RESEARCH

Open Access

Check for updates

Image-based and textbook-based virtual reality training on operational skills among junior residents: a proof of concept study

Wan-Ni Lin^{1,2†}, Hai-Hua Chuang^{3,4†}, Yi-Ping Chao^{1,5}, Li-Jen Hsin^{1,2}, Chung-Jan Kang^{1,2}, Tuan-Jen Fang^{1,2}, Hsueh-Yu Li^{1,2} and Li-Ang Lee^{1,2,4*}

Abstract

Background Operational training is a key component of resident education. Recently, innovative virtual reality (VR) training methods have been introduced to enhance training efficiency. Image-based VR (IBVR), which incorporates cognitive load, is theorized to improve task performance. However, the impact of IBVR on learning outcomes requires further investigation. This study aims to assess the efficacy of IBVR compared to textbook-based VR (TBVR) in teaching operational skills to junior residents.

Methods In a prospective cross-over pilot study, ten volunteers were randomly assigned to either the IBVR-TBVR or TBVR-IBVR group. Participants engaged in four learning sessions using either IBVR or TBVR modules during the first phase. Performance was assessed using quizzes, and Milestone/Direct Observation of Procedural Skills (DOPS) ratings on real patients. After one month, participants switched to the alternate VR module for further training. Cognitive load and stress were assessed during each session through questionnaires and heart rate variability (HRV). At the end of the study, learning satisfaction, experience, and overall effectiveness were evaluated using a global satisfaction scale, the AttrakDiff2 questionnaire, and group interviews. Qualitative data were analyzed using a thematic analysis framework.

Results The IBVR module yielded significantly better Milestone (p = 0.04), and DOPS (p < 0.01) scores compared to TBVR. There were no significant differences in knowledge gain, cognitive load, or HRV between the two modules. TBVR was favored in terms of global satisfaction (p = 0.03), hedonic stimulation (p = 0.01), and hedonic identification (p = 0.03), whereas IBVR was perceived as a more immersive and enriching experience. The majority (70%) of participants reported a positive experience with IBVR, while 50% expressed positive feedback regarding TBVR. Thematic analysis identified two key themes: usability of instructional content and ease of engagement.

Conclusion Although TBVR yielded higher learner satisfaction and hedonic appeal, IBVR resulted in greater improvements in operational performance and was positively received by most participants. This proof-of-concept

[†]Wan-Ni Lin and Hai-Hua Chuang contributed equally to this work.

*Correspondence: Li-Ang Lee 5738@cgmh.org.tw

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

study highlights the complementary strengths of both VR approaches and calls for further research to validate these preliminary findings and inform the design of effective VR-based surgical education strategies.

Trial registration Clinicaltrials.gov NCT03501641; https://clinicaltrials.gov/ct2/show/NCT03501641; date of registration: April 18, 2018.

Keywords Cognitive load, Direct observation of procedural skills, Heart rate variability, Milestone, Qualitative feedback, Technology acceptance, Virtual reality

Background

Surgery integrates knowledge, skills, behavior, and patient care, relying on physical interventions involving tissues or organ systems for diagnostic or therapeutic purposes. Key components of surgical practice include the coordination of staff, instruments, space, and systems for delivering effective care. The Surgical, Anaesthetic, and Obstetric healthcare system assesses global access to surgical services based on factors such as timeliness, capacity, safety, and affordability. It is estimated that around 5 billion people lack access to essential surgical care worldwide [1], with this disparity persisting, particularly in low- and middle-income countries as of 2020 [2].

Otorhinolaryngology-head and neck surgery (ORL-HNS) involves the surgical and medical management of conditions affecting the head and neck, representing a critical clinical competency within the field [3]. Improving training in geographic accessibility to surgical providers and procedural outcomes could significantly impact global ORL-HNS initiatives [4]. ORL-HNS encompasses a broad range of operative skills, including microscopic and endoscopic ear surgeries [5], image-guided sinus surgeries [5], robotic surgeries [6], and intraoperative nerve monitoring during parotidectomy [7]. While training programs, such as those for robotic surgery developed in 2012, have enhanced technical skills, they still face challenges like limited access to advanced technologies, potentially hindering the development of basic surgical skills [8, 9].

