Langenbeck's Archives of Surgery https://doi.org/10.1007/s00423-026-04013-6

Article in Press

Simulation unlocked: transforming thyroid surgery training through high-fidelity and curriculum-based innovation—a systematic review

Received: 13 October 2025

Accepted: 3 March 2026

Published online: 12 March 2026

Cite this article as: Zhang D., Brucchi F., Barbieri D. *et al.* Simulation unlocked: transforming thyroid surgery training through high-fidelity and curriculum-based innovation—a systematic review. *Langenbecks Arch Surg* (2026). https://doi.org/10.1007/s00423-026-04013-6

Daqi Zhang, Francesco Brucchi, Diego Barbieri, Hoon Yub Kim, Carla Colombo, Gianluca Donatini, Paolo Miccoli, Wu Che-Wei, Gianlorenzo Dionigi & Lanlan Wan

Title

SIMULATION UNLOCKED: TRANSFORMING THYROID SURGERY TRAINING THROUGH HIGH-FIDELITY AND CURRICULUM-BASED INNOVATION—A SYSTEMATIC REVIEW

Running Title

Training in Thyroid Surgery

AUTHORS

Daqi Zhang[1] , Francesco Brucchi[2,3], Diego Barbieri[4,5], Hoon Yub Kim[6], Carla Colombo[7], Gianluca Donatini[8], Paolo Miccoli[9], Wu Che-Wei[10], Gianlorenzo Dionigi[2,3], Lanlan Wan [11]*

AFFILIATIONS

1. Division of Thyroid Surgery, China-Japan Union Hospital of Jilin University, Jilin Provincial Key Laboratory of Surgical Translational Medicine, Jilin Provincial Precision Medicine Laboratory of Molecular Biology and Translational Medicine on Differentiated Thyroid Carcinoma, Changchun City, Jilin Province, China
2. Division of General Surgery, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Auxologico Italiano, Milan, Italy
3. Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy
4. Division of Head and Neck, San Raffaele Scientific Institute IRCCS (Istituto Di Ricovero E Cura a Carattere Scientifico), Otorhinolaryngology Unit, Milan, Italy.

5. School of Medicine, Vita-Salute San Raffaele University, Milan, Italy.
6. Department of Surgery, KUMC Thyroid Center, Korea University College of Medicine, Korea University Hospital, Seoul, Korea.
7. Division of Endocrinology, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Auxologico Italiano, Milan, Italy
8. Department of General and Endocrine Surgery, University of Poitiers, CHU Poitiers, Poitiers, France
9. Department of Surgical, Medical, Pathology, and Critical Care, The University of Pisa, 56124, Pisa, Italy.
10. Department of Otolaryngology-Head and Neck Surgery, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan.
11. Department of Anesthesia, The Second Hospital of Jilin University, Changchun City, Jilin Province, China

***CORRESPONDING AUTHORS**

Dr Lanlan Wan

No.218 Ziqiang Street, Nanguan District, Changchun, Jilin, 130000, China.

Email: wanll@jlu.edu.cn

PROSPERO registration number: CRD420251066856

ABSTRACT

Background: Simulation-based and structured training have transformed surgical education, providing safe environments for skill acquisition and performance assessment. However, evidence in thyroid surgery remains fragmented across modalities such as simulation, video-based learning, intraoperative neuromonitoring (IONM), and fellowship programs. This systematic review aimed to synthesize primary evidence on educational interventions in thyroid surgery and their impact on technical performance and patient outcomes.

Methods: A systematic search of PubMed, Embase, Scopus, and Web of Science was conducted up to April 2025. Eligible studies included randomized trials, prospective or retrospective cohorts, and validation studies assessing simulation-based, structured, or supervised training. Two reviewers independently screened and extracted data, assessing quality using RoB 2, ROBINS-I, or CASP. Due to heterogeneity, results were summarized through a structured narrative synthesis by training modality.

Results: Ten primary studies (2011–2025) were included. Simulation and virtual-reality training improved technical performance, task completion time, and perceived realism. Supervised resident training reduced operative time after 30–50 cases, approximating senior benchmarks. Video-assisted learning decreased error rates by 49 % and staff takeovers by 52 %. IONM-based education lowered recurrent laryngeal nerve injury rates and enhanced procedural standardization. Fellowship-trained surgeons showed shorter operative times and fewer complications compared with non-fellowship peers. Across all modalities, training interventions improved either performance or safety outcomes.

Conclusions: Structured educational programs—including simulation, supervised learning, and IONM or fellowship curricula—enhance surgical performance and safety in thyroid surgery. Despite limited quantitative evidence, findings support integrating multimodal, competency-based education into surgical training. Standardized outcome measures and multicenter validation are needed to define optimal learning pathways.

Keywords: thyroid surgery, surgical education, simulation, virtual reality, intraoperative neuromonitoring, fellowship training, learning curve, narrative synthesis

INTRODUCTION

Thyroid surgery requires precise anatomical dissection in a compact and delicate region of the neck, where critical structures such as the recurrent laryngeal nerve, the superior laryngeal nerve, the parathyroid glands, and important vascular structures are at risk [1–3]. Even minor technical errors can lead to complications such as vocal cord paralysis, hypocalcemia, or bleeding, which can significantly impact the patient's outcome and quality of life [4]. Traditionally, surgical expertise in thyroidectomy has been acquired through the apprenticeship model, which relies heavily on hands-on surgical experience under expert supervision [5].

