

A one-day simulation-based training program improves anesthesiologists' emergency cricothyrotomy performance and self-confidence: a multi-cohort quasi-experimental study of 460 physicians

Juan Zhi

Chinese Academy of Medical Sciences, Peking Union Medical College

Kaixuan Zhao

Tsinghua University Yuquan Hospital

Qianyu Wang

Chinese Academy of Medical Sciences and Peking Union Medical College

Quanle Liu

Chinese Academy of Medical Sciences and Peking Union Medical College

Ying Zhou

Chinese Academy of Medical Sciences and Peking Union Medical College

Jun Wu

Chinese Academy of Medical Sciences and Peking Union Medical College

Dong Yang

yangdongpumc@126.com

Chinese Academy of Medical Sciences and Peking Union Medical College

Research Article

Keywords: cricothyrotomy, emergency front-of-neck access, cannot intubate cannot oxygenate, simulation-based medical education, anesthesiology, global rating scale

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10774756/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

**A one-day simulation-based training program improves
anesthesiologists' emergency cricothyrotomy performance and self-
confidence: a multi-cohort quasi-experimental study of 460 physicians**

Authors: Juan Zhi ¹, Kaixuan Zhao², Qianyu Wang³, Quanle Liu⁴, Ying Zhou⁵, Jun Wu⁶, Dong
Yang⁷

Department of Anesthesiology, Plastic Surgery Hospital, Chinese Academy of Medical Sciences,
Peking Union Medical College, Shijingshan District, Beijing, 100144 China, Email:
honzhijuan@126.com

Department of Anesthesiology, Tsinghua University Yuquan Hospital Beijing 100040, China,
Email: zhaokaixuan0609@foxmail.com

Department of Anesthesiology, Plastic Surgery Hospital, Chinese Academy of Medical
Sciences and Peking Union Medical College, Shijingshan District, Beijing 100144,
China, Email: wqywqywangqianyu@163.com

Department of Anesthesiology, Plastic Surgery Hospital, Chinese Academy of Medical
Sciences and Peking Union Medical College, Shijingshan District, Beijing 100144,
China, Email: 491293701@qq.com

Department of Anesthesiology, Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shijingshan District, Beijing 100144, China, Email: zhouying84@whu.edu.cn

Department of Anesthesiology, Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shijingshan District, Beijing 100144, China, Email: wujun@psh.pumc.edu.cn

Department of Anesthesiology, Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shijingshan District, Beijing 100144, China, Email: yangdongpumc@126.com

Corresponding author: Dong Yang, Department of Anesthesiology, Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shijingshan District, Beijing 100144, China, Email: yangdongpumc@126.com

Abstract

Background. Emergency front-of-neck access (eFONA) by cricothyrotomy is the definitive rescue manoeuvre in a “cannot intubate, cannot oxygenate” (CICO) crisis, yet its extremely low incidence deprives anesthesiologists of real-world practice. We evaluated a structured one-day simulation-based cricothyrotomy curriculum.

Methods. In this quasi-experimental pre–post study, 460 anesthesiologists from 31 provincial-level regions of China completed six annual sessions (2019–2026) of a

standardized one-day program comprising a didactic lecture, ultrasound-guided cricothyroid membrane localization, expert demonstration, and small-group hands-on practice on a high-fidelity part-task trainer. Self-confidence in five procedural domains (5-point Likert) was measured at baseline (T0), after the didactic session (T1), and after hands-on practice (T2). Procedure time, a 33-item procedural checklist (score 0–66), and an 11-domain global rating scale (GRS, 11–77) were assessed at T1 and T2 by two trained, phase-blinded expert raters. Analyses included repeated-measures ANOVA and Friedman tests with Bonferroni-adjusted post-hoc comparisons, paired t-tests and Wilcoxon signed-rank tests with effect sizes, and linear mixed-effects models testing time × prior-experience interactions.

Results. All 460 participants completed the program. Mean procedure time decreased from 113.2 ± 16.5 s to 78.7 ± 10.9 s (mean difference –34.4 s, 95% CI –35.0 to –33.9; $p < 0.001$), checklist scores rose from 28.7 ± 4.8 to 52.7 ± 5.4 ($\Delta +23.9$, 95% CI 23.8 to 24.0; $p < 0.001$), and GRS scores rose from 32.6 ± 5.5 to 57.3 ± 5.7 ($\Delta +24.7$, 95% CI 24.6 to 24.9; $p < 0.001$). Total self-confidence increased from 13.7 ± 5.2 (T0) to 16.1 ± 4.7 (T1) and 20.2 ± 3.6 (T2) (Friedman $p < 0.001$; all Bonferroni-adjusted pairwise comparisons $p < 0.001$ except equipment-preparation confidence at T0 vs T1, which improved only after hands-on practice). Training gains were largest in participants with no prior cricothyrotomy experience (time × experience interaction, $p < 0.001$ for all outcomes). Confidence gains correlated with improvements in procedural speed but not with expert-rated quality. Post-training procedure times remained above the 40-s benchmark proposed for CICO rescue, indicating scope for mastery-oriented reinforcement.

Conclusions. A structured one-day simulation curriculum substantially improved anesthesiologists' cricothyrotomy performance and self-confidence, with the greatest absolute gains among inexperienced practitioners. Gains are best attributed to standardized goals, guided deliberate practice, and immediate feedback rather than simulator fidelity alone. Competency benchmarks and periodic refresher training should be incorporated to translate short-term gains into durable clinical readiness.

Keywords: cricothyrotomy; emergency front-of-neck access; cannot intubate cannot oxygenate; simulation-based medical education; anesthesiology; global rating scale

1. Introduction

The “cannot intubate, cannot oxygenate” (CICO) scenario is among the most time-critical emergencies in anesthesia, and emergency front-of-neck access (eFONA)—most commonly scalpel–bougie cricothyrotomy—is its definitive, life-saving resolution [1,2]. The Fourth National Audit Project (NAP4) of the Royal College of Anaesthetists and the Difficult Airway Society demonstrated that failure to attempt or delayed eFONA contributes to death and brain damage, underscoring that technical readiness, not merely theoretical knowledge, determines outcome [3]. Paradoxically, the very rarity of CICO—estimated at roughly 0.1–2.0 events per 10,000 general anesthetics [4,5]—means that most anesthesiologists will never perform a cricothyrotomy on a patient during training, and many never in an entire career [6,7]. Deliberate practice on real patients is therefore impossible, creating a structural competency gap that only simulation-based education can fill.

80 Simulation-based training is now a cornerstone of airway management education [8,9].
81 Randomized and observational studies have shown that even a single structured
82 simulation session can improve cricothyrotomy technique, compliance with difficult-
83 airway algorithms, and skill retention for up to one year [10,11]. National needs
84 assessments in emergency medicine and anesthesiology consistently rank cricothyrotomy
85 among the highest-priority technical skills for simulation-based curricula [12–14].
86 Nevertheless, important questions remain: how a comprehensive one-day curriculum—
87 integrating didactic teaching, ultrasound-guided surface anatomy, demonstration, and
88 deliberate hands-on practice—performs at scale; how its effects are distributed across
89 learners with different levels of prior experience; and how confidence gains relate to
90 objectively measured performance. Most published evaluations involve small single-cohort
91 samples (typically 20–40 trainees) [10,11,15] and rely on simple pre–post comparisons
92 that cannot separate the contributions of successive instructional components or test
93 whether training effects are moderated by baseline experience.