To address these challenges, simulation-based training methods, including part-time trainers, integrated simulators, and virtual reality (VR), are increasingly being adopted in competency-based medical education. VR offers an innovative and promising approach to creating immersive, emotionally engaging educational experiences that simulate medical examinations, procedures, and surgeries with varying degrees of interactivity and realism [10]. This technology allows for repeated practice while accommodating different learning styles, potentially reducing cognitive load and improving the retention of surgical skills [11].

Three distinct types of VR are commonly recognized: non-immersive, semi-immersive, and fully immersive. Non-immersive VR, such as screen-based VR, involves interaction through a computer or television screen, limiting the level of immersion [12]. Semi-immersive VR provides a more engaging experience using large projection screens or head-mounted displays (HMDs), facilitating deeper interaction with the virtual world [13]. Fully immersive VR leverages advanced HMDs and often incorporates auditory and haptic feedback, creating a highly realistic and immersive training environment [14].

Textbook-based VR (TBVR), a semi-immersive modality, enables learners to interact with textual and pictorial instructional content displayed on two-dimensional (2D) panels within a three-dimensional (3D) virtual environment. This approach uses hand controllers to turn virtual book pages, allowing structured and focused engagement with multimedia learning materials [15, 16]. The structured presentation of content in TBVR offers a cognitively manageable and pedagogically sound approach, which has been shown to help novice learners acquire procedural knowledge and spatial understanding in clinical education [17]. By integrating interactive elements with step-by-step, multimedia-enhanced instruction [18], TBVR has the potential to bridge the gap between traditional textbook-based education and hands-on intraoperative training, supporting the acquisition of foundational procedural knowledge and skills during early surgical education.

Recent advancements in image-based VR (IBVR), such as 360° videos, represent a form of immersive VR that has shown significant potential in improving learning outcomes and learner satisfaction. IBVR replicates real-world scenarios through high-fidelity immersive experiences, improving spatial reasoning and procedural memory in surgical training [19–21]. However, the effectiveness of IBVR varies depending on factors such as cognitive load, prior experience, and self-efficacy [22–24]. Despite these variations, research suggests that the virtual presence provided by VR generally improves learning outcomes compared to traditional desktop-based methods, regardless of the technology's cost or level of immersion [25].

While these findings are promising, the impact of instructional design strategies using high-cost VR systems without simulator equipment on user experience, usability, and performance in ORL-HNS surgical training remains underexplored. TBVR is valued for its simplicity and structured instructional design, which supports foundational learning, while IBVR excels in enhancing learner engagement and self-efficacy in complex surgical scenarios. However, the differences in surgical learning outcomes between semi-immersive and fully immersive VR modalities are not fully understood.

Existing literature has not adequately examined the impact of VR modalities (IBVR vs. TBVR) and training sequences on critical outcomes such as operational competency, cognitive load, and learner satisfaction. Additionally, the effects of these approaches on physiological metrics like heart rate variability (HRV) and their potential role in managing stress during training remain unexplored.

This proof-of-concept study evaluates the effectiveness of immersive IBVR training in enhancing operational skills and learner experience during representative ORL-HNS preoperative preparations and basic surgeries compared to semi-immersive TBVR training. It also investigates the influence of VR course design on training outcomes, offering insights for optimizing VR-based surgical education.

The research questions for this study include:

- 1. How does immersive IBVR training compare to semi-immersive TBVR training in enhancing operational skills among junior residents?
- 2. Does the sequence of VR training (TBVR followed by IBVR versus IBVR followed by TBVR) affect learning outcomes, cognitive load, and user satisfaction?
- 3. What are the specific impacts of IBVR and TBVR on HRV metrics, stress levels, and the overall learning experience?