However, the increasing importance of patient safety, ethical concerns regarding training with real patients, time constraints within surgical residency training, and the demand for standardised assessment of competency have led to the integration of simulation and structured training into surgical training [6–9]. Modern surgical training faces additional challenges, such as a reduced number of surgical procedures due to work time constraints, large fluctuations in caseloads, and unequal access to complex procedures, which impede equitable competency acquisition. Consequently, there is a need for reproducible, scalable, and safe training methods that guarantee competency prior to independent surgical practice.

Simulation provides a controlled, reproducible environment in which trainees can develop technical skills, anatomical knowledge, and intraoperative decision-making without compromising patient safety. This training approach supports deliberate practise, repeated exposure to rare or critical situations, immediate feedback, and the opportunity for objective assessment of skills, thus promoting both technical skills and non-technical skills such as communication and crisis management [7]. Furthermore, the integration of simulation into formal curricula enables transparent assessment of learning progress and supports remediation, helping to standardise training outcomes across institutions [10].

In other surgical fields such as laparoscopy, endoscopy, and robotic surgery, simulation-based training has been shown to accelerate the learning curve, improve technical skills, and reduce surgical errors [11–13]. Importantly, these approaches have been associated with improved training outcomes, higher confidence, and safer operative performance among trainees. Despite these advances, simulation remains less established in thyroid and endocrine surgery, and widespread standardisation is lacking. Available models range from lifelike silicone neck replicas to cadaver and porcine platforms to newer virtual reality-based systems. Structured curricula also vary considerably in terms of content, teaching methods, duration, and outcome assessment tools [12–16].

Initial studies have shown that simulations can improve anatomical knowledge, enhance surgical safety, improve technical skills, and reduce early operative time in thyroidectomy [17–19]. From an educational perspective, simulation can also help address inequalities in operative experience by providing all trainees with the opportunity to meet competency benchmarks regardless of clinical volume. However, these findings are based on small, heterogeneous studies with varying outcome measures. To date, there is no comprehensive quantitative synthesis definitively evaluating the effectiveness of simulation and structured training on clinical or educational outcomes in thyroid surgery.

This systematic review aims to systematically evaluate the effects of simulation and training curricula on operative performance, technical skill acquisition, knowledge retention, and surgical safety in thyroidectomy to inform future curricula and guidelines design for endocrine surgical training.

METHODS

This study was reported in accordance with the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and the guidelines of the A MeaSurement Tool to Assess systematic Reviews (AMSTAR II) [20, 21]. The protocol was designed a priori and included the development of a structured search strategy, defined inclusion and exclusion criteria, independent data extraction and an assessment of risk of bias. The protocol was also registered with PROSPERO (PROSPERO 2025: CRD420251066856).

Search strategy

A comprehensive systematic search of PubMed, Embase, Web of Science and Cochrane CENTRAL was conducted from the launch of the database until 11 October 2025. The search terms included controlled vocabulary (e.g. MeSH and EMTREE) and free text terms such as: "thyroid surgery", "simulation", "surgical training", "skill acquisition", "learning curve", "structured curriculum" and "training outcomes". Boolean operators and truncation strategies were used to maximise sensitivity. In addition to the database searches, the references of the included articles and relevant reviews were manually screened to identify further suitable studies. The detailed search strategies can be found in the Supplementary Materials (Supplementary Table 1).

Selection of studies

Two independent reviewers (author DZ and author GD) screened all titles and abstracts, followed by a review of the full texts of potentially relevant articles. Discrepancies were resolved by discussion or by the decision of a third reviewer (CC).

The inclusion criteria were:

Eligible studies met the following criteria:

1. **Population:** surgical trainees or practicing surgeons involved in open or minimally-invasive thyroidectomy;
2. **Intervention:** any structured educational or training program, including simulation-based, video-assisted, intraoperative supervised, or fellowship-level training;
3. **Comparator:** standard practice, pre-training baseline, or non-trained participants;

4. **Outcomes:** measures of operative performance (e.g., operative time, errors, OSATS), safety (e.g., RLN injury, complications), or educational metrics (knowledge, confidence, satisfaction);
5. **Study design:** randomized controlled trials, prospective or retrospective cohorts, or validation studies reporting original data.

Systematic reviews, editorials, and case reports were excluded. Paper not written in English were excluded. When multiple publications referred to overlapping datasets, only the most complete version was retained. This selection strategy is in line with previously validated meta-analytical approaches in surgical training research [11, 12].

