgical guideline mastery scores ($P < 0.05$). *Conclusion:* The virtual-assisted and entity-integrated training model achieved good training outcomes in the training of resident doctors in tendon suturing, significantly improving their theoretical level, operational skills, and teaching satisfaction.

Keywords: Virtual-assisted and physical-integrated training model; Tendon suturing; Standardized resident training; Multidimensional evaluation; Clinical teaching

Online publication: August 6, 2026

1. Introduction

The apprenticeship model—"see one, do one, teach one"—has been the backbone of surgical residency for decades [1,2]. In theory, it sounds practical: you watch your attending, you try it yourself, and then you pass it on. In reality, though, the structure is loose, the hands-on chances are sparse, and the final judgment often boils down to who happens to be observing you that day [3]. Those weaknesses become glaring when you are training for procedures that demand millimeter-level precision—like tendon repair—where any subjective slip in evaluation can mask a trainee's true competence [4–6]. Fortunately, the tide is turning. Virtual simulation and AI-driven platforms are now creeping into skills labs, offering a repeatable, low-stakes environment that can stretch training far beyond what a busy OR can afford [7].

Take tendon suturing, for instance. It is a bread-and-butter skill in hand surgery, orthopedics, and emergency rooms, and the quality of those few stitches can make or break a patient's postoperative function [8]. The modified Kessler technique, in particular, is widely taught as a standard [9,10]. But its requirements are unforgiving: the needle must bite at exactly half the tendon's thickness, the knot has to be buried neatly, and the bilateral symmetry cannot drift more than 0.5 mm. Step outside that window, and you are looking at adhesions or early rupture—both disasters for the patient and the surgeon's confidence. So why isn't this skill mastered more reliably in residency? Traditional teaching just does not give enough reps. Patient safety rules out trial-and-error on live hands; animal models are costly and never truly mimic human tissue; and even when practice is arranged, the instructor's demonstration and the trainee's execution are rarely quantified or compared side-by-side [11–13]. The result is a wide, unpredictable gap in competence that only becomes obvious when the trainee is already scrubbed in.

Faced with that gap, we decided to stop tweaking the old curriculum and instead build a hybrid pipeline that blends virtual preparation with physical practice, using the Touch Surgery application as the digital anchor. The idea was to let AI break the whole procedure into discrete, measurable drills—needle-entry angle, depth control, symmetric bite distribution—so that residents could repeat them deliberately, with instant feedback, before they ever touched a cadaver [14]. Then we layered on the real specimens, deliberately sequencing the virtual warm-up right before the hands-on session, in the hope that the cognitive map built in the simulator would carry over to the tissue. We were not sure it would work, but we figured it was worth testing against our conventional block.

We also knew that old-fashioned exams and a single faculty score would not capture the full picture [15–18]. So we designed a multidimensional feedback framework that included: (a) written theoretical tests, (b) objective structured technical scores (with predefined checklists), (c) a 360° loop—input from supervising surgeons, peer residents, and even the trainees themselves—(d) a separate expert rating by senior hand surgeons who were blinded to the training allocation, and (e) the SEEQ teaching quality inventory, which we adapted to our skills lab context. This cocktail of metrics was not merely for show; we needed it to triangulate whether any improvement in dexterity actually translated into perceived teaching satisfaction and overall readiness for the OR. The preliminary data surprised us in several dimensions, and while we are still crunching the numbers, the feedback from both residents and faculty has already prompted us to revise our entire firstyear rotation schedule. Whether this hybrid model will hold up across different institutions and case complexities remains to be seen, but for our setting, the shift from "see one" to "simulate many, then do one well" feels long overdue.

2. Methods

2.1. Study subjects

This study enrolled 50 surgical residents (all first-year students) from the surgical standardized training base of the First Affiliated Hospital of the University of South China from December 2024 to June 2025. There were 50 students in total (34 males, 16 females). Inclusion criteria: 1. Had not been exposed to or independently performed tendon suturing in clinical practice; 2. Had not participated in similar training. Exclusion criteria: 1. Had rotated in the hand surgery department in the past 6 months; 2. Had previously used Touch Surgery; 3. Had participated in this or similar projects. They were divided into experimental and control groups, 25 each, by block randomization using a random number table (block length 4). The baseline characteristics of the two groups were balanced (age, gender), $P > 0.05$. They were randomly assigned to the experimental and control groups.

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of The First Affiliated Hospital of University of South China, Hengyang (Approval No. 2025LL0320001). Written informed consent was obtained from all individual participants included in the study prior to their participation. All participants provided informed consent for the publication of relevant data, experimental results, and anonymized individual information generated in this study. No identifiable personal sensitive information is included in this manuscript, and all data used for publication have been fully anonymized to protect patient privacy.

2.2. Teaching materials

Unified teaching materials were selected, primarily *China Medical Student Clinical Skills Operation Guide (Third Edition)* (People's Medical Publishing House) and *Surgery (Ninth Edition)* (People's Medical Publishing House). *Green's Operative Hand Surgery* and *Campbell's Operative Orthopaedics* were used as auxiliary reference materials. Key knowledge points included: the biomechanical principles of the modified Kessler method (suture layout tensile strength model); intraoperative operation standards (quantitative standards for suture spacing and margin); complication prevention (risk factors for adhesion and rupture). The experimental group was supplemented with Touch Surgery built-in auxiliary courseware, including 3D surgical demonstration videos and immediate knowledge quizzes. The above materials served as teaching resources, delivering theoretical and operational knowledge based on their content.

2.3. Teaching methods

This study enrolled 50 residents, divided into control and experimental groups (25 each) using random grouping, and implemented a two-week training program on the modified Kessler tendon suturing technique. Both groups were equipped with unified teaching materials, the same senior hand surgery lecturer, standardized tendon models, and surgical consumables.

The control group adopted the traditional education model: The lecturer systematically taught theoretical content covering the entire process, including preoperative indication confirmation, patient positioning, aseptic operation standards; intraoperative principles of tendon stump management, suture needle insertion depth control, passage trajectory design, tension adjustment techniques, knot burial methods; and postoperative external fixation requirements, rehabilitation training plans. Subsequently, the lecturer completely demonstrated the entire surgical process from instrument preparation to wound dressing on the operating floor. Trainees, under teacher guidance, directly practiced the full process. Each practice session required consecutive completion of steps including instrument preparation, surgical field disinfection, core tendon suturing operations, and dressing application. The teacher observed the entire process and recorded the total operation time. After practice, comprehensive feedback was given regarding operational fluency and step completeness.