94 The present study addresses these gaps. Drawing on six consecutive annual sessions of a
95 standardized one-day cricothyrotomy training program delivered to 460 anesthesiologists,
96 we evaluated (i) changes in self-confidence across three measurement points, allowing the
97 didactic and hands-on components to be distinguished; (ii) changes in objectively rated
98 performance (procedure time, procedural checklist, global rating scale) before and after
99 deliberate practice; and (iii) the moderating effect of prior cricothyrotomy experience on
100 training gains, using repeated-measures and mixed-effects models with interaction terms.

2. Methods

2.1 Study design, setting, and ethics

This was a quasi-experimental, single-arm, pre–post intervention study embedded in six annual sessions (2019–2026) of the “Establishing an Emergency Front-of-Neck Airway” simulation training course at the Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, a tertiary academic medical center. The study design, curriculum sequence, and measurement timeline are summarized in Figure 1. The study involved exclusively simulation-based educational activities with no involvement of patients or animals. According to institutional policy, purely educational simulation studies of this nature are exempt from formal institutional review board approval; this exemption applies uniformly to all statements in this manuscript. The study was conducted in accordance with the Declaration of Helsinki. All participants volunteered after receiving an explanation of the educational objectives, provided informed consent for participation and for the analysis and publication of fully anonymized data; no identifiable personal data were collected. The reporting of this study follows, where applicable, the STROBE recommendations for observational studies and the GREET checklist for reporting educational interventions [16,17].

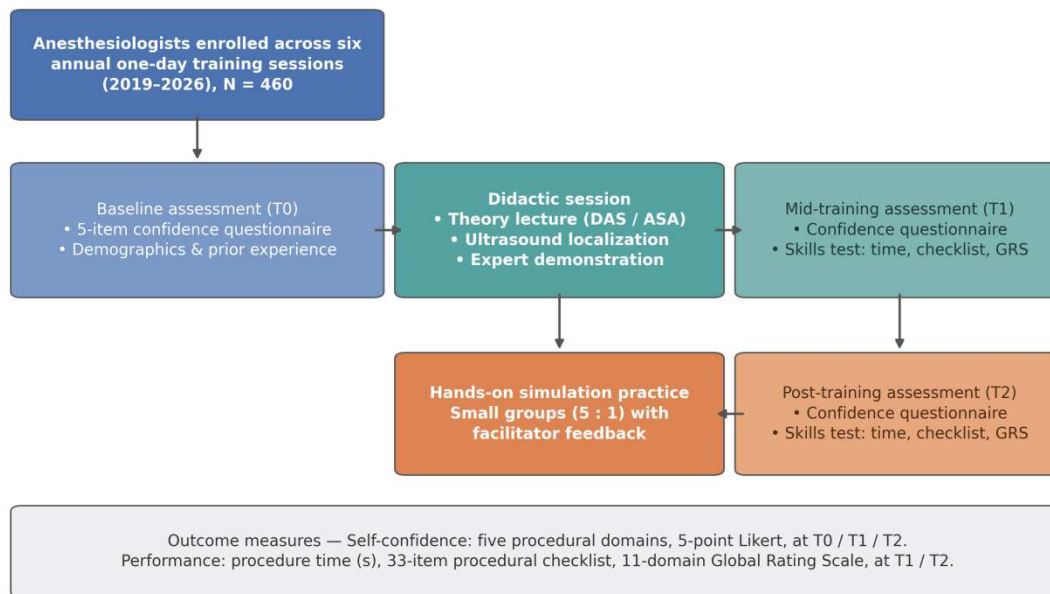


Figure 1

Figure 1. Study design, curriculum sequence, and assessment timeline. Self-confidence was measured at baseline (T0), after the didactic session (T1), and after hands-on practice (T2); objective skills tests (procedure time, 33-item checklist, 11-domain GRS) were performed at T1 and T2 with identical start (skin incision) and stop (confirmed ventilation) points.

2.2 Participants and sampling

Practicing anesthesiologists (residents and attending physicians) were recruited on a voluntary basis through institutional e-mail invitations and professional-network announcements sent to all eligible physicians before each session. We acknowledge that this constitutes convenience sampling; no random or stratified sampling was applied. To mitigate recruitment bias, invitations explicitly stated that participation was voluntary, unrelated to any performance evaluation, and open to all eligible anesthesiologists regardless of seniority. The inclusion criterion was a valid medical license in

anesthesiology. To preserve the validity of the pre-test as a baseline, individuals who had attended any previous session of this course series were ineligible. All 460 enrolled participants were first-time attendees, completed the full one-day program, and contributed complete data; there were no withdrawals and no missing outcome data.

2.3 Training intervention

The standardized one-day curriculum (see the session timetable in Supplementary Table S1) comprised the following sequential components.

Registration and baseline assessment (T0). Participants completed the baseline questionnaire, including demographics, prior cricothyrotomy experience, and the first administration of the self-confidence questionnaire.

Clinical case report. A facilitated review of a real case of post-extubation upper-airway obstruction rescue, anchoring the clinical relevance of eFONA.

Didactic session (60 min). A standardized lecture covering indications, contraindications, surface and sono-anatomy of the cricothyroid region, the scalpel–bougie–tube technique, needle cricothyrotomy, and complication management. Procedural steps were standardized according to the Difficult Airway Society (DAS) 2015 guidelines for the management of unanticipated difficult intubation in adults [1], updated in the most recent sessions to reflect the DAS 2025 guidelines [2], and the 2022 American Society of Anesthesiologists Practice Guidelines for Management of the Difficult Airway [18], adapted to local practice.

Ultrasound-guided localization (30 min). Demonstration and supervised practice of point-of-care ultrasound identification of the cricothyroid membrane (Figure 2), reflecting evidence that digital palpation alone is frequently inaccurate, particularly in females and patients with obesity [19], and the emphasis on airway ultrasound in contemporary guidance [2].