Data extraction and quality assessment

Data were extracted independently by two reviewers using a standardised data collection form. Variables extracted included: author, year of publication, country, study design, sample size, participant level (e.g. residents, fellows), simulation modality (e.g. porcine, synthetic, VR), number and duration of training sessions, and outcomes measured. The Cochrane Risk of Bias Tool (RoB II) was used to assess the quality of RCTs in five domains: randomisation, deviations from planned interventions, missing outcome data, outcome measurement and selection of reported outcomes [22]. Non-randomised comparative studies were assessed using the Newcastle-Ottawa Scale (NOS), which assesses the domains of selection, comparability and outcomes [23] (Supplementary Table 2). The overall quality of the evidence was assessed using the GRADE (Grading of Recommendations, Assessment, Development and Evaluations) approach [24]. Based on the overall assessment, the quality was categorised into four levels (high, moderate, low or very low). The studies were either downgraded or upgraded in quality depending on whether the criteria of risk of bias, inconsistency, indirectness, imprecision, publication bias, large magnitude, dose dependence or effect of all plausible confounders were met. The authors DZ and GD performed the GRADE assessment (Supplementary Table 3). We also use the Medical Education Research Study Quality Instrument (MERSQI), a tool specifically designed to assess the methodological quality of education and training interventions [25].

Outcomes

The review focused on outcomes reflecting both technical performance and educational effectiveness.

Primary endpoints included improvements in operative performance—such as reduction in operative time and enhancement of technical skills—when assessed through validated rating scales (e.g., global rating scales, task-specific checklists, Objective Structured Assessment of Technical Skills).

Secondary outcomes comprised indices of patient safety (e.g., recurrent laryngeal nerve injury, hypocalcemia, overall complication rate) and learning-related outcomes, including knowledge acquisition, trainee confidence, and progression along the learning curve.

Additional educational endpoints considered were retention of knowledge, self-assessment of readiness, and perceived procedural safety. When available, studies reporting validation of simulation models, trainee satisfaction, or impact of structured curricula on standardization of outcomes were also analyzed.

Data Synthesis

Given the heterogeneity of study designs, interventions, and reported outcomes, a quantitative meta-analysis was not performed. Instead, a structured narrative synthesis was conducted, grouping studies into five major training domains:

1. Simulation and virtual/augmented reality;
2. Supervised resident training and learning curve studies;
3. Video-assisted learning modules;
4. Intraoperative neuromonitoring (IONM) educational programs;
5. Fellowship and structured curricula.

Within each domain, results were compared qualitatively in terms of technical performance, safety outcomes, and educational effectiveness. Trends and convergences among studies were described narratively, emphasizing consistencies and gaps in the current literature.

RESULTS

Study Identification and Selection

The comprehensive search strategy initially identified 540 records through database and manual searches. Of these, 60 full-text articles were assessed for eligibility according to predefined criteria. Ultimately, 9 studies met all inclusion criteria and were included in the qualitative and quantitative synthesis (Table 1, Figure 1) [19, 26–33]

Study Characteristics

The included studies consisted of one randomized controlled trials (RCTs) [28], 4 prospective comparative studies [27, 29, 30, 32], 2 validation studies [19, 33] and 2 retrospective studies [26, 31] (Table 1). Geographically, the studies originated primarily from Europe (n=5), Asia (n=1), and America (n=3), with publication dates ranging from 2011 to 2025. Significant heterogeneity existed among studies regarding both the nature of the training interventions and the composition of the trainee populations.

The methodological quality of the included studies was heterogeneous.

According to the RoB 2 assessment (Supplementary Table S1), one randomized trial presented some concerns mainly related to assessor blinding and sample size. Among non-randomized studies, most exhibited moderate-to-serious risk of bias due to confounding and lack of blinding, particularly in retrospective cohorts and educational validations. Validation and feasibility studies were generally at high risk of bias in outcome measurement because of self-reported endpoints and absence of control groups.

Overview of Educational Modalities (Table 2)

1. *Simulation and Virtual/Augmented Reality Training [19, 28, 33]*. Simulation-based interventions consistently improved technical performance and perceived realism.

- In the only randomized trial, Basios et al. [28] reported significantly shorter task completion time (27.3 ± 3.8 vs 35.3 ± 6.5 min, $p < 0.05$) and higher OSATS scores in the VR group compared with textbook learning.
- Feitosa et al. [19] validated a low-cost TOETVA simulator, demonstrating high realism (Likert 4.6/5) and strong content validity.
- Pukas et al. [33] developed an augmented-reality platform allowing enhanced visualization of the recurrent laryngeal nerve, confirming feasibility and usability.
- Wu et al. [29] developed a high-fidelity simulation scenario for the management of postoperative cervical hematoma, a rare but life-threatening complication of thyroidectomy. Ninety-six junior residents participated across multiple boot camp sessions

between 2012 and 2017. Learners rated the simulation highly for realism (4.3/5), communication effectiveness (4.5/5), and debriefing value (4.5/5). Performance improvements were observed in non-technical skills (NOTSS scale, $p < .01$). The scenario provided an effective environment for developing both technical (airway management, hematoma decompression) and cognitive-behavioral skills under crisis conditions.

Overall, simulation and VR/AR environments were found to be effective, safe, and well accepted by trainees, although most studies were small and heterogenous in outcome measures.

2. *Supervised Resident Training and Learning Curve****s [30]*****:**

- Structured supervision models showed measurable learning progression.
- Tarallo et al. [30] prospectively evaluated resident performance across 400 thyroidectomies in two centers. Operative time decreased after 30–50 procedures, with complication rates comparable to senior surgeons.

These findings indicate that stepwise supervision and structured exposure can ensure operative safety while accelerating the learning curve.

