Extended Reality for Needle-Based Anesthesia Training: A Meta-Analysis of Randomized Controlled Trials

Xin He*
Institute of Medical Technology, Peking University Health Science Center
National Institute of Health Data Science, Peking University

Zhaohui Jiang
Department of Anesthesiology, Beijing Chao-Yang Hospital

Jiawan Wang
Department of Anesthesiology, Beijing Chao-Yang Hospital

Liang Zhou†
Institute of Medical Technology, Peking University Health Science Center
National Institute of Health Data Science, Peking University

Abstract

Needle-based procedures in anesthesiology training (e.g., neuraxial puncture and ultrasound-guided regional techniques) require coordinated visuomotor skills and anatomical reasoning. Compared to the total amount of extended reality (XR) studies on anesthesia training and simulation, validation of the effect on needle-based and neuraxial procedures is far less explored. We report on a meta-analysis to evaluate whether XR-based educational interventions improve needle-based procedural performance relative to non-XR instruction. We systematically identified randomized controlled trials or randomized pilot studies and synthesized two outcomes: (1) procedural skill proficiency, measured by Global Rating Scales (GRS), and (2) procedural efficiency, quantified by time-based metrics. Random-effects meta-analyses (standardized mean difference, SMD) are performed for the seven studies meeting the inclusion criteria (4 for proficiency and 3 for efficiency). XR interventions do not demonstrate a statistically significant improvement (pooled SMD = 0.47, 95% CI [-0.21, 1.15], $p = 0.117$; $I^2 = 0.0\%$) for procedural skill proficiency; whereas for procedural efficiency, there is a trend toward favoring XR, though not reaching statistical significance (pooled SMD = 0.63, 95% CI [-0.01, 1.28], $p = 0.051$; $I^2 = 0.0\%$). Given the small number of samples, mixed intervention designs, and variability in procedural contexts, the evidence supports consideration of the trend toward efficiency gains with caution, whereas consistent proficiency improvements are not demonstrated.

Index Terms: Extended reality, anesthesia education, neuraxial puncture, meta-analysis.

1 Introduction

Anesthesia education and training in general have long faced numerous challenges. For example, the need to cover a wide spectrum of conditions ranging from acute to chronic with diverse subsettings, the operating room is not suitable for teaching, and conducting teaching and practice on patients may compromise patient safety [16]. Among those, needle-based procedures such as neuraxial puncture and ultrasound-guided regional techniques are complex psychomotor tasks that integrate anatomical knowledge, spatial reasoning, and fine control. Early skill acquisition can be stressful for novices and may be associated with inefficient performance in real clinical settings. Traditional training models, often presented as an apprenticeship approach, have inherent limitations: the unpredictability of clinical opportunities, ethical constraints associated with practicing invasive procedures on patients, and the difficulty of teaching complex spatial anatomy without risk. To this end, the use of simulation-based education to support competency development is motivated [10, 5].

XR-based training can potentially address aforementioned issues by offering safe and repeatable environments for deliberate practice [11]. While XR is widely explored in many sub-domains in anesthesia such as dental education [18, 17, 24], airway management [23], and crisis simulation [1], its application in needle-based regional and neuraxial anesthesia, e.g., spinal, epidural, and nerve blocks, is understudied due to unique challenges. Unlike visual-heavy procedures, these interventions often involve blind needle placement, requiring the practitioner to construct a complex mental model of internal 3D anatomy and rely on subtle tactile cues. Although several empirical studies have demonstrated well-designed AR/MR-based simulation systems and evaluated XR techniques for these specific tasks, results regarding their superiority over traditional methods remain mixed [25, 21, 3, 2, 15].

To better understand the effectiveness of XR in neuraxial and regional anesthesia training, we conducted a focused meta-analysis of randomized studies of needle-based anesthesia procedures in this paper. We aim to quantify the impact of XR training on two primary outcomes: skill proficiency and procedural efficiency to provide evidence-based recommendations for future simulator design. Additionally, we provide suggestions for future study designs.

2 Methods

A systematic literature search was conducted with a set of selection criteria formed by the authors, comprised of anesthesiologists and XR experts. A meta-analysis was then performed with the extracted data from the included studies.

2.1 Search Strategy and Selection Criteria

The following expression was used for a search in the PubMed database:
(Virtual Reality OR Augmented Reality OR Mixed Reality) AND (Epidural OR Spinal Anesthesia OR Nerve Block OR Lumbar Puncture) AND (Training OR Simulation OR Education) AND (Randomized Controlled Trial OR RCT).