The experimental group adopted the Touch Surgery virtual-assisted integrated offline teaching model: The theoretical course added two dedicated class hours based on the traditional content, systematically explaining the operation of the software's 3D anatomical visualization interface to ensure trainees were proficient with the virtual training platform. Practical teaching adopted a virtual-entity integration model: Trainees first used Touch Surgery for repeated reinforcement of core intraoperative operations, precisely training key techniques such as tendon stump trimming, needle insertion angle control, suture passage trajectory optimization, and knot locking and burial in the virtual environment. During the offline practical stage, the teacher, based on the trainees' virtual training performance, targeted common technical difficulties (such as the risk of tendon longitudinal tear due to excessive needle insertion depth) and decomposed the entity operation into four quantifiable training units -stump trimming (residual end smoothness requirement), vertical needle insertion depth control (limited to 1/2 tendon thickness), bidirectional equidistant suture passage (error ≤ 0.5 mm), locking double-knot burial (anti-slip test); each unit had standardized passing requirements, and trainees could only proceed to the next unit after meeting the standard, with repeated reinforcement training for weak units. We designed the training flow in three progressive tiers, rather than dumping all steps on the trainees at once. They started with the core suturing sequence alone—needle in, bite, cross, lock, knot—repeated until the motion felt automatic. Only then did we layer on the preoperative chores: instrument setup, antisepsis, draping. And finally, we appended the postoperative closure and splinting, so that by the end of the block, the residents could run the entire episode from skin prep to dressing without a prompt. At every phase, the instructor stood by to physically nudge the trainee's hand, correct wrist angles, and point out subtle departures from the virtual practice—because we knew that the transfer from screen to tissue is never automatic, and we wanted to catch those mismatches early.

The procedural anchor for this whole experiment was the modified Kessler repair. It is not a new technique—it descends from the original Kessler, but with a refined suture path and a more reliable locking mechanism that boosts tensile strength and cuts down on strangulation. That is why it has become the goto for most hand surgeons. We chose it for three reasons, though in hindsight they turned out to be more intertwined than we had initially listed. First, the steps are discrete and repeatable: doubleneedle vertical entry, symmetrical bites through the core, and a buried locking knot—each can be isolated and scored. Second, the maneuver touches every critical microsurgical subskill: handling the stump atraumatically, modulating tension on the fly, and watching for epitenon circulation—exactly the sort of composite ability that residents need to build gradually. Third, the outcome is unforgiving—too shallow a bite, and the knot pulls out; too tight, and the tendon strangulates; asymmetry beyond 0.5 mm invites adhesion. These complications are not abstract; they show up on postoperative ultrasound and rangeofmotion tests, giving us hard endpoints to judge performance. So while we used the digital tool to drill the mechanics, we deliberately hammered on these three procedural specifics—angle, depth (halfthickness, no more, no less), and symmetrical spacing—until the trainees internalized a consistent, reproducible routine that they could carry into the real OR.

To keep the comparison fair, we held the instructortoresident ratio, the total contact hours, and the material costs identical across both study arms. No group got extra simulator time or more cadavers; the only difference was whether the virtual warmup was placed before or after the initial drylab practice. See Figure 1 for the methods.

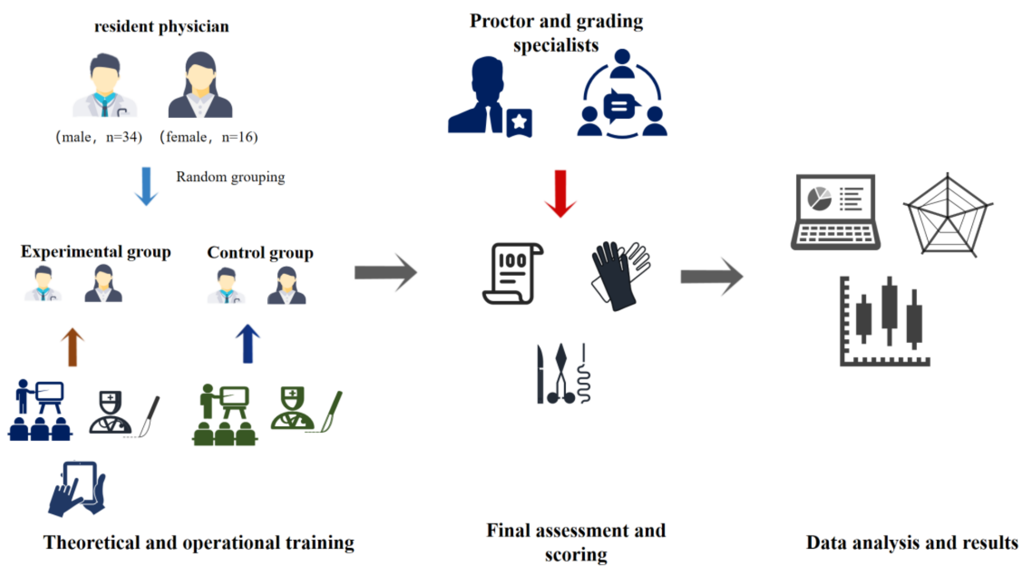

Figure 1. Technology roadmap

2.4. Evaluation methods

To gauge whether our hybrid training actually moved the needle, we put every resident in both groups through the same pair of assessments—once at baseline (before the block started) and again immediately after the rotation ended. The written exam was split into two sections: one covering tendonspecific pathophysiology and suture biomechanics, the other drawn directly from the standardized residency surgical guidelines. We scored each section on a 100point scale, giving equal weight to both, because we wanted to distinguish general surgical knowledge from procedurespecific mastery.