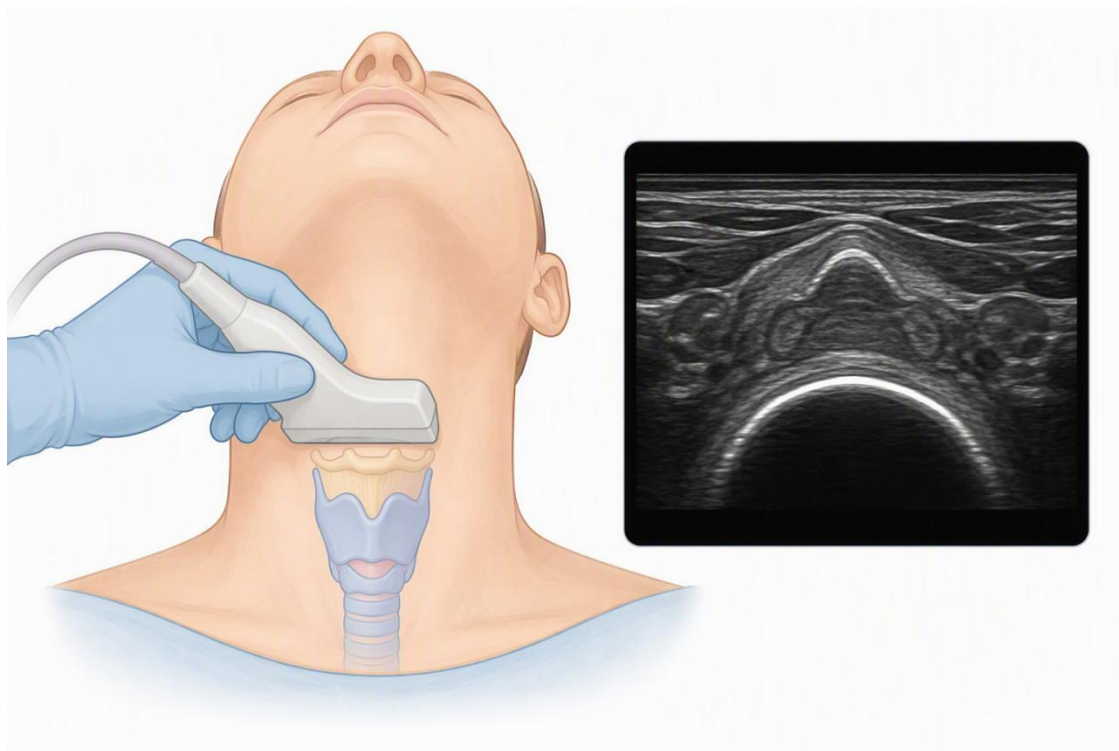
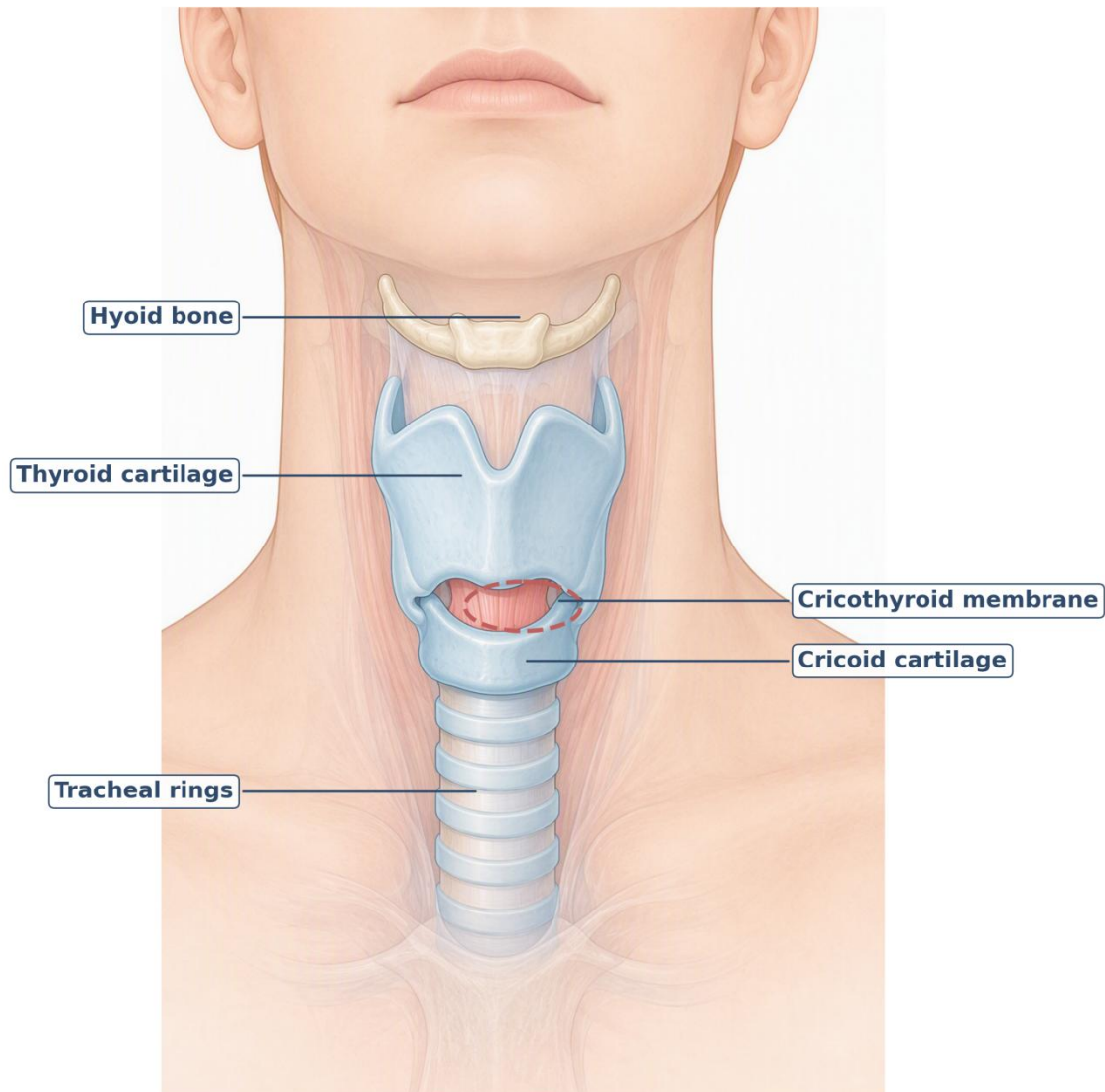


Figure 2

Figure 2. Point-of-care ultrasound identification of the cricothyroid membrane, integrated into the didactic phase of the curriculum. (BioRender-style medical illustration.)

Expert demonstration. An experienced anesthesiologist demonstrated the complete scalpel–bougie–tube technique on the simulator in real time. The relevant anatomy and the taught technique are illustrated in Figures 3 and 4.



164

165 *Figure 3*

166 *Figure 3. Anatomy of the anterior neck relevant to emergency front-of-neck access. The*
167 *cricothyroid membrane (highlighted) is bounded by the thyroid cartilage above and the*
168 *cricoid cartilage below. (BioRender-style medical illustration.)*

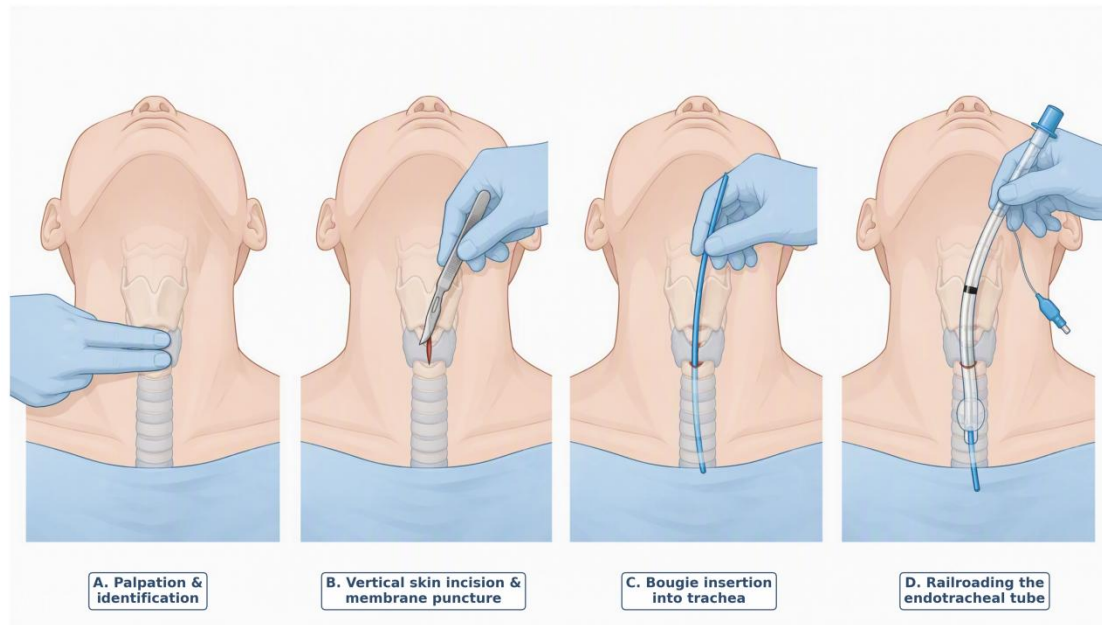


Figure 4

Figure 4. The scalpel-bougie-tube technique taught in the curriculum: (A) palpation and identification of the cricothyroid membrane; (B) vertical skin incision and membrane puncture; (C) bougie insertion into the trachea; (D) railroaded the endotracheal tube over the bougie. (BioRender-style medical illustration.)