Methods

Study design

We conducted a randomized, controlled, cross-over pilot study to validate basic VR-based surgical training among convenience-sampled junior residents. This design helped increase the sample size and reduce the risk of inadequate training. The study was carried out from August 1, 2019, to July 31, 2020, at Linkou Chang Gung Memorial Hospital, a tertiary medical center in Taoyuan, Taiwan. It was approved by the Institutional Review Board of Chang Gung Medical Foundation (No: 201601821B0), and all procedures involving human participants were conducted in accordance with the ethical standards of the institutional and/or national research committee and with the 1975 Helsinki Declaration [26] and the CONSORT guidelines [27]. All participants provided written informed consent. The clinical trial is registered with ClinicalTrials.gov (NCT03501641) and can be accessed at http://clinicaltrials.gov/show/NCT03501641. Figure 1 illustrates the study flowchart.

Participants

We recruited ten junior residents who were novices to the targeted surgeries, each having participated in fewer than ten instances of the specific procedures. Eligible participants were over 20 years of age and held the status of junior resident (R1 or R2). Exclusion criteria included pregnancy, hypertension, recent motion sickness, inner ear infections, claustrophobia, recent surgery, pre-existing binocular vision abnormalities, heart conditions, epileptic symptoms, or unwillingness to participate. Additionally, we evaluated the cognitive style of each participant using the group embedded figures test (GEFT) [28]. The GEFT categorizes cognitive styles into "fielddependent" (GEFT score≤12) and "field-independent" (GEFT score > 12) [29]. Given the significant interaction between cognitive style (field-dependence versus fieldindependence) and the training/testing environment (VR versus real) on learning outcomes [30], we ensured a balanced representation of cognitive styles in this study.

Setting

The surgical training program was developed with two key learning objectives: (1) "Residents should be able to precisely and proficiently prepare for complex surgeries pre-operatively" and (2) "Learners should be able to precisely and proficiently perform basic surgeries." The training focused on three complex procedures for ORL-HNS patients and one fundamental procedure—ventilation tube placement—suitable for junior residents.

Participants were randomly assigned to either the immersive IBVR or semi-immersive TBVR group and received training in four different procedures (Fig. 2). HRV was monitored during each VR session. After completing all four sessions, participants immediately filled out cognitive load questionnaires, including the Paas cognitive load scale (CLS) [31], the NASA task load index (TLX) [32], and the cognitive load component (CLC) [33].

The first phase of training spanned one month, during which participants used VR modules for 10 min, followed by 10 to 15 min performing the corresponding procedures or preparing for operations in real environments.

Two evaluators, blinded to the training module each participant received, assessed participants' precision and proficiency using the Milestone instruments [3] and the direct observation of procedural skills (DOPS) [34], with patient safety being the top priority throughout.

After completing the procedural learnings and assessments, participants received immediate bidirectional feedback from evaluators and reflected on their performance. To minimize the risk of overlapping training effects, a wash-out period of at least one month was implemented before the second training phase. During



Fig. 1 CONSORT flowchart illustrating participant recruitment, randomization, and study completion



Fig. 2 Virtual reality (VR) sessions in two training modules including image-based VR (IBVR) (left panels) and textbook-based VR (TBVR) (right panels)

the second phase, participants switched to the alternate VR module for further training.

At the end of this study, participants rated their global satisfaction score (GSS) [35] and the AttrakDiff2 questionnaire [36]. The study concluded with group interviews where participants provided detailed feedback and reflections, which were qualitatively analyzed to assess the overall effectiveness of the IBVR and TBVR surgical training approaches.

Surgical training program overview

The training program provided comprehensive instruction on key surgical procedures:

- 1. Navigation for image-guided endoscopic sinus surgery.
- 2. The use of the da Vinci system for transoral robotic surgery.
- 3. Facial nerve detection in parotid gland surgery.
- 4. Microscopic tympanostomy with ventilation tube insertion.

The curriculum was developed following the guidelines of the American Board of Otolaryngology [3].