Inclusion Criteria Studies were selected based on the Population, Intervention, Comparison, Outcomes and Study designs (PICOS) criteria:

- **Participants:** Anesthesiologists, residents, medical students, or nurses undergoing technical skills training.
- **Intervention:** XR technologies, including VR, AR, and MR simulators and systems.
- **Comparison:** Traditional training methods (e.g., video or textbooks).
- **Outcomes:** Objective quantitative metrics, specifically skill proficiency (Global Rating Scale) and procedural efficiency (quantified time measurements).
- **Study Design:** Randomized controlled trials and randomized pilot studies.

*e-mail: hexin@stu.pku.edu.cn
†e-mail: zhoulng@pku.edu.cn. L. Zhou is the corresponding author.

Table 1: Characteristics of Included Studies

Study	Year	*N*	Type	Device&Platform	Haptics	Control Group	Outcome Subgroup
Kulcsar et al. [14]	2013	23	VR	SenseGraphic Workbench + Phantom Desktop	✓	Conventional teaching	Skill proficiency
OSullivan et al. [19]	2014	8	VR	Phantom Desktop	✓	No further simulation training	Skill proficiency
Kim et al. [13]	2023	20	VR	Oculus Quest 2	✓	Video, text, self-learning	Skill proficiency
Li et al. [15]	2024	21	VR	Oculus Quest 2 + 3D Organon Anatomy®		2D image anatomy learning	Skill proficiency
Keri et al. [12]	2015	24	AR	Perk Tutor		Conventional method only	Procedural efficiency
Gao et al. [7]	2024	12	MR	Microsoft HoloLens 2		Landmark-guided puncture	Procedural efficiency
Felten et al. [6]	2025	53	AR	Microsoft HoloLens 2 + Touch™	✓	Theory + bedside training	Procedural efficiency

Note. N denotes the effective sample size for the corresponding outcome measures.

Exclusion Criteria For homogeneity, we strictly excluded:

- studies focusing on dental or maxillofacial anesthesia and other fields unrelated to puncture or needling procedure (e.g., anxiety management and sedation);
- interventions related to general anesthesia or airway management;
- studies utilizing non-immersive mobile AR applications (e.g., AR books) ;
- studies measuring only subjective outcomes (e.g., satisfaction, confidence).

2.2 Data Extraction

We extracted data focusing on two key subgroups: Skill Proficiency (higher scores mean better performance) and Procedural Efficiency (shorter times reflect greater efficiency).

Outcome Selection Although the included trials used different GRS instruments or equivalent skill scores, we treated them as continuous measures of overall procedural performance and synthesized them using standardized mean differences (SMD; Hedges' *g*). When multiple skill scores were reported in a single study, we selected the outcome measure based on a larger sample size and assessed at a time point closer to the completion of training.

Special Cases For crossover trials with repeated measures, we used the number of participants as the sample size. For studies reporting medians and interquartile ranges, we estimated means and standard deviations using Wan et al. [22].

Statistical Analysis We performed random-effects meta-analyses using the DerSimonian–Laird estimator [4]. Effect sizes were calculated as Hedges' *g* with small-sample correction [8], and confidence intervals and *p*-values were calculated using the *t*-distribution due to the limited number of studies. All analyses were conducted in Python (v3.13) using `statsmodels` and `matplotlib`.

3 Results

Our analysis is reported with a summary of characteristics of the studies, and detailed meta-analysis results for the two subgroups, respectively.

3.1 Study Characteristics

A total of 7 randomized studies were included. Key study characteristics, including participant level, task type, XR modality, comparator, and outcome reporting, are summarized in Table 1.

All studies were conducted in clinical anesthesia training contexts with trainee populations (e.g., anesthesia novices/residents) and targeting neuraxial and/or regional anesthesia–related needle procedures. XR interventions (VR/AR/MR systems) with varying levels of interaction fidelity (e.g., tracked instruments and/or haptic interfaces in some systems), while comparators included standard instruction and conventional simulation/teaching approaches.

3.2 Skill Proficiency

Four trials contributed proficiency outcomes with GRS results. As shown in the forest plot in Figure 1, the random-effects meta-analysis yielded a pooled effect size of SMD = 0.47 with the 95% CI of $[-0.21, 1.15]$ ($p = 0.117$). Heterogeneity was estimated as $I^2 = 0\%$, indicating no detectable between-study inconsistency in effect direction; however, given the small number of studies, heterogeneity estimates should be interpreted cautiously. Overall, the pooled estimate favors XR, but the confidence interval crosses zero, suggesting that current randomized evidence does not establish a clear improvement in proficiency scores.