Mid-training assessment (T1). Participants completed the second self-confidence questionnaire and then performed the pre-practice skills test: a single timed cricothyrotomy on the simulator, scored by expert raters (see Section 2.4).

Hands-on simulation practice (60 min). Deliberate practice in small groups (five participants per facilitator; six parallel groups) on the cricothyrotomy part-task trainer, with immediate structured feedback against the procedural checklist. The training materials, practice setting, and assessment procedures are illustrated in Figure 5.



182

183 *Figure 5*

184 *Figure 5. Training delivery and assessment environment. (A) Standardized cricothyrotomy kit*
 185 *(scalpel, bougie, cuffed endotracheal tube, syringe, and drapes) prepared for each*
 186 *workstation; (B) equipment inspection and assembly before practice; (C) hands-on*
 187 *cricothyrotomy practice on the high-physical-fidelity part-task trainer; (D) an expert rater*
 188 *completing the procedural checklist and global rating scale during a timed assessment.*
 189 *Photographs from the 2026 course session; no identifiable participants are shown.*

190 **Post-training assessment (T2).** A second timed and rated performance (post-training
 191 skills test) followed by the third self-confidence questionnaire.

192 **Consolidation.** A friendly skills competition and award of completion certificates
 193 concluded the day.

The cricothyrotomy simulator is a high-fidelity part-task trainer replicating the anterior neck anatomy (palpable thyroid cartilage, cricoid cartilage, cricothyroid membrane, and tracheal rings) with layered tissue haptics that reproduce skin resistance and stepwise dissection feel. It is not embedded in a full-body manikin and does not simulate bleeding or airway misting; we therefore characterize it as a high-physical-fidelity part-task trainer rather than a high-environmental-fidelity simulator (see Discussion).

2.4 Outcome measures and instruments

All instruments were administered in Mandarin Chinese. The measurement timeline and assessment endpoints are shown in Figure 1.

Self-confidence questionnaire (T0, T1, T2). Self-confidence was assessed with a study-developed, five-item questionnaire (Supplementary File S1), each item rated on a 5-point Likert scale (1 = very unconfident, 5 = very confident). The items address distinct procedural domains rather than global confidence alone: (1) identifying the cricothyroid membrane; (2) preparing the required equipment; (3) following the emergency front-of-neck airway algorithm; (4) performing scalpel-bougie cricothyrotomy; and (5) managing a CICO crisis. The item set was drafted by an expert panel of three senior anesthesiologists based on the procedural steps of the DAS guidelines [1,2], reviewed for content validity by two independent airway-management experts, and piloted with 12 anesthesiologists before the 2019 session, with minor wording adjustments for clarity. The questionnaire is not copyrighted and is freely available for research use. Because it was developed directly in Chinese, no translation procedure was required. In the present sample, internal consistency was excellent (Cronbach's α = 0.983 at T0, 0.986 at T1, and 0.965 at T2).

216 **Procedural checklist (T1, T2).** Technical performance was scored with a 33-item
217 procedural checklist (Supplementary File S2) covering preparation, landmark
218 identification, incision, membrane puncture, bougie insertion, tube railroading, cuff
219 inflation, ventilation confirmation, and post-procedure checks. Each step was scored 0 (not
220 performed), 1 (performed partially or out of sequence), or 2 (performed correctly),
221 yielding a total score of 0–66. The checklist was adapted from the DAS scalpel–bougie–tube
222 sequence [1] and previously published cricothyrotomy assessment tools [10,20].

223 **Global rating scale (GRS) (T1, T2).** Overall performance was rated on an 11-domain
224 global rating scale (Supplementary File S2), each domain scored on a 7-point anchored
225 scale (1 = very poor, 7 = excellent; total 11–77), covering preparation, respect for tissue,
226 instrument handling, procedural flow, knowledge of steps, economy of movement, use of
227 assistants, communication, response to difficulty, overall technical skill, and overall
228 performance. The instrument was adapted from the OSATS global rating scale, whose
229 reliability and construct validity are well established [21,22], and from airway-specific
230 rating tools [20].

231 **Procedure time (T1, T2).** Elapsed time was measured with a digital stopwatch by a
232 dedicated research assistant who was independent of the raters, from the moment of the
233 initial skin incision to confirmed ventilation through the placed endotracheal tube (inflated
234 cuff, manual ventilation with visible chest excursion of the test lung). Raters began
235 checklist and GRS scoring at the same start point and completed scoring at the same
236 endpoint, so that time and quality metrics referred to identical task boundaries.

Rater procedures. All performances were assessed live in the simulation room by two expert raters with clinical and teaching experience in cricothyrotomy; the final score was the mean of the two raters. Before the study, raters completed a standardized calibration session comprising detailed review of the checklist and GRS anchor descriptors, joint live scoring of two pilot performances, and consensus discussion of discrepancies. Raters were blinded to whether a performance was a pre-practice (T1) or post-training (T2) test and did not time the procedures. Because raters and trainees belonged to the same academic network, prior familiarity with individual participants could not be fully excluded; this is addressed as a limitation.

2.5 Statistical analysis

Analyses were performed with SPSS Statistics 26.0 (IBM Corp., Armonk, NY, USA) and verified in Python 3.12 (statsmodels 0.14). Distributional assumptions were examined with the Shapiro–Wilk test; because several outcomes deviated from normality, all primary inferences were confirmed with the corresponding non-parametric tests, and conclusions were unchanged. Self-confidence across the three time points was analyzed with one-way repeated-measures ANOVA per item and for the total score, complemented by Friedman tests; pairwise comparisons (T0–T1, T1–T2, T0–T2) used Wilcoxon signed-rank tests with Bonferroni adjustment (three comparisons; adjusted $\alpha = 0.0167$). Pre- versus post-practice performance (procedure time, checklist, GRS) was compared with paired t-tests and Wilcoxon signed-rank tests; effect sizes are reported as Cohen’s d_z with 95% confidence intervals (CIs) for mean differences. To test whether training effects were moderated by baseline characteristics, we fitted linear mixed-effects models with a random intercept per

participant: each outcome was regressed on measurement time, prior-experience group (none / 1 / 2–4 / ≥ 5 prior procedures), their interaction, and gender; joint significance of interaction terms was assessed with Wald χ^2 tests. Between-subgroup differences in gain scores were additionally examined with Kruskal–Wallis tests. Associations between changes in self-confidence and changes in objective performance were explored with Spearman rank correlations. A post-hoc power analysis for the primary paired comparisons was conducted. Two-sided $p < 0.05$ was considered statistically significant. No a-priori sample-size calculation was performed because the study analyzed a consecutive educational cohort; the achieved precision is reported through 95% CIs, and post-hoc power exceeded 0.99 for all primary outcomes.