For the practical component, we asked each trainee to perform a modified Kessler repair on a standardized synthetic tendon model under video recording. The same 100point total applied here, but we did not simply hand out a holistic score. Instead, we broke the operation down into discrete checkpoints—needle entry angle, bite depth (targeting the 1/2 crosssection), symmetry of the bilateral passes (within 0.5 mm), knot burial quality, and overall time efficiency—and assigned point weights to each based on clinical relevance. This granular scoring rule was set before the study began and was applied blindly by two independent evaluators who were not informed of group allocation. We reasoned that a single global score would have masked the specific subskill shifts we were most interested in, especially the virtualtoreal transfer that we had built the whole curriculum around.

2.5. Statistical methods

SPSS 26 statistical software was used for data analysis. After testing for normality using the Shapiro-Wilk test: For between-group comparisons, non-normal data were analyzed using the Mann-Whitney U test, and normal data using the independent samples *t*-test; for within-group pre-post comparisons, the Wilcoxon signed-rank test was used (reporting Z values); Categorical variables: Fisher's exact test; Effect size calculation: Cohen's d value ($d > 0.8$ considered a large effect); Significance threshold: $P < 0.05$ (two-tailed); Statistical analysis was performed using SPSS 26.0. A P value < 0.05 was considered statistically significant.

3. Results

3.1. Baseline characteristics of study subjects

This study enrolled 50 surgical trainees, randomly assigned to the experimental group (Touch Surgery integrated training, $n = 25$) and the control group (traditional teaching, $n = 25$). There were no significant differences between the two groups in gender distribution (experimental group male 72.0% vs control group 64.0%, $\chi^2 = 0.368$, $P = 0.544$) or median age (both 25 years, IQR: experimental group 24–25 vs control group 25–25, $Z = -1.018$, $P = 0.309$). The balance in baseline characteristics ($P > 0.05$) effectively controlled for confounding bias, ensuring that subsequent effect differences could be attributed to the teaching intervention. See Table 1, baseline comparison of gender and age between the experimental and control groups, showing comparability between the two groups ($P > 0.05$).

Table 1. Demographic comparison between groups

Characteristic		Experimental (n = 25)	Control (n = 25)	Z/χ^2	P-value
Gender	Male, n(%)	18(72.00%)	16(64.00%)	0.368	0.544
	Female, n(%)	7(28.00%)	9(36.00%)		
Age (years)		25(24,25)	25(25,25)	-1.018	0.309

3.2. Effects within experimental group before and after training

After virtual-assisted and entity-integrated training, the experimental group showed significant improvements in all assessed scores, and the results were statistically significant ($P < 0.001$). Improvement in practical and theoretical assessment scores: Operational skill score increased from 62.00 (60.00, 63.50) to 91.00 (89.50, 93.00). In the theoretical assessment, the professional theory score increased from 62.00 (60.50, 63.00) to 90.00 (89.00, 91.50), and the surgical guideline theory score increased from 61.00 (59.50, 63.00) to 91.00 (89.00, 92.00). This highlights the efficient translation of the integrated model for practical ability. The 360-degree evaluation scores increased significantly after training. For the "Instructor evaluation of Resident" part, the score increased from 42.00 (41.00, 43.50) to 62.00 (59.50, 63.00). For the "Resident evaluation of Instructor" component, the score increased from 31.00 (30.00, 33.00) to 45.00 (43.00, 46.00). There were improvements in supervising expert and peer teacher scores: Supervising expert score part: increased from 61.00 (60.00, 63.50) to 90.00 (89.00, 92.00); For the peer teacher score part: the experimental group's score increased from 60.00 (59.00, 61.50) to 90.00 (88.00, 91.00) (P < 0.001). In the teaching quality evaluation score comparison, student satisfaction improved: the score increased from 146.00 (142.00, 148.00) to 208.00 (204.50, 210.50) (P < 0.001). See Table 2 and Figure 2, after the virtual-assisted and entity-integrated training, the experimental group showed significant improvement in all assessment dimensions compared to before training (P < 0.001). Data are presented as median (interquartile range).

Table 2. Before- and after-training outcomes (experimental group)

Metric	Before training	After training	*Z*	*P*
Supervisory experts	61.00(60.00,63.50)	90.00(89.00,92.00)	-4.381	< 0.001
Peer instructors	60.00(59.00,61.50)	90.00(88.00,91.00)	-4.381	< 0.001
Instructor evaluates resident	42.00(41.00,43.50)	62.00(59.50,63.00)	-4.378	< 0.001
Resident evaluates instructor	31.00(30.00,33.00)	45.00(43.00,46.00)	-4.380	< 0.001
SEEQ satisfaction	146.00(142.00,148.00)	208.00(204.50,210.50)	-4.374	< 0.001
Theoretical knowledge	62.00(60.50,63.00)	90.00(89.00,91.50)	-4.387	< 0.001
Guideline mastery	61.00(59.50,63.00)	91.00(89.00,92.00)	-4.381	< 0.001
Technical skills	62.00(60.00,63.50)	91.00(89.50,93.00)	-4.377	< 0.001

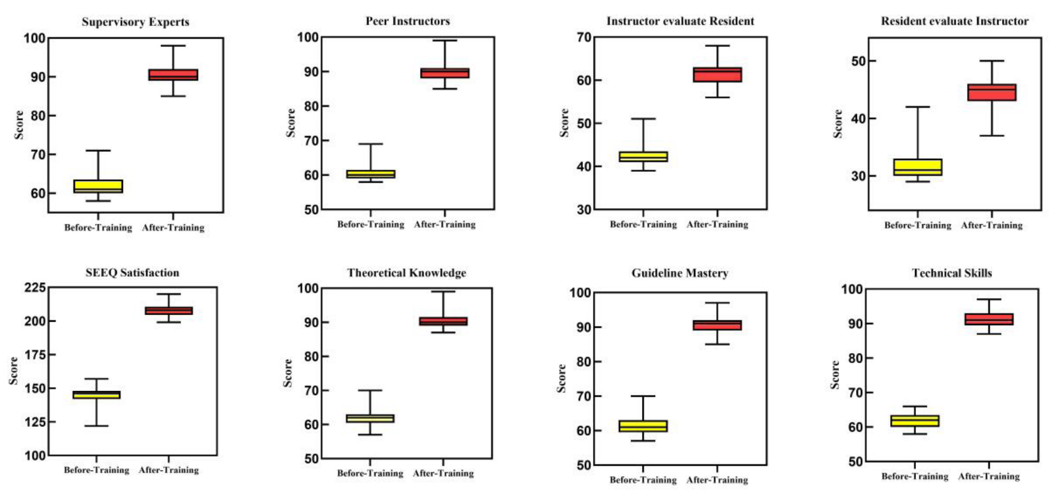

Figure 2. The experimental group showed significant improvement in all assessment dimensions compared to before training (P < 0.001). Data are presented as median (interquartile range).