Multiverse analysis across all 8 possible outcome-selection combinations yielded consistently positive pooled effects (SMD range 0.30–0.73), but confidence intervals crossed zero in all cases ($p =$ 0.054–0.429). Regardless of which GRS outcome is chosen, the effect did not reach statistical significance. Further details are provided in Supplementary Figure S1.

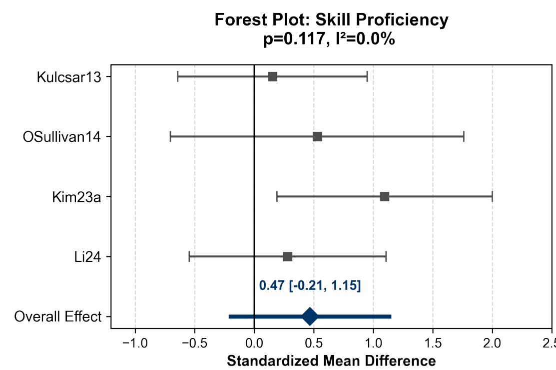

Figure 1: The Forest Plot of Skill Proficiency

3.3 Procedural Efficiency

Three trials contributed time-based outcomes. As shown in Figure 2, the pooled random-effects estimate was SMD = 0.63 with 95% CI being $[-0.01, 1.28]$ ($p = 0.051$). The SMD result suggests that XR training may reduce procedural completion time, corresponding to roughly a 67% probability that a randomly selected XR-trained participant completes the procedure faster than a randomly selected control participant. We back-translated the pooled SMD to each study's original time scale, yielding illustrative reductions of approximately 38 s, 101 s, and 316 s (about 5.3 min) for the three trials, respectively (Keri15; Gao24; Felten25). Note that these back-translations use each study's pooled SD and are intended as illustrative magnitudes; the actual means and SDs were estimated from reported medians and IQRs where raw data were unavailable. Given the wide 95% CI, the true effect could range from negligible to substantial.

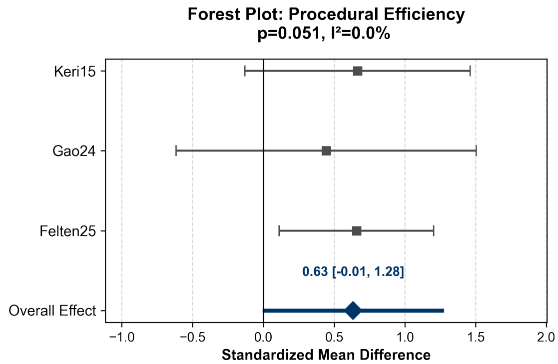

Figure 2: The Forest Plot of Procedural Efficiency

4 Discussion

We discuss potential explanations of the meta-analysis results, implications for future work, and limitations of this work.

4.1 Explanations

The meta-analysis result suggests that XR systems for needle-based procedures education showed more consistent signals for efficiency than for global proficiency.

As for the time efficiency, a possible mechanistic explanation is that visuospatial coordination enables rapid coordination and coupled responses among the hand, eye, and brain. Keri suggested that overlaying ultrasound images with augmented-reality information helps trainees establish the spatial relationship of the needle tip relative to the ultrasound plane and the anatomical structures, thereby making puncture performance faster and more stable [12]. Gao suggested that mixed-reality instruction allows trainees to obtain a three-dimensional understanding of the entry point, direction, depth, and trajectory [7]. At the subjective level, augmented-reality simulation teaching can provide immediate feedback and error correction, which enhances students' sense of composure during execution and may in turn influence efficiency [6]. Another possible reason is that XR provides an environment in which trainees can practice repeatedly without consequences, which leads to faster performance in clinical settings.