3. *Video-Assisted and Multimedia Training* [32]

- Hamour et al. [32] demonstrated that a high-definition video module reduced intraoperative error rates by 49 % and staff takeovers by 52 % among residents. This suggests that video-based learning is an effective adjunct for preoperative skill acquisition and confidence building, particularly when simulation resources are limited.

4. *Intraoperative Neuromonitoring (IONM) Education [27, 31]*

Two studies evaluated structured training in IONM:

- Dionigi et al. [27] reported significant improvement in knowledge and procedural standardization among 75 surgeons attending 17 formal courses.
- Kuryga et al. [31] demonstrated a reduction in recurrent laryngeal nerve (RLN) palsy from 2.8 % to 0.6 % following certified IONM training (RR ≈ 0.21).

These results confirm that formal IONM education can enhance safety and consistency in thyroid surgery.

5. *Fellowship and Advanced Structured Programs* [26]

- In a cohort of 498 patients, Lin et al. [26] found that fellowship-trained surgeons had shorter operative times (mean difference −23 min), lower complication rates (4.5 % vs 8.9 %), and reduced blood loss compared with non-fellowship peers. These data underscore the value of advanced, competency-based fellowships in improving both technical and patient-related outcomes.

Narrative Synthesis of Outcomes

- Across modalities, all studies reported improvements in at least one domain of surgical education:
- Technical performance: enhanced efficiency and skill acquisition (VR/AR, supervised learning).
- Safety: lower complication or RLN injury rates (IONM, fellowship).
- Educational quality: improved knowledge, confidence, and realism (simulation, video-based learning).

However, heterogeneity in study design, assessment tools, and reporting standards precluded quantitative synthesis. Most studies were small, single-center, and varied in outcome definitions, highlighting the need for standardized evaluation frameworks.

DISCUSSION

This systematic review provides a comprehensive synthesis of current evidence regarding simulation and structured training in thyroid surgery. The findings support integrating high-fidelity simulation and curriculum-based training as effective adjuncts to traditional surgical education, with consistent benefits across technical, cognitive, and educational outcomes.

Interpretation of Findings

Across the 10 included studies, simulation-based and structured training programs consistently improved operative performance, technical skills, learner self-confidence, and anatomical knowledge, without evidence of increased complication rates. Simulation-trained cohorts demonstrated shorter operative times and earlier attainment of procedural autonomy, supporting the ability of structured curricula to accelerate the early learning curve. These benefits were particularly relevant for junior surgical trainees (residents, postgraduate years 1–3), who achieved operative benchmarks and confidence more rapidly following repeated, supervised simulation sessions [34].

Technical skills assessed through validated instruments such as OSATS and task-specific checklists showed consistent qualitative improvement among trainees exposed to simulation-based learning. Simulation also played a central role in standardizing assessments and offering targeted remediation opportunities for underperforming trainees—an essential step toward competency-based education [35].

Importantly, the educational benefit extended beyond technical domains. Simulation-based and structured curricula were associated with improved self-reported confidence, greater readiness to participate in live surgeries, and enhanced identification and preservation of critical anatomical structures. This is particularly significant in thyroid surgery, where intraoperative errors may lead to substantial morbidity. Evidence from both prospective and retrospective studies confirmed that simulation provides a safe, reproducible environment in which novice surgeons can progress toward operative autonomy under controlled supervision.

Influence of Simulation Modality and Curriculum Design

Differences in simulation modality influenced learning outcomes. High-fidelity models—cadaveric, porcine, or advanced virtual reality—generally produced greater educational effects than low-fidelity or unsupervised synthetic simulations. Repeated exposure, structured instructor feedback, and objective assessment tools enhanced performance more effectively than single, unstructured sessions. However, these conclusions derive from a limited number of studies, and their generalizability remains uncertain. This methodological limitation should be addressed through future multicenter investigations [34–36].

Implications for Surgical Education

Collectively, the evidence underscores the transformative role of structured educational programs in thyroid surgery training. Simulation enables deliberate practice, accelerates skill acquisition, and safeguards patient outcomes. Supervised resident training, IONM educational programs, and fellowship curricula further demonstrate that structured mentorship and multimodal learning can sustain competency and standardization across diverse healthcare systems.

Simulation-based curricula can also equalize training opportunities, foster accountability, and provide targeted remediation—critical features in specialties characterized by high anatomical variability, steep learning curves, and low tolerance for error, such as thyroid and endocrine surgery [34, 36–38]. Although most included studies originated from high-income settings, several interventions relied on low-cost or modular training solutions, suggesting potential scalability beyond tertiary centers.

Cost Considerations and Resource Allocation

None of the included studies reported standardized cost data, precluding formal cost-effectiveness or budget-impact analyses within this review. Nevertheless, structured simulation-based training may offer indirect institutional value by reducing operative time, complications, and supervision requirements, as suggested in other surgical domains. Future studies should incorporate standardized economic outcomes to support evidence-based educational and investment decisions.