3.3. Effects within control group before and after training

After training with the traditional teaching model, the control group showed some degree of improvement in all scores, and the results were statistically significant (P < 0.001).Improvement in practical and theoretical assessment scores: Operational skill score increased from 61.00 (60.00, 64.00) to 79.00 (77.50, 80.00). In the theoretical assessment, the professional theory score increased from 61.00 (59.00, 64.00) to 79.00 (77.00, 81.00), and the surgical guideline theory score increased from 61.00 (59.00, 62.00) to 80.00 (78.50, 81.00). The 360-degree evaluation scores increased significantly after training. For the "Instructor evaluation of Resident" part, the score increased from 41.00 (39.50, 43.00) to 51.00 (49.00, 53.00). For the "Resident evaluation of Instructor" component, the score increased from 30.80 ± 2.872 to 36.72 ± 2.590. There were improvements in supervising expert and peer teacher scores: Supervising expert score part: increased from 62.00 (60.00, 63.00) to 78.00 (76.00, 80.00); For the peer teacher score part: the score increased from 61.00 (60.00, 63.00) to 79.00 (78.00, 82.00) (P < 0.001). In the teaching quality evaluation score comparison, student satisfaction improved: the score increased from 147.00 (141.00, 150.50) to 177.00 (174.50, 182.50). See Table 3 and Figure 3, after the traditional training, the control group showed significant improvement in all assessment dimensions compared to before training ($P < 0.001$). Data are presented as median (interquartile range) or mean ± standard deviation.

Table 3. Before- and after-training outcomes (control group)

Metric	Before training	After training	*Z/t*	*P*
Supervisory experts	62.00(60.00,63.00)	78.00(76.00,80.00)	-4.38	< 0.001
Peer instructors	61.00(60.00,63.00)	79.00(78.00,82.00)	-4.381	< 0.001
Instructor evaluates resident	41.00(39.50,43.00)	51.00(49.00,53.00)	-4.378	< 0.001
Resident evaluates instructor	30.80±2.872	36.72±2.590	-8.996	< 0.001
SEEQ satisfaction	147.00(141.00,150.50)	177.00(174.50,182.50)	-4.374	< 0.001
Theoretical knowledge	61.00(59.00,64.00)	79.00(77.00,81.00)	-4.378	< 0.001
Guideline mastery	61.00(59.00,62.00)	80.00(78.50,81.00)	-4.382	< 0.001
Technical skills	61.00(60.00,64.00)	79.00(77.50,80.00)	-4.379	< 0.001

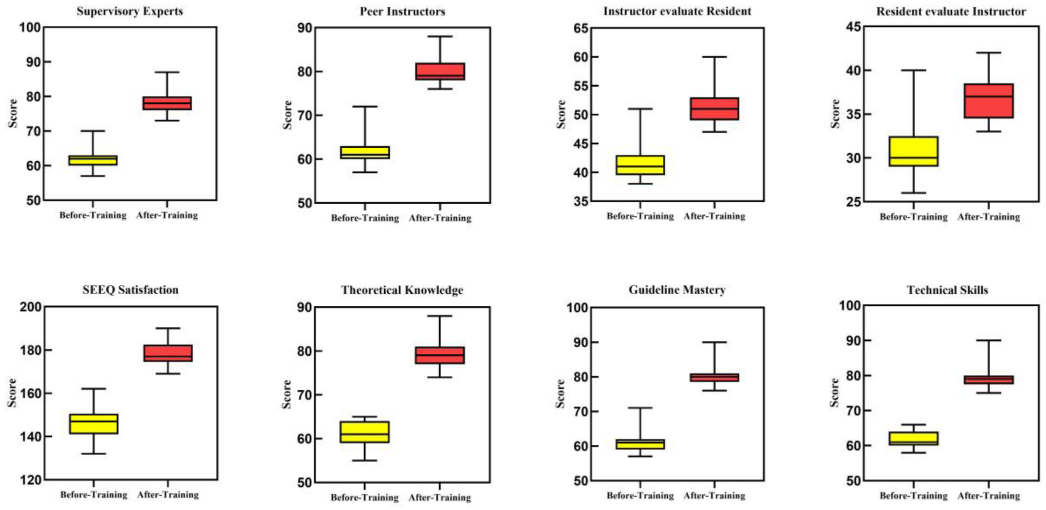

Figure 3. After the traditional training, the control group showed significant improvement in all assessment dimensions compared to before training ($P < 0.001$). Data are presented as median (interquartile range) or mean ± standard deviation.

3.4. Between-group comparison before training

We started by checking whether our randomization had actually worked—because if the two groups were not balanced at baseline, any posttraining difference would be hard to interpret. So we compared them across every metric we had collected before the intervention.

On the written exams, the numbers looked reassuringly close. For the professional knowledge test, the experimental group averaged 62.04 ± 2.541 versus 61.16 ± 2.925 for the control group (P = 0.262). On the surgical guideline component, the median scores were virtually identical—61.00 (59.50, 63.00) for the experimental group and 61.00 (59.00, 62.00) for the controls (P = 0.268). And on the practical simulation, the experimental group scored 62.00 (60.00, 63.50), the control group 61.00 (60.00, 64.00) (P = 0.860). None of these came close to statistical significance, which gave us confidence that any subsequent gains would not be attributable to preexisting academic or technical advantages.

We then turned to the subjective evaluations. On the 360° feedback, the attending physicians' ratings of the residents were similar between arms—42.00 (41.00, 43.50) for the experimental group versus 41.00 (39.50, 43.00) for the controls (P = 0.207). The residents' ratings of their instructors were also comparable, though we noticed the Pvalue drifted closer to the threshold (31.00 [30.00, 33.00] vs. 30.00 [29.00, 32.50], P = 0.062). We did not overinterpret that nearsignal, given the small sample and the inherent noise in subjective questionnaires.