The proficiency result is somewhat below our expectations, and this can be partly explained by the small sample size and the short VR intervention exposure. But the root reason may stem from the inherent nature of neuraxial and needle-based anesthesia procedures that can be divided into the identification phase and the execution phase. The identification phase is where traditional methods, such as palpation and landmark-based techniques, already perform reasonably well for many patients with normal anatomy. However, even with perfect identification, subsequent needle placement requires additional skills that XR may not address as effectively [20]. Once the target location is identified, executing a precise needle trajectory demands complex hand-eye coordination and proprioceptive control. The inherent complexity of this motor task means that while XR can simulate the visual aspects, it struggles to fully replicate the tactile and proprioceptive feedback essential for mastery. An advanced visualization method for needle guidance on physical phantoms is available [9] to partly address the issue, but it does not support the fine anatomical structures encountered in anesthesia. Several included trials also highlight limitations in simulator fidelity and representation, and O'Sullivan even discontinued recruitment because the virtual sono-anatomy was overly simplified and lacked key structures, raising concern for negative learning [19, 13].

4.2 Implications

Across current evidence, XR training showed clearer improvements in time than in global proficiency scores. This may indicate that the enhancement of XR technology in visuospatial perception and comprehension is more pronounced from a temporal perspective, whereas its overall effectiveness is constrained by a variety of factors. Admittedly, the effects of the gains in spatial perceptual coordination afforded by XR cannot be strictly disentangled from additional learning time at present [19, 6].

Another advantage of XR is scalability, which can be helpful in public health emergencies and other resource-limited scenarios along with its potential in speed raising. Platforms that provide repeated practice with reproducible setup and rapid preparation may also ease teaching logistics in simulation centers [12]. However, it is worth noting that speed should not be treated as a primary educational target.

Several trials explicitly note technology limitations that can restrain proficiency learning, including imperfect haptics, limited realism, reduced imaging resolution, and a risk of misleading anatomical learning when simulators are simplified or inaccurate [19, 13]. Short intervention duration, small samples, and rater dependent scoring further reduce sensitivity for broad GRS style outcomes [14, 15]. Future work should therefore focus on improving anatomical and instrument fidelity and on using assessments that separate identification from execution, combining targeted rating scales with objective motion or trajectory metrics [12, 6]. For anatomy identification part, the incremental benefit of XR over conventional approaches may be limited, as many learners can reach adequate performance with landmark-based teaching in typical anatomy. Larger gains may be more likely in difficult cases, where patient-specific guidance has been proposed and tested as a clinically oriented direction for subsequent trials [7]. In this context, XR could serve as an effective medium for sharing rare-case datasets and supporting scalable anesthesia education across sites. Also, from the methodology perspective, adequately powered trials with blinded or video blinded assessment, equal training time controls, and clinical transfer and retention endpoints are needed.

4.3 Limitations

First, the lack of standardized frameworks for XR-based training interventions and the inherent differences among distinct XR technologies pose significant constraints. The included studies employed heterogeneous XR modalities with varying interaction fidelity, haptic feedback configurations, and training durations—for instance, some used immersive VR headsets while others adopted AR overlays or MR devices, with 4 systems integrating tactile feedback. Treating XR as a monolithic intervention obscures the specific technical features that may drive training outcomes, hindering meaningful comparisons of effectiveness and limiting the generalizability of our findings.

Second, this meta-analysis is constrained by small sample sizes, heterogeneous interventions and procedures, and inconsistent outcome reporting, which together limit precision and generalizability. Some included studies occur in simulated settings, while others involve clinical performance by novices under supervision; these contexts may yield different sensitivity to XR-based training. Finally, publication bias cannot be excluded.

Third, heterogeneity in outcome definitions exists, particularly in the Procedural Efficiency subgroup, where metrics ranged from specific needle-insertion time to total procedure duration. While standardized effect sizes were used to mitigate scale differences, future research should adopt standardized reporting guidelines (e.g., distinguishing technical performance time from setup time) to allow for more granular analyses. The GRS measurements of the included articles were also not strictly consistent.

5 Conclusion

Evidence of randomized controlled studies on XR for needling procedure in anesthesia education suggests potential time efficiency benefits (shorter completion time) but no clear, consistent improvement in global proficiency as measured by GRS across different interventions. This highlights the strengths of XR for visuospatial coordination enhancement and low-cost, low-risk simulation, while also pointing to possible limitations in mechanical feedback and high fidelity simulation that achieve clinical requirements. Future XR training should prioritize high-fidelity anatomical modeling, pair it with anesthesia-specific sensory support, and data-driven fine-grained evaluation. XR's potential in complex cases is also a promising yet underexplored direction. Larger RCTs with more rigorous controls, broader evidence synthesis and standardized, validated evaluation scales will be essential to clarify XR's real-world impact on needle-based anesthesia training.