3. Results

3.1 Participant characteristics

Four hundred sixty anesthesiologists were enrolled across six annual sessions (2019–2026) and all completed the program (Figure 1). Participants came from 31 provincial-level regions of China; 52.4% practiced in Beijing. Baseline characteristics are summarized in Table 1. The mean clinical experience was 6.3 ± 4.2 years (median 5); 218 participants (47.4%) were male and 242 (52.6%) female. Regarding prior cricothyrotomy experience—defined broadly as any involvement in the procedure in clinical practice or simulation, including assisting as a second operator—260 participants (56.5%) reported none, 103 (22.4%) reported one prior procedure, 49 (10.7%) two to four, and 48 (10.4%) five or more.

280 **Table 1. Participant characteristics (N = 460).**

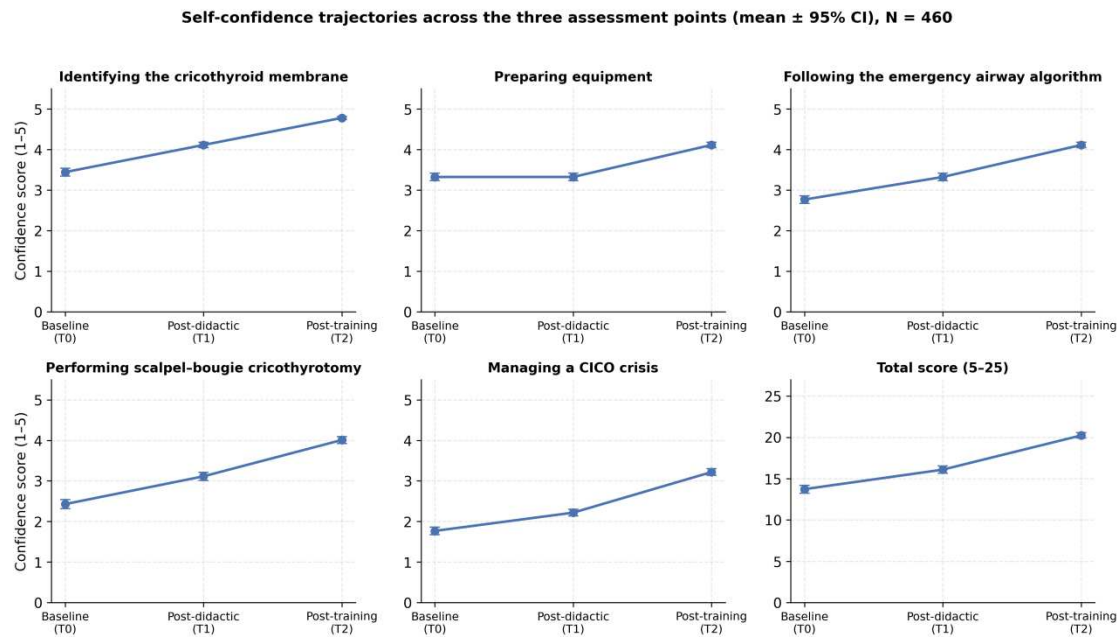
Characteristic	n (%) or mean $\pm$ SD
Gender	
Male	218 (47.4%)
Female	242 (52.6%)
Clinical experience, years	
	6.3 $\pm$ 4.2 (median 5)
< 5 years	201 (43.7%)
5–10 years	157 (34.1%)
> 10 years	102 (22.2%)
Provincial-level region of practice	
	31 regions; Beijing 241 (52.4%)
Prior cricothyrotomy experience*	
None	260 (56.5%)
1 procedure	103 (22.4%)
2–4 procedures	49 (10.7%)
$\geq$ 5 procedures	48 (10.4%)

281 * Defined as any involvement in cricothyrotomy in clinical practice or simulation, including
 282 assisted participation (see Limitations).

283 **3.2 Self-confidence across the three measurement points**

284 Self-confidence increased monotonically across T0, T1, and T2 for the total score and for
 285 four of five items (Table 2, Figure 6). Repeated-measures ANOVA showed a significant main

286 effect of time for every item and for the total score (all $p < 0.001$); Friedman tests yielded
287 identical conclusions. Bonferroni-adjusted pairwise comparisons showed significant
288 increases from baseline to post-didactic ($T0 \rightarrow T1$) and from post-didactic to post-training
289 ($T1 \rightarrow T2$) for all items (all adjusted $p < 0.001$), with one informative exception: confidence
290 in **equipment preparation** did not change after the didactic session alone ($T0\ 3.32 \pm 1.04$
291 vs $T1\ 3.32 \pm 1.04$; adjusted $p > 0.9$) and improved only after hands-on practice ($T2\ 4.11 \pm$
292 0.74 ; adjusted $p < 0.001$). Total confidence rose from 13.7 ± 5.2 ($T0$) to 16.1 ± 4.7 ($T1$) and
293 20.2 ± 3.6 ($T2$).



294
295 **Figure 6**

296 *Figure 6. Self-confidence trajectories across the three measurement points for the five*
297 *procedural domains and the total score (mean $\pm$ 95% CI, N = 460). Note the flat $T0 \rightarrow T1$*
298 *segment for equipment preparation, which improved only after hands-on practice. Error bars*
299 *represent 95% confidence intervals.*

300 **Table 2. Self-confidence scores (5-point Likert) at the three measurement points (N =**
 301 **460).**

	T0	T1 Post-	T2 Post-			
Procedural domain	Baseline	didactic	training	Overall test	T0→T1	T1→T2
Identifying the	3.44 ±	4.11 ±	4.78 ±	$\chi^2(2) =$	< 0.001	<
cricothyroid	1.05	0.74	0.41	670.5, p <		0.001
membrane				0.001		
Preparing equipment	3.32 ±	3.32 ±	4.11 ±	$\chi^2(2) =$	> 0.9	<
	1.04	1.04	0.74	726.0, p <	(n.s.)	0.001
				0.001		
Following the eFONA	2.77 ±	3.32 ±	4.11 ±	$\chi^2(2) =$	< 0.001	<
algorithm	1.01	1.04	0.74	733.4, p <		0.001
				0.001		
Performing scalpel-	2.43 ±	3.11 ±	4.01 ±	$\chi^2(2) =$	< 0.001	<
bougie	1.24	1.08	0.93	784.6, p <		0.001
cricothyrotomy				0.001		
Managing a CICO	1.77 ±	2.22 ±	3.22 ±	$\chi^2(2) =$	< 0.001	<
crisis	1.01	0.91	0.91	854.0, p <		0.001
				0.001		
Total score (5-25)	13.73 ±	16.09 ±	20.23 ±	$\chi^2(2) =$	< 0.001	<
	5.19	4.71	3.59	897.9, p <		0.001

	T0	T1 Post-	T2 Post-			
Procedural domain	Baseline	didactic	training	Overall test	T0→T1	T1→T2

0.001

Values are mean ± SD. Overall test: Friedman test across the three time points (repeated-measures ANOVA was concordant, all $p < 0.001$). Pairwise columns: Bonferroni-adjusted Wilcoxon signed-rank p-values (three comparisons per row; adjusted $\alpha = 0.0167$). T0→T2 comparisons were significant for all rows (all adjusted $p < 0.001$). n.s. = not significant.