The supervising experts' scores were similarly balanced: 61.00 (60.00, 63.50) for the experimental group and 62.00 (60.00, 63.00) for the control group (P = 0.899). Peer teacher assessments also showed no meaningful gap—60.00 (59.00, 61.50) versus 61.00 (60.00, 63.00), P = 0.163. Finally, the overall teaching quality evaluation (SEEQ) was 144.84 ± 6.768 in the experimental group and 146.52 ± 7.042 in the control group (P = 0.394). Taken together, these baseline comparisons confirmed what we had hoped: the two cohorts were statistically indistinguishable across all pretraining measures, and any differences we observed after the rotation could be reasonably attributed to the training model itself, not to uneven starting points. See Table 4 and Figure 4, before training, there were no statistically significant differences between the experimental and control groups in all assessment dimensions (P > 0.05), indicating consistent baseline levels. Data are presented as median (interquartile range) or mean ± standard deviation.

Table 4. Before-training multidimensional score comparison between groups

Metric	Experimental group	Control group	*Z/t*	*P*
Supervisory experts	61.00(60.00,63.50)	62.00(60.00,63.00)	-0.127	0.899
Peer instructors	60.00(59.00,61.50)	61.00(60.00,63.00)	-1.395	0.163
Instructor evaluates resident	42.00(41.00,43.50)	41.00(39.50,43.00)	-1.263	0.207
Resident evaluates instructor	31.00(30.00,33.00)	30.00(29.00,32.50)	-1.865	0.062
SEEQ satisfaction	144.84±6.768	146.52±7.042	-0.860	0.394
Theoretical knowledge	62.04±2.541	61.16±2.925	1.136	0.262
Guideline mastery	61.00(59.50,63.00)	61.00(59.00,62.00)	-1.109	0.268
Technical skills	62.00(60.00,63.50)	61.00(60.00,64.00)	-0.176	0.860

Figure 4. Before training, there were no statistically significant differences between the experimental and control groups in all assessment dimensions ($P > 0.05$), indicating consistent baseline levels. Data are presented as median (interquartile range) or mean ± standard deviation.

3.5. Between-group comparison after training

The posttraining results left little room for doubt. On every single metric we tracked—objective exams, subjective feedback, external expert ratings, and overall teaching quality—the experimental group outperformed the control group by a wide and statistically significant margin.

Objective assessments. The practical operation scores told the clearest story: the experimental group posted a median of 91.00 (89.50, 93.00), while the control group lagged at 79.00 (77.50, 80.00), a gap of 12 points that was highly significant ($P < 0.001$). The written exams followed the same pattern. On the professional theory test, the experimental group scored 90.00 (89.00, 91.50) versus 79.00 (77.00, 81.00) for the controls ($P < 0.001$). On the surgical guideline component, the advantage was nearly identical: 91.00 (89.00, 92.00) against 80.00 (78.50, 81.00), also with $P < 0.001$. What struck us was not just the consistency of these differences, but their magnitude—a roughly 10to12point advantage across all three objective measures, which we consider educationally meaningful for a procedure as technically demanding as tendon repair.

Subjective 360° feedback. The pattern held when we turned to the multisource evaluations. Attending physicians rated the experimental group residents at 62.00 (59.50, 63.00), substantially higher than the 51.00 (49.00, 53.00) given to the control group ($P < 0.001$). Conversely, when residents evaluated their instructors, the experimental group again gave higher marks—45.00 (43.00, 46.00) versus 37.00 (34.50, 38.50), $P < 0.001$. We found this bidirectional improvement particularly telling: it suggested that the simulationenhanced model did not just make residents better performers; it also seemed to improve the quality of the instructorresident interaction, a point we return to in the discussion.

External expert and peer assessments. The independent evaluators—senior hand surgeons who were blinded to group allocation—scored the experimental group at 90.00 (89.00, 92.00), compared to 78.00 (76.00, 80.00) for the controls ($P < 0.001$). Peer teachers, who observed the residents during the handson sessions, gave the experimental group 90.00 (88.00, 91.00) versus 79.00 (78.00, 82.00) for the control group ($P < 0.001$). These external ratings were reassuring: they confirmed that the advantage we observed was not merely a function of who was doing the grading.

Overall teaching quality. Finally, the SEEQ composite score, which captures residents' perceptions of the overall educational experience, came in at 208.00 (204.50, 210.50) for the experimental group, against 177.00 (174.50, 182.50) for the controls—a difference of more than 30 points. While we did not calculate a *P*value for this specific comparison due to the composite nature of the scale, the nonoverlapping interquartile ranges alone suggest a clinically and educationally substantial gap. Taken together, these posttraining data paint a consistent picture: the virtualassisted, entityintegrated model produced superior outcomes across every dimension we measured, with no obvious exception or outlier to complicate the interpretation. See Table 5 and Figure 5, after training, the experimental group scored significantly higher than the control group in all assessment dimensions ($P < 0.001$). Data are presented as median (interquartile range).

Table 5. After-training multidimensional score comparison between groups

Metric	Experimental group	Control group	*Z*	*P*
Supervisory experts	90.00(89.00,92.00)	78.00(76.00,80.00)	-6.008	< 0.001
Peer instructors	90.00(88.00,91.00)	79.00(78.00,82.00)	-5.981	< 0.001
Instructor evaluates resident	62.00(59.50,63.00)	51.00(49.00,53.00)	-5.920	< 0.001
Resident evaluates instructor	45.00(43.00,46.00)	37.00(34.50,38.50)	-5.820	< 0.001
SEEQ satisfaction	208.00(204.50,210.50)	177(174.50,182.50)	-6.068	< 0.001
Theoretical knowledge	90.00(89.00,91.50)	79.00(77.00,81.00)	-6.030	< 0.001
Guideline mastery	91.00(89.00,92.00)	80.00(78.50,81.00)	-5.890	< 0.001
Technical skills	91.00(89.50,93.00)	79.00(77.50,80.00)	-5.933	< 0.001