Limitations

While this synthesis includes high-quality, recent studies from multiple regions and training systems, several limitations should be acknowledged. The overall risk-of-bias profile reflects the early developmental stage of structured educational research in thyroid surgery. While individual studies consistently reported positive outcomes, the predominance of small, single-center, non-randomized designs and subjective endpoints limits the certainty of the evidence. Restriction to English-language publications may have led to the exclusion of relevant regional experiences and represents a potential source of selection bias. The only randomized trial [28] supports the feasibility and potential benefit of simulation, but larger multicenter RCTs with blinded assessment and standardized metrics are needed to confirm these findings. Nevertheless, the convergence of results across heterogeneous, moderate-to-high-risk studies strengthens the qualitative signal that structured and multimodal training enhances both technical and educational performance.

The heterogeneity of study designs, training modalities, and outcome measures precluded quantitative pooling of data. Most studies originated from tertiary centers with advanced educational resources, limiting generalizability to smaller or resource-constrained settings. Sample sizes in prospective studies were often small, reducing statistical power. Publication bias and restriction to English-language studies may also have influenced the available evidence. Finally, long-term retention of skills and translation of simulated performance into independent clinical competence were rarely evaluated, highlighting an important direction for future research [38].

Conclusion and Future Perspectives

Simulation and structured curricula represent powerful tools for competency-based education in thyroid surgery. The accumulated evidence indicates that such programs accelerate skill acquisition, improve safety, and help close existing gaps in traditional apprenticeship models. Broader adoption—supported by curriculum standardization, validated assessment tools, and shared outcome measures—will promote greater training equity and consistency across institutions.

Future research should prioritize multicenter, randomized designs, standardized reporting, and longitudinal follow-up to better define the long-term effectiveness, efficiency, and cost-effectiveness of simulation and multimodal training in endocrine surgery.

Funding: None

Declaration of interests: The authors declare that they have no competing interests.

Acknowledgments: We would like to thank the clinicians in our department for their support.

Data Availability. The data supporting the results of this study are available on request from the corresponding author

REFERENCES

1. Randolph GW, Dralle H, with the International Intraoperative Monitoring Study Group, Abdullah H, Barczynski M, Bellantone R, Brauckhoff M, Carnaille B, Cherenko S, Chiang F, Dionigi G, Finck C, Hartl D, Kamani D, Lorenz K, Miccolli P, Mihai R, Miyauchi A, Orloff L, Perrier N, Poveda MD, Romanchishen A, Serpell J, Sitges-Serra A, Sloan T, Van Slycke S, Snyder S, Takami H, Volpi E, Woodson G (2011) Electrophysiologic recurrent laryngeal nerve monitoring during thyroid and parathyroid surgery: International standards guideline statement. The Laryngoscope 121:. https://doi.org/10.1002/lary.21119

2. Bergenfelz A, Jansson S, Kristoffersson A, Mårtensson H, Reihnér E, Wallin G, Lausen I (2008) Complications to thyroid surgery: results as reported in a database from a multicenter audit comprising 3,660 patients. Langenbecks Arch Surg 393:667–673. https://doi.org/10.1007/s00423-008-0366-7

3. Sitges-Serra A, Ruiz S, Girvent M, Manjón H, Dueñas JP, Sancho JJ (2010) Outcome of protracted hypoparathyroidism after total thyroidectomy. Br J Surg 97:1687–1695. https://doi.org/10.1002/bjs.7219

4. Lombardi CP, Raffaelli M, Princi P, Santini S, Boscherini M, De Crea C, Traini E, D’Amore AM, Carrozza C, Zuppi C, Bellantone R (2004) Early prediction of postthyroidectomy hypocalcemia by one single iPTH measurement. Surgery 136:1236–1241. https://doi.org/10.1016/j.surg.2004.06.053

5. Anders Ericsson K (2008) Deliberate Practice and Acquisition of Expert Performance: A General Overview. Acad Emerg Med 15:988–994. https://doi.org/10.1111/j.1553-2712.2008.00227.x

6. Zendejas B, Brydges R, Hamstra SJ, Cook DA (2013) State of the Evidence on Simulation-Based Training for Laparoscopic Surgery: A Systematic Review. Ann Surg 257:586–593. https://doi.org/10.1097/SLA.0b013e318288c40b

7. Stefanidis D, Korndorffer JR, Markley S, Sierra R, Heniford TB, Scott DJ (2007) Closing the Gap in Operative Performance Between Novices and Experts: Does Harder Mean Better for Laparoscopic Simulator Training? J Am Coll Surg 205:307–313. https://doi.org/10.1016/j.jamcollsurg.2007.02.080

8. Palter VN, Grantcharov TP (2010) Simulation in surgical education. Can Med Assoc J 182:1191–1196. https://doi.org/10.1503/cmaj.091743

9. Brucchi F, De Troyer A, Gori A, Dionigi G, Vanderstraeten E, Mottrie A, Van Herzeele I, Rashidian N, Muysoms F (2025) Structured Training in Robotic Abdominal Wall Surgery: A Systematic Review of Educational Models, Methodologies, Existing Gaps and Unmet Needs. J Abdom Wall Surg 4:15190. https://doi.org/10.3389/jaws.2025.15190

10. Brucchi F, Vierstraete M, Vanderstraeten E, Mottrie A, Rashidian N, Muysoms F (2025) Proposal for the Hernia ASCEND Hugo™ RAS training pathway: acquisition of skills by comprehensive exercise-based nimbleness and dexterity training. J Robot Surg 19:516. https://doi.org/10.1007/s11701-025-02704-8