3.3 Objective performance before and after hands-on practice

All three objective performance measures improved substantially between the pre-practice (T1) and post-training (T2) tests (Table 3, Figure 7). Mean procedure time decreased by 34.4 s (95% CI -35.0 to -33.9), from 113.2 ± 16.5 s to 78.7 ± 10.9 s (paired $t(459) = 122.5$, $p < 0.001$; Wilcoxon $p < 0.001$; Cohen's $d_z = 5.71$). Checklist scores increased by 23.9 points (95% CI 23.8 to 24.0), from 28.7 ± 4.8 to 52.7 ± 5.4 ($p < 0.001$; $d_z = 22.1$), and GRS scores increased by 24.7 points (95% CI 24.6 to 24.9), from 32.6 ± 5.5 to 57.3 ± 5.7 ($p < 0.001$; $d_z = 16.5$). These effect sizes are exceptionally large, which is typical when trainees are re-tested on the same task and device immediately after guided practice; we therefore emphasize the mean differences and their CIs rather than the standardized metrics (see Limitations).

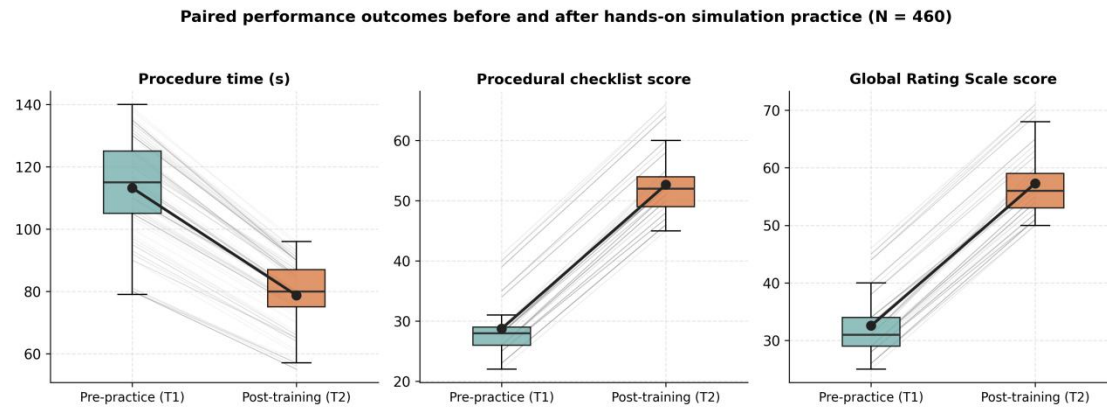


Figure 7

Figure 7. Paired objective performance before (T1) and after (T2) hands-on simulation practice (N = 460): procedure time (left), procedural checklist score (middle), and global rating scale score (right). Boxes show medians and interquartile ranges; faint lines connect individual paired performances (random subset of 160 participants shown for clarity); bold lines connect means. All comparisons $p < 0.001$.

Table 3. Objective performance outcomes before and after hands-on simulation practice (N = 460).

Outcome	Pre-practice (T1)	Post-training (T2)	Mean difference (95% CI)	Paired t-test / Wilcoxon	Cohen's d_z
Procedure time, s	113.2 ± 16.5	78.7 ± 10.9	-34.4 (-35.0 to -33.9)	$p < 0.001$ / $p < 0.001$	5.71
Checklist score (0-66)	28.7 ± 4.8	52.7 ± 5.4	+23.9 (23.8 to 24.0)	$p < 0.001$ / $p < 0.001$	22.09

Outcome	Pre- practice (T1)	Post- training (T2)	Mean difference (95% CI)	Paired t-test / Wilcoxon	Cohen's d_z
GRS score (11–77)	32.6 ± 5.5	57.3 ± 5.7	+24.7 (24.6 to 24.9)	p < 0.001 / p < 0.001	16.50

Values are mean ± SD. All pre–post comparisons were significant with both the paired t-test and the Wilcoxon signed-rank test. The timing endpoint was confirmed ventilation; checklist and GRS scoring covered the identical task segment (Section 2.4).

3.4 Moderation by prior experience and the confidence–competence relationship

Prior experience strongly moderated the magnitude—but not the direction—of training gains (Figure 8). Mixed-effects models showed significant time × experience interactions for procedure time (Wald $\chi^2(3) = 647.7$, $p < 0.001$), checklist ($\chi^2(3) = 121.8$, $p < 0.001$), and GRS ($\chi^2(3) = 215.8$, $p < 0.001$). The reduction in procedure time was greatest among inexperienced participants (none: -37.8 ± 4.9 s; 1 prior: -33.2 ± 2.4 s; 2–4: -29.2 ± 1.8 s; ≥ 5 : -24.0 ± 1.2 s; Kruskal–Wallis $p < 0.001$), and the same gradient held for total confidence gains (7.8 ± 1.1 , 7.4 ± 1.5 , 3.0 ± 0.0 , and 1.0 ± 0.0 points across the four groups; $p < 0.001$), consistent with a ceiling effect. In the mixed-effects model of total confidence, the time × experience interaction was highly significant (Wald $\chi^2(6) = 2910.1$, $p < 0.001$). Checklist and GRS gains were uniformly large across subgroups, indicating that procedural quality improved irrespective of baseline experience, whereas efficiency (time) and self-confidence depended on the starting level. An equivalent pattern was observed across strata of clinical seniority (years of practice; all Kruskal–Wallis $p < 0.001$). Gender was

associated with small absolute differences in gain scores (e.g., confidence gain: male 5.2 vs female 7.6 points, $p < 0.001$) that are unlikely to be educationally meaningful.

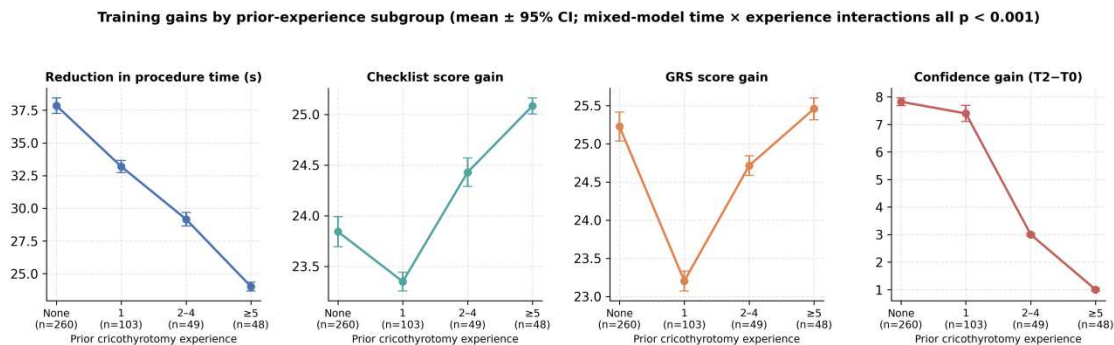


Figure 8

Figure 8. Training gains by prior-experience subgroup (mean $\pm$ 95% CI). Reduction in procedure time and confidence gains decreased with increasing prior experience, whereas checklist and GRS gains were uniformly large; all time $\times$ experience interactions were significant ($p < 0.001$, linear mixed-effects models).

Exploratory correlation analyses further dissociated confidence from competence. The gain in total self-confidence correlated moderately with the reduction in procedure time (Spearman $\rho = -0.47$, $p < 0.001$) but was essentially uncorrelated with the gain in expert-rated global performance (GRS gain: $\rho = 0.03$, $p = 0.50$) and only weakly with the checklist gain ($\rho = -0.21$, $p < 0.001$). Participants thus felt more confident largely in proportion to how much faster they became—not in proportion to how much their rated procedural quality improved.