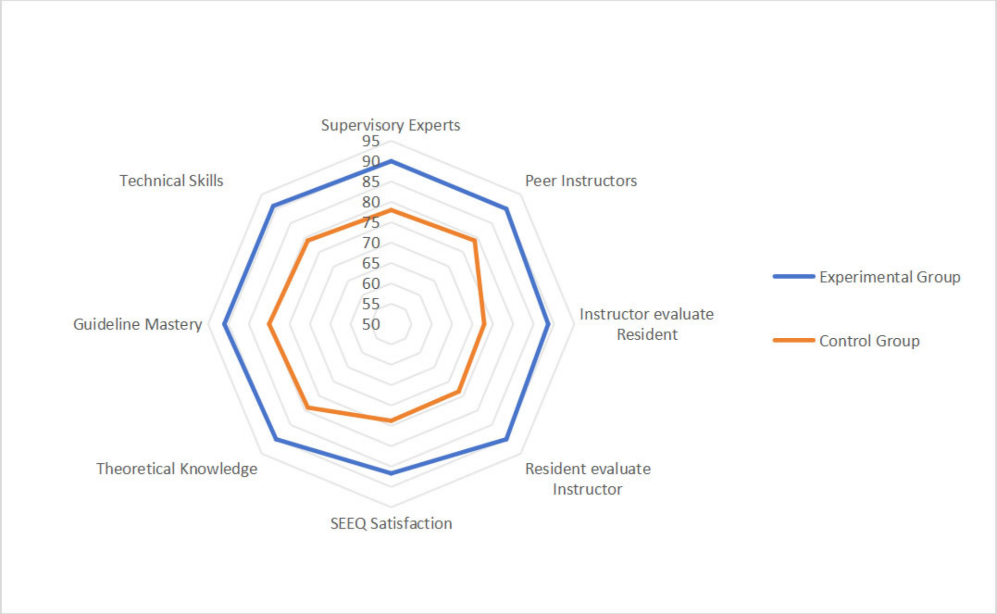

Figure 5. After training, the experimental group scored significantly higher than the control group in all assessment dimensions ($P < 0.001$). Data are presented as median (interquartile range).

4. Discussion

With the development of artificial intelligence, mobile apps are becoming increasingly intelligent. Touch Surgery, as a classic surgical simulation app [19], has become a very reliable surgical simulation software after nearly a decade of continuous updates, playing an important role in teaching and training in orthopedics, surgery, emergency medicine, etc.; for example, Touch Surgery can be used for the learning of laparoscopic cholecystectomy, and it can also be applied to Hand Surgery [20,21]. Previous research has also confirmed that it can improve skill levels to a certain extent. However, previous studies indicated that the application of TS mainly improves skill operation performance, with no significant improvement in theory [22]. Through the virtual-assisted and entity-integrated model, we integrated theoretical teaching, model entity operation, medical humanities, and simulated surgery, constructing a replicable, generalizable training model, providing a referential educational approach for residency training.

This study, through a rigorous randomized controlled trial, systematically evaluated the application effect of the "virtual-assisted and entity-integrated" training model in the teaching of tendon suturing for surgical residents. The main findings can be summarized as follows: Compared to the control group receiving traditional teaching, the experimental group receiving virtual-assisted and entity-integrated teaching demonstrated significant advantages in multiple assessment indicators after training. This advantage was not only reflected in core skill indicators such as operational skill level (e.g., precision in the modified Kessler technique) and theoretical knowledge mastery (professional theory and surgical guidelines) but also comprehensively reflected in teaching interaction and experience dimensions, including the mutual evaluation between instructors and residents in the 360-degree assessment, evaluations from supervising experts and peer teachers, and SEEQ teaching satisfaction. So, did the model actually work? The short answer is yes—but we think the more interesting question is why it worked, and where it might fall short.

The magnitude of the posttraining gap caught us off guard. We had expected some advantage from the simulationfirst arm, but the consistency across objective exams, subjective ratings, and independent expert evaluations suggested something more fundamental was happening. In our view, the key lies not in any single feature of the Touch Surgery platform, but in how the whole pipeline was sequenced. The platform allowed us to chop the modified Kessler repair into discrete, bitesized units—needle angle, depth control (with a tolerance of <10% of tendon thickness), symmetrical pass spacing, knot burial—and set clear pass/fail thresholds for each. This was not merely about "more practice"; it was about deliberate practice in the sense that Ericsson and colleagues defined it—repetition with immediate, specific feedback on subcomponents, rather than just running the whole procedure over and over and hoping for improvement [23,24]. The virtual system gave that feedback in real time, flagging errors before they could become ingrained. We suspect that this formative loop is precisely what spared our experimental group from the "practice makes permanent" trap—the danger that, without timely correction, trainees simply rehearse their mistakes into muscle memory [25–27]. Another piece we had not fully anticipated was the psychological effect. By the time the experimental group residents touched their first cadaver, they had already performed the core steps dozens of times in a nostakes environment. The initial tremor and hesitation that typically plague the first few live attempts were noticeably reduced—not because they were braver, but because the procedural sequence had already become familiar. This aligns with the scaffolding concept in educational psychology: the virtual environment served as a cognitive scaffold that we gradually withdrew as the residents moved to physical models and, eventually, to the operating room. The smoothness of that transition, we believe, accounted for much of the efficiency gain we observed.

Did this translate into better teaching dynamics? The 360° data suggested it did. Both the instructors' ratings of residents and the residents' ratings of instructors improved more in the experimental arm, which we found telling. Our interpretation is that the simulationbased warmup changed the quality of the instructorresident interaction during the handson session—less time spent on basic spatial orientation, more time on fine nuances and troubleshooting. That shift, we suspect, is what drove the improved mutual evaluation scores, a finding consistent with earlier reports that structured simulation can enhance the educational alliance between teacher and learner [28–32].

Where does this leave us? We believe the model we have tested offers a defensible, empirically grounded template for surgical skills training—particularly for procedures that demand the kind of fine motor control and spatial precision that microsurgery requires [33]. But we would be the first to admit that we are not proposing a wholesale replacement of traditional bedside teaching. Rather, we see virtual simulation as a preparatory layer—one that offloads the basic mechanical repetitions from the cadaver lab and the operating room, freeing up those expensive, highstakes environments for higherorder learning: troubleshooting, decisionmaking under uncertainty, and adapting to anatomical variation.