References

[1] L. W. Babus, H. Gurnaney, A. K. Doshi, H. Liu, A. Nishisaki, D. Singh, R. J. Daly Guris, and CHOP Virtual Reality Group. The utility of virtual reality and manikin crisis scenario simulations for anaesthesia trainee education: A randomised crossover pilot study. *Anaesthesia Reports*, 12(2):e12316, 2024. doi: 10.1002/anr3.12316 1

[2] A. Chuan, A. Bogdanovych, B. Moran, S. Chowdhury, Y. C. Lim, M. T. Tran, T. Y. Lee, J. Duong, J. Qian, T. Bui, A. M. H. Chua, B. Jeyaratnam, S. Siu, C. Tiong, M. McKendrick, and G. A. McLeod. Using Virtual Reality to teach ultrasound-guided needling skills for regional anaesthesia: A randomised controlled trial. *Journal of Clinical Anesthesia*, 97:111535, Oct. 2024. doi: 10.1016/j.jclinane.2024.111535 1

[3] A. Chuan, J. Qian, A. Bogdanovych, A. Kumar, M. McKendrick, and G. McLeod. Design and validation of a virtual reality trainer for ultrasound-guided regional anaesthesia. *Anaesthesia*, 78(6):739–746, June 2023. doi: 10.1111/anae.16015 1

[4] R. DerSimonian and N. Laird. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3):177–188, Sept. 1986. doi: 10.1016/0197-2456(86)90046-2 2

[5] C. Elendu, D. C. Amaechi, A. U. Okatta, E. C. Amaechi, T. C. Elendu, C. P. Ezeh, and I. D. Elendu. The impact of simulation-based training in medical education: A review. *Medicine*, 103(27):e38813, July 2024. doi: 10.1097/MD.0000000000038813 1

[6] R. Felten, K. Bigaut, T. Wirth, L. Kremer, L. Gauer, C. Arnold, I. Ollivier, J. Godet, M. Scherlinger, M. Dubois, E. Sebbag, J. De Sèze, and J.-E. Gottenberg. Advancing medical training with augmented reality and haptic feedback simulator: Outcomes of a randomized controlled trial on lumbar puncture. *BMC medical education*, 25(1):1231, Aug. 2025. doi: 10.1186/s12909-025-07536-6 2, 3

[7] L. Gao, Y. Xu, X. Zhang, Z. Jiang, J. Wu, Y. Dong, M. Li, L. Jin, J. Qiu, L. You, C. Qin, and W. Gu. Comparison of Mixed Reality-Assisted Spinal Puncture with Landmark-Guided Spinal Puncture by Novice Practitioners: A Pilot Study. *Journal of Pain Research*, 17:2701–2712, 2024. doi: 10.2147/JPR.S470285 2, 3

[8] L. V. Hedges. Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics*, 6(2):107–128, 1981. doi: 10.2307/1164588 2

[9] F. Heinrich, F. Joeres, K. Lawonn, and C. Hansen. Comparison of projective augmented reality concepts to support medical needle insertion. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2157–2167, 2019. doi: 10.1109/TVCG.2019.2903942 3

[10] M. J. V. Henriksen, T. Wienecke, J. Kristiansen, Y. S. Park, C. Ringsted, and L. Konge. Opinion and Special Articles: Stress when performing the first lumbar puncture may compromise patient safety. *Neurology*, 90(21):981–987, May 2018. doi: 10.1212/WNL.0000000000005556 1

[11] K. Hoek, C. Martini, M. van Velzen, and E. Sarton. From Theory to Practice: Simulation in Anesthesiology Through Active Learning Experiences. In D. Singh, J.-W. van 't Klooster, and U. S. Tiwary, eds., *Intelligent Human Computer Interaction*, pp. 29–40. Springer Nature Switzerland, Cham, 2025. doi: 10.1007/978-3-031-88705-5_3 1

[12] Z. Keri, D. Sydor, T. Ungi, M. S. Holden, R. McGraw, P. Mousavi, D. P. Borschneck, G. Fichtinger, and M. Jaeger. Computerized training system for ultrasound-guided lumbar puncture on abnormal spine models: A randomized controlled trial. *Canadian Journal of Anaesthesia*, 62(7):777–784, July 2015. doi: 10.1007/s12630-015-0367-2 2, 3