4. Discussion

4.1 Principal findings

In this large multi-cohort evaluation, a standardized one-day simulation-based curriculum was associated with significant and substantial improvements in every measured outcome: procedural efficiency (34.4 s faster), procedural quality (checklist +23.9 of 66 points; GRS +24.7 of 77 points), and domain-specific self-confidence. Three findings add granularity to the existing literature. First, decomposing the course into didactic and hands-on phases showed that most confidence domains improved after the lecture, but confidence in equipment preparation increased only after deliberate hands-on practice—empirically supporting the intuition that psychomotor readiness, not declarative knowledge, drives preparedness for equipment handling. Second, prior experience moderated gains: novices improved the most in speed and confidence, while all groups improved similarly in procedural quality, implying that a single standardized curriculum can serve heterogeneous trainee populations but that refresher value for experienced practitioners lies mainly in quality consolidation rather than speed. Third, despite a 30% relative reduction, post-training procedure times (mean 78.7 s) remained well above the ≤ 40 -s benchmark achieved after five consecutive mannequin attempts in the classic study of Wong and colleagues [23]. This suggests that our one-session, single-attempt format should evolve toward mastery learning with repeated attempts to a time criterion [24], a design we have begun to explore with the end-of-course skills competition.

4.2 Comparison with previous work

Our pre-training median procedure time (115 s) closely mirrors the 117-s pretest median reported by Hubert et al. in 27 anesthesiology residents [10], and our post-training times (75–80 s interquartile range) resemble their 3-, 6-, and 12-month post-test values (52–69 s), although our participants were re-tested immediately rather than after a delay. Boet et al. demonstrated that cricothyrotomy skills acquired in a single simulation session are retained for at least one year [11]; conversely, quantitative reviews of skill decay caution that complex procedural skills deteriorate without reinforcement [25], and Wong et al. explicitly noted that optimal retraining intervals remain undetermined [23]. Our results therefore support the feasibility of large-scale short courses while reinforcing the case for scheduled refresher training. The observed experience-dependent moderation is consistent with Friedman et al., who found that model fidelity mattered little once instructional goals and guidance were held constant [26], and with meta-analyses showing comparable learning outcomes across fidelity levels [27–29].

4.3 Interpreting confidence alongside competence

Self-confidence is a slippery metric when isolated from performance: confidence without competence risks harm, and simulation itself can inflate self-assessment [28]. Framed within Miller’s pyramid, our instruments assess the “shows how” level (checklist and GRS under observation) while the questionnaire captures learner perception—Kirkpatrick level 1–2 rather than behavior or patient outcomes [30,31]. Three features of our confidence data deserve explicit comment. First, the uniform direction of change (no participant reported lower confidence after training) is expected in an unblinded educational

intervention in which learners know they have just been taught and wish to report benefit; it does not by itself constitute evidence of learning. Second, and more informatively, confidence gains tracked improvements in speed but not in expert-rated quality—a dissociation that mirrors reports that simulation can breed overconfidence [28] and reinforces the primacy of observed performance metrics. Third, because self-ratings are vulnerable to social-desirability and response-shift bias, the confidence results should be read as motivational outcomes that complement, rather than substitute for, the objective assessments. Composite evaluation frameworks that jointly weigh speed, correctness, and global quality—and ultimately clinical transfer—should anchor future studies [21,30,32].

4.4 Fidelity reconsidered

We deliberately avoid attributing the observed gains to simulator fidelity. Participants were trained and tested on the same device; comparable gains would be expected on a cadaver or an inexpensive model given identical goals, guidance, and feedback [26–29]. Indeed, fidelity should be construed as a property of the overall training environment rather than of the physical simulator alone [32]. In this spirit, we acknowledge that our environment lacked several fidelity elements—the part-task trainer was not embedded in a full-body manikin, no bleeding or physiological cues were simulated, and participants did not routinely wear gloves—so our trainer is best described as a high-physical-fidelity part-task trainer embedded in a moderate-fidelity environment. What likely drove learning was the alignment of standardized objectives, expert demonstration, ultrasound-verified landmark identification, small-group deliberate practice, and immediate checklist-anchored feedback.

4.5 Limitations

Several limitations merit emphasis. (1) The single-arm quasi-experimental design without a control group precludes causal attribution; history, maturation, and test-retest effects are plausible alternative explanations. (2) Participants were re-tested on the same simulator and task immediately after practice, which inflates measured gains; the exceptionally large effect sizes should be interpreted accordingly. (3) Convenience sampling of volunteers may over-represent motivated learners, limiting generalizability. (4) Prior experience was self-reported and broadly defined (any involvement, including assistance), which likely explains the higher-than-expected proportion of participants reporting prior cricothyrotomy exposure despite the procedure's rarity; future work should distinguish independent performance from assisted participation. (5) Raters, although trained, calibrated, and blinded to test phase, belonged to the same academic network and could have known some participants; video-based scoring by external raters would strengthen future validity evidence. (6) Confidence was self-assessed without blinding and is subject to social-desirability and response-shift bias; as expected under these conditions, ratings moved only upward. (7) Only immediate outcomes were measured; retention and clinical transfer were not assessed [23,25]. (8) The checklist, GRS, and confidence questionnaire were developed or adapted locally and, although internally consistent and content-validated, warrant formal psychometric validation in independent samples.

4.6 Implications and future directions

For educators, the curriculum demonstrates that a reproducible one-day format can deliver uniform procedural-quality gains to hundreds of anesthesiologists per year, with the

largest efficiency and confidence returns in the least experienced. Embedding a ≤ 40 -s mastery criterion [23], spaced refresher sessions [11,25], and longitudinal tracking of transfer to clinical CICO events are logical next steps. For researchers, adequately powered randomized comparisons of curriculum components (e.g., ultrasound localization vs palpation-only; one vs multiple attempts) and formally validated composite outcome instruments are needed.

5. Conclusions

A structured one-day simulation-based cricothyrotomy curriculum combining didactic teaching, ultrasound-guided landmark identification, expert demonstration, and facilitated deliberate practice was followed by significant improvements in anesthesiologists' procedural speed, procedural quality, and domain-specific self-confidence, with the greatest absolute gains among practitioners without prior experience. Improvements should be credited to the structured instructional design rather than to simulator fidelity per se. Because confidence gains tracked speed rather than expert-rated quality, and post-training times remained above recommended CICO benchmarks, curricula should pair confidence-building with objective competency criteria and periodic retraining to convert immediate gains into durable readiness for this rare but life-saving procedure.

Declarations

Ethics approval and consent to participate.

This study involved only simulation-based educational activities without patient or animal involvement and was exempt from formal ethics review according to the the Review Board

of Plastic Surgery Hospital, Chinese Academy of Medical Sciences. All participants provided written informed consent for participation in the study.

Consent for publication

Not Applicable

Availability of data and materials

The datasets generated and analyzed during this study are available in the Zenodo repository: <https://doi.org/10.5281/zenodo.14632368>. The complete confidence questionnaire and the checklist/GRS instruments are provided as Supplementary Files S1 and S2.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

JZ: Conceptualization, Methodology, Investigation, Resources, Writing – review & editing.

KXZ: Data acquisition, analysis and interpretation, drafting the manuscript.

QYW: Investigation, Data curation, Writing – review & editing.

QLL: Investigation, Data curation, Writing – review & editing.

YZ: Investigation, Data curation, Visualization, Writing – review & editing.

JW: Investigation, Resources, Writing – review & editing.

DY: Conceptualization, Methodology, Supervision, Writing – review & editing.