11. Dawe SR, Pena GN, Windsor JA, Broeders JAJL, Cregan PC, Hewett PJ, Maddern GJ (2014) Systematic review of skills transfer after surgical simulation-based training. Br J Surg 101:1063–1076. https://doi.org/10.1002/bjs.9482

12. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB (2011) Does Simulation-Based Medical Education With Deliberate Practice Yield Better Results Than Traditional Clinical Education? A Meta-Analytic Comparative Review of the Evidence: Acad Med 86:706–711. https://doi.org/10.1097/ACM.0b013e318217e119

13. Barry Issenberg S, Mcgaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ (2005) Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 27:10–28. https://doi.org/10.1080/01421590500046924

14. Donatini G, Bakkar S, Leclere FM, Dib W, Suaud S, Oriot D, Breque C, Richer JP, Faure JP, Danion J (2021) SimLife model: introducing a new teaching device in endocrine surgery simulation. Updat Surg 73:289–295. https://doi.org/10.1007/s13304-020-00871-x

15. Graham-Stephenson A, Gabrysz-Forget F, Yarlagadda B (2022) Development of a novel 3D-printed and silicone live-wire model for thyroidectomy. Am J Otolaryngol 43:103410. https://doi.org/10.1016/j.amjoto.2022.103410

16. Gimm O, Barczyński M, Mihai R, Raffaelli M (2019) Training in endocrine surgery. Langenbecks Arch Surg 404:929–944. https://doi.org/10.1007/s00423-019-01828-4

17. Bell RH (2009) Why Johnny cannot operate. Surgery 146:533–542. https://doi.org/10.1016/j.surg.2009.06.044

18. Libutti SK (2010) Endocrine surgery specialty training: Opportunities for growth. Surgery 148:1073–1074. https://doi.org/10.1016/j.surg.2010.09.034

19. Feitosa VN, Cidrão TV, Castro IA, Souza Júnior KCD, Martins SRC, Silva FEPD, Moura Junior LG (2024) Development and validation of the thyroidectomy training box: cervical simulator for training endoscopic transoral thyroidectomy. Acta Cirúrgica Bras 39:e397624. https://doi.org/10.1590/acb397624

20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ n71. https://doi.org/10.1136/bmj.n71

21. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ j4008. https://doi.org/10.1136/bmj.j4008

22. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng H-Y, Corbett MS, Eldridge SM, Emberson JR, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT (2019) RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ l4898. https://doi.org/10.1136/bmj.l4898

23. Wells, G.A., Wells, G., Shea, B., Shea, B., O’Connell, D., Peterson, J., Welch, Losos, M., Tugwell, P., Ga, S.W., Zello, G.A., & Petersen, J.A. (2014) The Newcastle–Ottawa Scale (NOS) for Assessing the Quality of Non-Randomized Studies in Meta-Analysis

24. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336:924–926. https://doi.org/10.1136/bmj.39489.470347.AD

25. Al Asmri M, Haque MS, Parle J (2023) A Modified Medical Education Research Study Quality Instrument (MMERSQI)

developed by Delphi consensus. BMC Med Educ 23:63. https://doi.org/10.1186/s12909-023-04033-6

26. Lin R, Huang S, Guo X, Gao S, Zheng F, Zheng Z (2024) Impact of fellowship training for specialists on thyroidectomy outcomes of patients with thyroid cancer. Sci Rep 14:9033. https://doi.org/10.1038/s41598-024-59864-0

27. Dionigi G, Bacuzzi A, Barczynski M, Biondi A, Boni L, Chiang FY, Dralle H, Randolph GW, Rausei S, Sacco R, Sitges-Serra A (2011) Implementation of systematic neuromonitoring training for thyroid surgery. Updat Surg 63:201–207. https://doi.org/10.1007/s13304-011-0098-z

28. Basios A, Voloudakis N, Atmatzidis S, Velikoudi M, Bellou E, Vamvakidis K, Papaziogas B, Koutelidakis I (2025) The impact of immersive virtual reality training in thyroid surgery: a prospective randomized controlled trial. Updat Surg. https://doi.org/10.1007/s13304-025-02387-8

29. Wu K, Kim S, Rajasingham SJ, Bruni I, Fung K, Roth KE (2019) Simulation of Urgent Airway Management in a Postthyroidectomy Hematoma. MedEdPORTAL 10802. https://doi.org/10.15766/mep_2374-8265.10802

30. Tarallo M, Crocetti D, Gurrado A, Iorio O, Iossa A, Caruso D, Bononi M, Stabilini C, Bracale U, Chiappini A, Testini M, Avenia N, Polistena A, Cavallaro G (2022) Achieving the learning curve in total thyroidectomy: a prospective evaluation on resident's training by CUSUM and KPSS analysis. Ann R Coll Surg Engl 104:414–420. https://doi.org/10.1308/rcsann.2021.0318

31. Kuryga D, Wojskowicz P, Szymczuk J, Wojdyla A, Milewska A, Barczynski M, Dadan J, Rogowski M, Mysliwiec P (2021) Training in intraoperative neuromonitoring of recurrent laryngeal nerves reduces the risk of their injury during thyroid surgery. Arch Med Sci 17:1294–1302. https://doi.org/10.5114/aoms.2019.85737