[13] J. Y. Kim, J. S. Lee, J. H. Lee, Y. S. Park, J. Cho, and J. C. Koh. Virtual reality simulator's effectiveness on the spine procedure education for trainee: A randomized controlled trial. *Korean Journal of Anesthesiology*, 76(3):213–226, June 2023. doi: 10.4097/kja.22491 2, 3

[14] Z. Kulcsár, E. O'Mahony, E. Lövquist, A. Aboulafia, D. Sabova, K. Ghori, G. Iohom, and G. Shorten. Preliminary evaluation of a virtual reality-based simulator for learning spinal anesthesia. *Journal of Clinical Anesthesia*, 25(2):98–105, Mar. 2013. doi: 10.1016/j.jclinane.2012.06.015 2, 3

[15] X. Li, S. Ye, Q. Shen, E. Liu, X. An, J. Qin, Y. Liu, X. Xing, J. Chen, and B. Lu. Evaluating virtual reality anatomy training for novice anesthesiologists in performing ultrasound-guided brachial plexus blocks: A pilot study. *BMC anesthesiology*, 24(1):474, Dec. 2024. doi: 10.1186/s12871-024-02865-3 1, 2, 3

[16] S. D. Marshall and M. C. Turner. Anesthesia Education: Trends and Context. In *Clinical Education for the Health Professions*, pp. 69–85. Springer, Singapore, 2023. doi: 10.1007/978-981-15-3344-0_8 1

[17] R. Mladenovic, D. Dakovic, L. Pereira, V. Matvijenko, and K. Mladenovic. Effect of augmented reality simulation on administration of local anaesthesia in paediatric patients. *European Journal of Dental Education: Official Journal of the Association for Dental Education in Europe*, 24(3):507–512, Aug. 2020. doi: 10.1111/eje.12529 1

[18] R. Mladenovic, L. a. P. Pereira, K. Mladenovic, N. Videnovic, Z. Bukumiric, and J. Mladenovic. Effectiveness of Augmented Reality Mobile Simulator in Teaching Local Anesthesia of Inferior Alveolar Nerve Block. *Journal of Dental Education*, 83(4):423–428, Apr. 2019. doi: 10.21815/JDE.019.050 1

[19] O. O'Sullivan, G. Iohom, B. D. O'Donnell, and G. D. Shorten. The effect of simulation-based training on initial performance of ultrasound-guided axillary brachial plexus blockade in a clinical setting - a pilot study. *BMC anesthesiology*, 14:110, 2014. doi: 10.1186/1471-2253-14-110 2, 3

[20] A. Sadeghi, R. Patel, and J. Carvalho. Ultrasound-facilitated neuraxial anaesthesia in obstetrics. *BJA Education*, 21(10):369–375, Oct. 2021. doi: 10.1016/j.bjae.2021.06.003 3

[21] S. P. Shevlin, L. Turbitt, D. Burckett-St.Laurent, A. J. Macfarlane, S. West, and J. S. Bowness. Augmented Reality in Ultrasound-Guided Regional Anaesthesia: An Exploratory Study on Models With Potential Implications for Training. *Cureus*, 15(7):e42346, 2023. doi: 10.7759/cureus.42346 1

[22] X. Wan, W. Wang, J. Liu, and T. Tong. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC medical research methodology*, 14:135, Dec. 2014. doi: 10.1186/1471-2288-14-135 2

[23] D. T. Wong, A. Mehta, K. P. Singh, S. M. Leong, A. Ooi, A. Niazi, E. You-Ten, A. Okrainec, R. Patel, M. Singh, and J. Wong. The effect of virtual reality bronchoscopy simulator training on performance of bronchoscopic-guided intubation in patients: A randomised controlled trial. *European Journal of Anaesthesiology*, 36(3):227–233, Mar. 2019. doi: 10.1097/EJA.0000000000000890 1

[24] A. Yari, P. Fasih, A. Goodarzi, A. Nouralishahi, and D. Nikeghbal. The effect of augmented reality book on the proficiency of local anesthesia administration of the inferior alveolar nerve. *Journal of Dental Education*, 88(7):1000–1008, July 2024. doi: 10.1002/jdd.13522 1

[25] T. Zheng, H. Xie, F. Gao, C. Gong, W. Lin, P. Ye, Y. Liu, B. He, and X. Zheng. Research and application of a teaching platform for combined spinal-epidural anesthesia based on virtual reality and haptic feedback technology. *BMC Medical Education*, 23:794, Oct. 2023. doi: 10.1186/s12909-023-04758-4 1