All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgement

We thank the course faculty, facilitators, and expert raters of the six training sessions for their dedication and contribution to this study.

References

- 491 Frerk C, Mitchell VS, McNarry AF, Mendonca C, Bhagrath R, Patel A, et al. Difficult Airway
492 Society 2015 guidelines for management of unanticipated difficult intubation in adults. Br J
493 Anaesth. 2015;115(6):827-848.
- 494 Ahmad I, et al. Difficult Airway Society 2025 guidelines for management of unanticipated
495 difficult tracheal intubation in adults. Br J Anaesth. 2026 (in press).
496 doi:10.1016/j.bja.2025.10.006.
- 497 Cook TM, Woodall N, Frerk C; Fourth National Audit Project. Major complications of airway
498 management in the UK: results of the Fourth National Audit Project of the Royal College of
499 Anaesthetists and the Difficult Airway Society. Part 1: anaesthesia. Br J Anaesth.
500 2011;106(5):617-631.
- 501 Nagaro T, Yorozya T, Sotani M, Adachi N, Tabo E, Arai T, et al. Survey of patients whose
502 lungs could not be ventilated and whose trachea could not be intubated in university
503 hospitals in Japan. J Anesth. 2003;17(4):232-240.
- 504 Tachibana N, Niiyama Y, Yamakage M. Incidence of cannot intubate-cannot ventilate
505 (CICV): results of a 3-year retrospective multicenter clinical study in a network of
506 university hospitals. J Anesth. 2015;29(3):326-330.
- 507 Bainbridge D, Cheng DC. Cricothyrotomy, simulated and real: a review. Anesth Analg.
508 2020;130(6):1720-1731.
- 509 Lockey D, Crewdson K, Weaver A, Davies G. Observational study of the success rates of
510 intubation and failed intubation airway rescue techniques in 7256 attempted intubations of
511 trauma patients by pre-hospital physicians. Br J Anaesth. 2014;113(2):220-225.

512 Kennedy CC, Cannon EK, Warner DO, Cook DA. Advanced airway management simulation
513 training in medical education: a systematic review and meta-analysis. Crit Care Med.
514 2014;42(1):169-178.

515 Barsuk JH, Cohen ER, Caprio T, McGaghie WC, Simuni T, Wayne DB. Simulation-based
516 education with mastery learning improves residents' lumbar puncture skills. Neurology.
517 2012;79(2):132-137.

518 Hubert V, Duwat A, Deransy R, Mahjoub Y, Dupont H. Effect of simulation training on
519 compliance with difficult airway management algorithms, technical ability, and skills
520 retention for emergency cricothyrotomy. Anesthesiology. 2014;120(4):999-1008.

521 Boet S, Borges BCR, Naik VN, Siu LW, Riem N, Chandra D, et al. Complex procedural skills
522 are retained for a minimum of 1 yr after a single high-fidelity simulation training session.
523 Br J Anaesth. 2011;107(4):533-539.

524 Husted A, Rölting JD, Ingeman ML, et al. Identifying technical skills and clinical procedures
525 for simulation-based training in emergency medicine: A nationwide needs assessment. Am
526 J Emerg Med. 2022;62:140-143.

527 Kester-Greene N, Hall AK, Walsh CM. Simulation curricular content in postgraduate
528 emergency medicine: A multicentre Delphi study. CJEM. 2019;21(5):667-675.

529 Bessmann EL, Østergaard HT, Nielsen BU, et al. Consensus on technical procedures for
530 simulation-based training in anaesthesiology: A Delphi-based general needs assessment.
531 Acta Anaesthesiol Scand. 2019;63(6):720-729.

532 Aho JM, Thiels CA, AlJamal YN, Ruparel RK, Rowse PG, Heller SF, et al. Every surgical
533 resident should know how to perform a cricothyrotomy: an inexpensive cricothyrotomy
534 task trainer for teaching and assessing surgical trainees. J Surg Educ. 2015;72(4):658-661.

535 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE
536 Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology
537 (STROBE) statement: guidelines for reporting observational studies. *Lancet*.
538 2007;370(9596):1453-1457.

539 Phillips AC, Lewis LK, McEvoy MP, Galipeau J, Glasziou P, Moher D, et al. Development and
540 validation of the guideline for reporting evidence-based practice educational interventions
541 and teaching (GREET). *BMC Med Educ*. 2016;16:237.

542 Apfelbaum JL, Hagberg CA, Connis RT, Abdelmalak BB, Agarkar M, Dutton RP, et al. 2022
543 American Society of Anesthesiologists practice guidelines for management of the difficult
544 airway. *Anesthesiology*. 2022;136(1):31-81.

545 Kristensen MS, Teoh WH, Rudolph SS. Ultrasonographic identification of the cricothyroid
546 membrane: best evidence, techniques, and clinical impact. *Br J Anaesth*. 2016;117(Suppl
547 1):i39-i48.

548 Ahmed A, Khan FA, Ismail S. Reliability and validity of a tool to assess airway management
549 skills in anesthesia trainees. *J Anaesthesiol Clin Pharmacol*. 2016;32(3):333-338.

550 Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective
551 structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*.
552 1997;84(2):273-278.

553 Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an
554 innovative “bench station” examination. *Am J Surg*. 1997;173(3):226-230.

555 Wong DT, Prabhu AJ, Coloma M, Imasogie N, Chung FF. What is the minimum training
556 required for successful cricothyroidotomy? A study in mannequins. *Anesthesiology*.
557 2003;98(2):349-353.

558 McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based
559 mastery learning with translational outcomes. *Med Educ.* 2014;48(4):375-385.

560 Arthur W Jr, Bennett W Jr, Stanush PL, McNelly TL. Factors that influence skill decay and
561 retention: a quantitative review and analysis. *Hum Perform.* 1998;11(1):57-101.

562 Friedman Z, You-Ten KE, Bould MD, Naik V. Teaching lifesaving procedures: the impact of
563 model fidelity on acquisition and transfer of cricothyrotomy skills to performance on
564 cadavers. *Anesth Analg.* 2008;107(5):1663-1669.

565 Zeng Q, Wang K, Liu W, et al. Efficacy of high-fidelity simulation in advanced life support
566 training: a systematic review and meta-analysis of randomized controlled trials. *BMC Med*
567 *Educ.* 2023;23:746.

568 Massoth C, Röder H, Ohlenburg H, Hessler M, Zarbock A, Pöpping DM, et al. High-fidelity is
569 not superior to low-fidelity simulation but leads to overconfidence in medical students.
570 *BMC Med Educ.* 2019;19:29.

571 Kim J, Park JH, Shin S. Effectiveness of simulation-based nursing education depending on
572 fidelity: a meta-analysis. *BMC Med Educ.* 2016;16:152.

573 Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.*
574 1990;65(9 Suppl):S63-S67.

575 Yardley S, Dornan T. Kirkpatrick's levels and education 'evidence'. *Med Educ.*
576 2012;46(1):97-106.

577 Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-
578 based training. *Acad Med.* 2014;89(3):387-392.

579 Rai MR, Popat MT. Evaluation of airway equipment: man or manikin? Anaesthesia.
580 2011;66(1):1-3.

581 Kuduvalli PM, Jervis A, Tighe SQ, Robin NM. Unanticipated difficult airway management in
582 anaesthetised patients: a prospective study of the effect of mannequin training on
583 management strategies and skill retention. Anaesthesia. 2008;63(4):364-369.

Supplementary Files

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

- [SupplementaryMaterials.docx](#)