32. Hamour AF, Mendez AI, Harris JR, Biron VL, Seikaly H, Côté DWJ (2018) A High-Definition Video Teaching Module for Thyroidectomy Surgery. J Surg Educ 75:481–488. https://doi.org/10.1016/j.jsurg.2017.07.019

33. Pukas A, Smal V, Dyvak A, Voytyuk I, Deikalo I, Hrynkiv N (2024) Augmented Reality Simulator for Recurrent Laryngeal Nerve Identification during Thyroid Surgery. In: 2024 14th International

Conference on Advanced Computer Information Technologies (ACIT). IEEE, Ceske Budejovice, Czech Republic, pp 576–580

34. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM (2002) Virtual Reality Training Improves Operating Room Performance: Results of a Randomized, Double-Blinded Study. Ann Surg 236:458–464. https://doi.org/10.1097/00000658-200210000-00008

35. Reznick RK, MacRae H (2006) Teaching Surgical Skills — Changes in the Wind. N Engl J Med 355:2664–2669. https://doi.org/10.1056/NEJMra054785

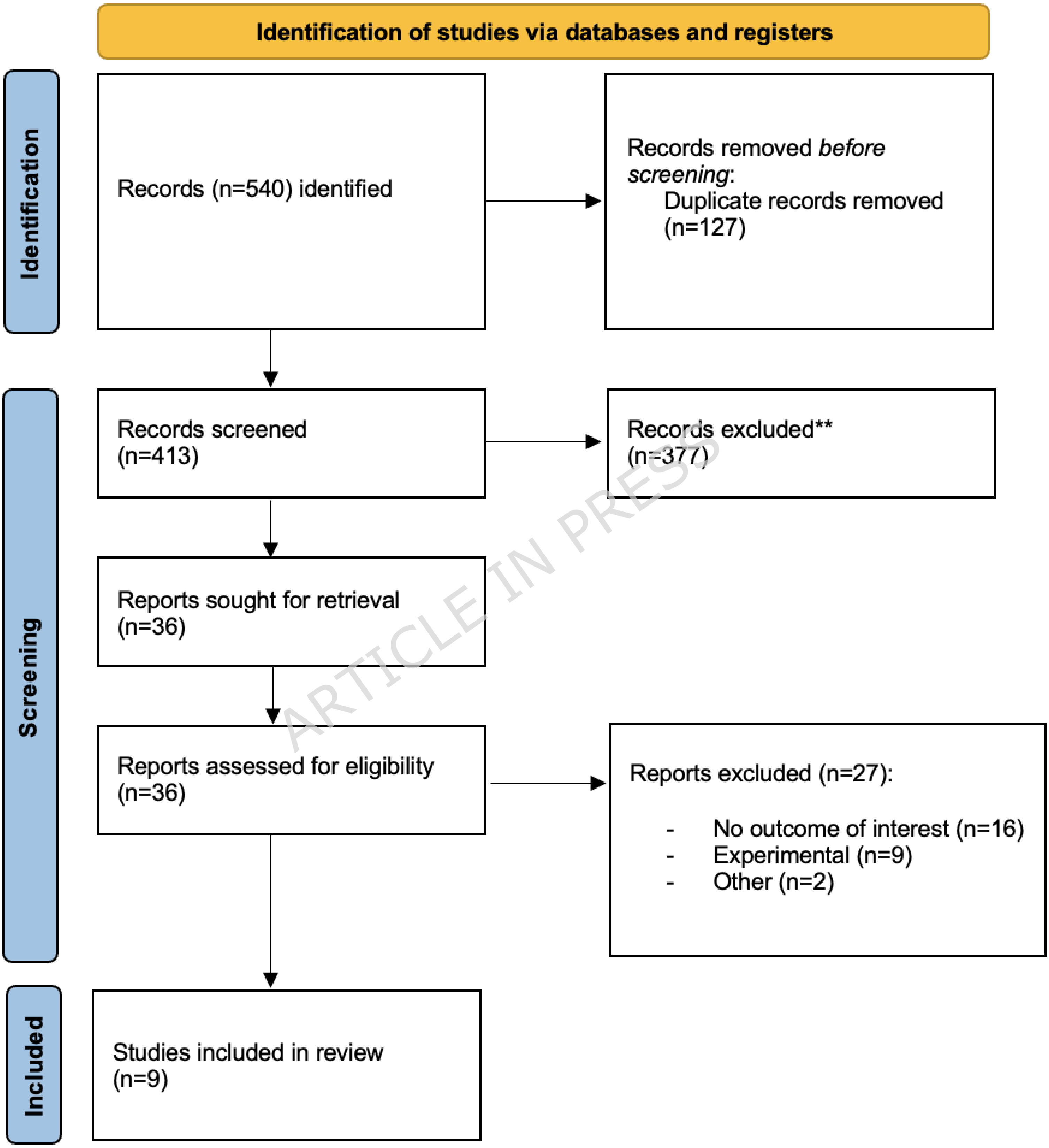
36. Ahmed K, Keeling AN, Fakhry M, Ashrafian H, Aggarwal R, Naughton PA, Darzi A, Cheshire N, Athanasiou T, Hamady M (2010) Role of Virtual Reality Simulation in Teaching and Assessing Technical Skills in Endovascular Intervention. J Vasc Interv Radiol 21:55–66. https://doi.org/10.1016/j.jvir.2009.09.019