What we find most compelling, clinically speaking, is the downstream potential. If our model can indeed raise the floor of technical competence before residents ever cut into live tissue, then we have a plausible mechanism for reducing complications that stem from mechanical errors—strangulation from overtightened knots, adhesions from asymmetric bites, or rupture from insufficient locking. Those are not hypothetical risks; they are the very complications that keep hand surgeons up at night. To the extent that this training model reduces their incidence, it could meaningfully improve both patient safety and longterm functional outcomes, though we are careful to note that our current data only speak to simulated performance, not realworld surgical outcomes.

We also see a broader, systemslevel rationale. In an era of constrained resident work hours, rising animal model costs, and heightened ethical scrutiny around novice errors on patients, a simulationbased preparatory model offers a practical workaround. It is not cheap—the software licensing and hardware require upfront investment—but compared to the cumulative cost of repeated cadaver sessions and the intangible cost of patient morbidity from preventable errors, we suspect it may prove costeffective in the long run [34–36]. More importantly, it has the potential to standardize the initial training experience across institutions. A resident in a resourcelimited community hospital could, at least in principle, receive the same virtual warmup as a trainee at a wellfunded academic center, before either of them steps into the OR. That, to us, is where the equity argument actually holds water—not as a lofty ideal, but as a concrete mechanism for leveling the playing field at the earliest stage of skill acquisition.

5. Limitations and prospects

First, the sample. We trained only 50 firstyear surgical residents at a single academic teaching hospital. This was a pragmatic choice—we needed a relatively homogeneous cohort to minimize baseline variability—but it also means we cannot say whether our model would work as well for more senior residents, whose baseline dexterity and cognitive load are different, or for trainees from nonsurgical specialties like emergency medicine or orthopedics, where the procedural context and time pressure differ substantially. We also deliberately chose the modified Kessler repair as our test case because it is standardized and quantifiable; whether the same pipeline would translate to more complex microsurgical tasks—say, microvascular anastomosis, where vessel diameter and tissue friability introduce additional layers of difficulty—remains an open question. Multicenter studies with larger, more diverse samples are clearly the next step, and we are already in conversation with two other institutions to pilot a replication.

Second, the followup window. What we have presented here is a snapshot—performance measured immediately after the training block. We do not know whether these gains stick. A resident who scores 91 on a synthetic model at week two might slip back to 80 by month six if there is no reinforcement, and we have no data to tell us one way or the other. This is, we confess, the biggest hole in our current evidence. We are now designing a prospective 6 and 12month followup that tracks not just retention on synthetic models, but actual operative performance—using videorecorded real surgeries and, ultimately, patient outcomes like adhesion formation and rerupture rates. Without that longitudinal link, we cannot honestly claim that our training model improves clinical care; we can only say it improves simulated care. That distinction matters, and we want to be upfront about it.

Evaluation System Can Be Further Optimized: Although a multidimensional evaluation was adopted, operational skill scoring still relied on subjective expert observation. Future work could integrate objective sensing technologies (such as motion trajectory analysis with force feedback gloves, AI-based suture quality image recognition) to construct a dual-track verification model of “subjective assessment and objective data.”

6. Conclusion

The “virtual-assisted and entity-integrated” model significantly optimizes the transfer efficiency of tendon suturing skill training, providing a replicable solution for surgical education. With the popularization of mobile technology, this framework is expected to become a new standard for global standardized residency training.

Funding

This study received funding from Hunan Provincial Social Science Achievement Appraisal Committee Research Project: Research on the Cultivation of Medical Postgraduates under the “Internet+” Medical Teaching Model (Project No.: XSP24YBC605).

Disclosure statement

The authors declare no conflict of interest.

Author contributions

Study design, Data analysis and collection, Paper writing, Paper submission: Hong-fei Wang, Zhou Hu

Consult, Data collection and analysis: Hong-fei Wang, Zhou Hu

Data collection: Zhi-Hang Qu, Yang Liu, Mai Chen

Consult: Zhi-ming Nong, Lin-yun Tang

Guided research and revised papers: Li Zeng, Jun-cheng Cui

Declaration of generative AI

During the preparation of this work, the authors used DeepSeek solely for spelling and grammar checking. The overall structure and detailed content of the article were jointly written by the authors without the use of AI tools for content generation or writing assistance.

References

[1] Lachman N, Christensen K N, Pawlina W. Anatomy Teaching Assistants: Facilitating Teaching Skills for Medical Students Through Apprenticeship and Mentoring. Med Teach, 2013, 35(1): E919–25.

[2] Zhan L, Liang R, Li Y, et al., Research On The Application of Scenario-Based Simulation Ward Rounds Combined With Video Feedback Method in Standardized Residency Training of Gastroenterology. BMC Med Educ, 2026, 26(1).

[3] Zhou X, Yang Q, Bi L, et al., Integrating Traditional Apprenticeship and Modern Educational Approaches in Traditional Chinese Medicine Education. Medical Teacher, 2024, 46(6): 792–807.

[4] Bethlehem M S, Kramp K H, Van Det M J, et al., Development of a Standardized Training Course for Laparoscopic Procedures Using Delphi Methodology. J Surg Educ, 2014, 71(6): 810–6.

[5] Fagotti A, Petrillo M, Rossitto C, et al., Standardized Training Programmes for Advanced Laparoscopic Gynaecological Surgery. Curr Opin Obstet Gynecol, 2013, 25(4): 327–31.

[6] Vohra T T, Kinni H, Gardner-Gray J, et al., Teaching and Assessing Bedside Procedures: a Standardized Cross-Disciplinary Framework for Graduate Medical Education. Acad Med, 2024, 99(3): 266–72.

[7] Pulijala Y, Ma M, Pears M, et al., Effectiveness of Immersive Virtual Reality in Surgical Training—a Randomized Control Trial. Journal of Oral and Maxillofacial Surgery, 2018, 76(5): 1065–72.

[8] Nourissat G, Berenbaum F, Duprez D. Tendon injury: From Biology to Tendon Repair. Nature Reviews Rheumatology, 2015, 11(4): 223–33.