37. Van Dongen KW, Tournoij E, Van Der Zee DC, Schijven MP, Broeders IAMJ (2007) Construct validity of the LapSim: Can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc 21:1413–1417. https://doi.org/10.1007/s00464-006-9188-2

38. Norman G, Dore K, Grierson L (2012) The minimal relationship between simulation fidelity and transfer of learning. Med Educ 46:636–647. https://doi.org/10.1111/j.1365-2923.2012.04243.x

FIGURE LEGENDS

Figure 1. A streamlined overview of the study selection process and final categorization is provided, including randomized controlled trials, prospective studies, and systematic reviews. Each stage and inclusion criterion is clearly displayed, representing the analysis of the 540 reviewed papers.


Identification of studies via databases and registers
Identification
Records (n=540) identified
Records removed *before screening*:
Duplicate records removed (n=127)
Screening
Records screened (n=413)
Records excluded** (n=377)
Reports sought for retrieval (n=36)
Reports assessed for eligibility (n=36)
Reports excluded (n=27):
- No outcome of interest (n=16)
- Experimental (n=9)
- Other (n=2)
Included
Studies included in review (n=9)

Characteristics and Main Findings of Included Studies

Table 1. Characteristics of included studies

Author (year)	Country	Design	Participants (n)	Training modality	Comparator	Primary outcomes
Basios et al., 2025	Greece	Randomized controlled trial	16 residents (8 vs 8)	Immersive virtual-reality simulation	Textbook/manual instruction	Task completion time, OSATS, satisfaction
Feitosa et al., 2024	Brazil	Validation study	10 expert evaluators	TOETVA simulation box	—	Realism, ergonomics, educational value
Pukas et al., 2024	Ukraine	Technical validation	Prototype AR simulator	—	Technical feasibility, usability	Augmented-reality simulator for RLN identification
Lin et al., 2024	China	Retrospective cohort	498 patients	Fellowship-trained surgeons	Non-fellowship surgeons	Operative time, blood loss, complications
Kuryga et al., 2021	Poland	Retrospective cohort	2 351 nerves at risk	IONM training and certification	Non-trained surgeons	RLN injury
Tarallo et al., 2021	Italy	Prospective multicenter cohort	400 thyroidectomies	Stepwise resident supervision	Senior surgeons	Operative time, RLN injury, complications
Wu et al., 2019	USA	Prospective simulation validation	96 junior residents	High-fidelity simulation for post-thyroidectomy	Pre-training baseline	Non-technical skills (NOTSS), teamwork,

				hematoma management		realism
Hamour et al., 2018	Canada	Prospective controlled study	6 residents	High-definition video teaching module	Pre-training baseline	Error rate (OCHRA), staff takeovers
Dionigi et al., 2011	Italy	Prospective educational intervention	75 surgeons in 17 courses	Standardized IONM hands-on course	Pre-course baseline	Knowledge, confidence, standardization

Table 2. Main findings and relevance to the current review

Study	Key results	Educational relevance	Implication for present review
Basios et al., 2025	VR group faster (27.3 ± 3.8 vs 35.3 ± 6.5 min; $p < 0.05$) and higher OSATS scores.	Demonstrates efficacy of immersive simulation.	Supports integration of VR-based learning.
Feitosa et al., 2024	High realism (Likert 4.6/5) and strong face/content validity.	Confirms feasibility of TOETVA simulator.	Adds evidence for affordable simulation models.
Pukas et al., 2024	AR simulator feasible with accurate RLN visualization.	Expands training tools with AR applications.	Demonstrates technological innovation.
Lin et al., 2024	Fellowship surgeons had shorter LOS (2.1 vs 3.4 days) and fewer complications (4.5 % vs 8.9 %).	Real-world validation of structured fellowship programs.	Confirms benefit of formal advanced training.
Kuryga et al., 2021	RLN palsy decreased from 2.8 % to 0.6 % after IONM training (RR ≈ 0.21).	Strong evidence of safety improvement via structured training.	Supports clinical impact of IONM education.
Tarallo et al., 2021	Residents approached senior operative times after ~30 cases with similar	Validates supervised stepwise resident training.	Core evidence for structured resident learning.

	complication rates.		
Wu et al., 2019	High-fidelity simulation improved crisis management and teamwork (NOTSS, $p < 0.01$).	Introduces training for emergency complications and non-technical skills.	Broadens educational domains to include safety-critical scenarios.
Hamour et al., 2018	Error rate ↓ 49 %, staff takeovers ↓ 52 % post-video training.	Demonstrates effectiveness of audiovisual learning.	Reinforces preoperative video-based preparation.
Dionigi et al., 2011	90 % of participants improved knowledge; checklist enhanced uniformity.	Early foundation for structured IONM teaching.	Historical basis for training standardization.

Abbreviations: RLN = recurrent laryngeal nerve; LOS = length of stay; OSATS = Objective Structured Assessment of Technical Skills; NOTSS = Non-Technical Skills Scale; IONM = Intraoperative Neuromonitoring.