[9] Longo U G, Forriol F, Campi S, et al., A Biomechanical Comparison of the Primary Stability of Two Minimally Invasive Techniques for Repair of Ruptured Achilles Tendon. Knee Surgery, Sports Traumatology, Arthroscopy, 2012, 20(7): 1392–7.

[10] Mait J E, Hayes W T, Blum C L, et al., A Biomechanical Comparison of Different Tendon Repair Techniques. Journal of Long-Term Effects of Medical Implants, 2016, 26(2).

[11] Keilin C A, Farlow J L, Malloy K M, et al., Otolaryngology Curriculum During Residency Preparation Course Improves Preparedness for Internship. Laryngoscope, 2021, 131(7): E2143–E8.

[12] Sun W, Gregory D A, Tomeh M A, et al., Silk Fibroin as a Functional Biomaterial for Tissue Engineering. Int J Mol Sci, 2021, 22(3).

[13] Xie Y, Chen J, Celik H, et al., Evaluation of An Electrochemically Aligned Collagen Yarn for Textile Scaffold Fabrication. Biomed Mater, 2021, 16(2): 025001.

[14] Lohre R, Bois A J, Pollock J, et al., Effectiveness of Immersive Virtual Reality On Orthopedic Surgical Skills and Knowledge Acquisition Among Senior Surgical Residents: A Randomized Clinical Trial. Jama Network Open, 2020, 3(12): E2031217–E.

[15] Capa-Luque W, Bazán-Ramírez A, Mayorga-Falcón L E, et al., Effects of Psychology Teachers' Didactic

Performance on Student Didactic Performance. Front Psychol, 2025, 16: 1607024.

[16] Li J, Yin K, Wang Y, et al., Effectiveness of Generative Artificial Intelligence-Based Teaching Versus Traditional Teaching Methods in Medical Education: A Meta-Analysis of Randomized Controlled Trials. BMC Med Educ, 2025, 25(1): 1175.

[17] Liao W, He J, Yang C, et al., Application of a New Multi-Element Integrated Teaching Mode Based on Bite-Sized Teaching, Flipped Classroom, and MOOC in Clinical Teaching of Obstetrics and Gynaecology. BMC Med Educ, 2023, 23(1): 820.

[18] Yan D, Li G. Applying Multivariate Generalizability Theory to Compose and Rank The Evaluation Scores of College Teachers' Teaching Ability. Sci Rep, 2025, 15(1): 24091.

[19] Mandler A G. Touch Surgery: A Twenty-First Century Platform for Surgical Training. Journal of Digital Imaging, 2018, 31(5): 585–90.

[20] Kowalewski K-F, Hendrie J D, Schmidt M W, et al., Validation of the Mobile Serious Game Application Touch Surgery™ for Cognitive Training and Assessment of Laparoscopic Cholecystectomy. Surgical Endoscopy, 2017, 31(10): 4058–66.

[21] Tulipan J, Miller A, Park A G, et al., Touch Surgery: Analysis and Assessment of Validity of a Hand Surgery Simulation "App." Hand, 2019, 14(3): 311–6.

[22] Bunogerane G J, Taylor K, Lin Y, et al., Using Touch Surgery to Improve Surgical Education in Low-and Middle-income Settings: a Randomized Controlled Trial. Journal of Surgical Education, 2018, 75(1): 231–7.

[23] Bergsmann E, Schultes M T, Winter P, et al., Evaluation of Competence-Based Teaching in Higher Education: From Theory to Practice. Eval Program Plann, 2015, 52: 1–9.

[24] Stratton T D, Witzke D B, Freund M J, et al., Validating Dental and Medical Students' Evaluations of Faculty Teaching in an Integrated, Multi-instructor Course. J Dent Educ, 2005, 69(6): 663–70.

[25] Nagayo Y, Saito T, Oyama H. A Novel Suture Training System for Open Surgery Replicating Procedures Performed By Experts Using Augmented Reality. J Med Syst, 2021, 45(5): 60.

[26] Benjamin J C, Flores S, Jain P, et al., Virtual Deliberate Practice Module for Tracheostomy Change Training: An Application of Educational Design Research. Ats Sch, 2022, 3(1): 135–43.

[27] Velásquez-Plata D, Burkhardt R. Deliberate Practice: The Cornerstone of Microsurgical Therapy. Clin Adv Periodontics, 2026, 16 Suppl 1: S157–S64.

[28] Lu J, Yang X, Zhao W, et al., Effect Analysis of a Virtual Simulation Experimental Platform in Teaching Pulpotomy. BMC Med Educ, 2022, 22(1): 760.

[29] Wang L, Zhang F, Xie H. Application of Virtual Simulation in Clinical Skills and Operation Courses. Front Med (Lausanne), 2023, 10: 1184392.

[30] Wang Q, Wang F, Jiang C, et al., The Role of Cardiac Ultrasound Virtual Simulation Technology in The Construction of Clinical Diagnostic Reasoning of Structural Heart Diseases. BMC Med Educ, 2025, 25(1): 634.

[31] Wu S, Gao L, Fu J, et al., The Application of Virtual Simulation Technology in Scaling and Root Planing Teaching. Int Dent J, 2024, 74(2): 303–9.

[32] Zheng P, Yang J, Lou J, et al., Design and Application of Virtual Simulation Teaching Platform for Intelligent Manufacturing. Sci Rep, 2024, 14(1): 12895.

[33] Cao Z, Sciabola S, Wang Y. Large-Scale Pretraining Improves Sample Efficiency of Active Learning-Based Virtual Screening. J Chem Inf Model, 2024, 64(6): 1882–91.

[34] Yang B. Application of Practice-Based Learning and Improvement in Standardized Training of General

Practitioners. BMC Med Educ, 2024, 24(1): 214.

[35] Chen Q, Li Z, Tang S, et al., Development of a Blended Emergent Research Training Program for Clinical Nurses (Part 1). BMC Nurs, 2022, 21(1): 2.

[36] Morgan S. Pqrst: a Framework for Case Discussion and Practice-Based Teaching in General Practice Training. Aust J Gen Pract, 2021, 50(8): 603–6.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.