To enhance learning, four 10-minute instructional videos were produced, offering detailed demonstrations grounded in authoritative textbooks and manufacturer manuals. These videos were segmented into key sessions to provide step-by-step guidance. Using multimedia demonstrations of diverse worked examples [37] and enriched with self-explanation prompts [38], the videos effectively illustrated both the setup of surgical instruments and operative techniques.

Real-patient demonstrations were recorded using a 360° camera (Garmin VIRB 360, Garmin Ltd., Kansas City, MO, United States) and a digital camera (D650, Nikon Imaging Japan Inc., Minato-Ku, Tokyo, Japan).

IBVR module

The immersive IBVR module was developed by converting video recordings into an interactive IBVR format using the VIVEPAPER[™] program (HTC Corp., New Taipei, Taiwan). Participants used hand controllers to navigate the 360° virtual environment and activate session markers, allowing them to access 2D video sessions focused on instrumental preparation and essential surgical skills. This module featured immersive 3D and 2D video content, accessible through a VR HMD, providing participants with a fully immersive and engaging learning experience (Fig. 3, left panel).

TBVR module

The semi-immersive TBVR module allowed participants to interact with textual and pictorial learning materials using hand controllers to turn virtual book pages. The content was displayed within a focused 120° field of view using the same VR HMD, ensuring consistency in hardware across modules (Fig. 3, right panel). Instructional



IBVR

TBVR

Fig. 3 Representative virtual reality (VR) scenes in the two training modules: image-based VR (IBVR) (left panel) and textbook-based VR (TBVR) (right panel). In the IBVR module, participants use hand gestures or controllers to select labeled subsections in the virtual environment and watch corresponding two-dimensional videos. In the TBVR module, participants use hand gestures or controllers to turn virtual book pages, enabling interaction with text and photographs

materials, adapted from authoritative textbooks and manufacturer manuals, were integrated into the TBVR format through the same VR software. This module provided clear, step-by-step explanations of sequential procedures using a combination of text and photographs, offering a structured and accessible learning experience. Importantly, participants in the TBVR condition did not move within the virtual environment to view 2D video sessions, which distinguishes the active interaction in TBVR from that in IBVR.

Both IBVR and TBVR systems were developed in accordance with the Analysis, Design, Development, Implementation, and Evaluation model, ensuring a systematic and structured approach to the design and implementation of instructional strategies [39]. Each module underwent a thorough review by two experienced instructors to verify the accuracy and consistency of the educational material.

Randomization and blinding

Participants were randomly assigned in a 1:1 ratio to either the IBVR-TBVR group or the TBVR-IBVR group. Randomization was conducted using the Random Number Generator in IBM SPSS software (version 25; IBM Corp., Armonk, NY, USA). Stratification based on age, sex, and cognitive style was applied to ensure balanced group characteristics. The allocation sequence remained concealed until the implementation of the video modules to maintain blinding.

To preserve the study's integrity and minimize potential bias, all assessors evaluating participant performance using Milestone and DOPS rating forms were blinded to group allocation. Assessors were provided with anonymized participant identification numbers and were not informed of the specific VR module (IBVR or TBVR) used during training sessions. This ensured that evaluations were conducted independently of group assignment.

Intervention

After randomization, participants were unblinded and given 10 min to engage with their assigned modules using a VIVE Pro VR headset (HTC Corp., New Taipei, Taiwan) to learn a procedure. Before starting the intervention, participants received instructions on how to use the VR headset and controllers. In the IBVR group, learners could freely explore the instructor's demonstrations, which detailed the knowledge and skills required for targeted surgeries. Conversely, in the TBVR group, learners accessed text and pictorial materials at their discretion. Within several hours following the intervention, each learner was tasked with performing the corresponding procedure on a real patient in the operating room, with the session lasting between 10 and 20 min.

Assessment of learning outcomes

Small quiz

Before each training session, participants completed a single-question quiz designed to assess their prior knowledge related to the session's content (Fig. 4). The interactive models used in the quiz, a machine gun and a crossbow, were default assets provided by the VR development software during the prototype stage. These models were chosen as placeholders to facilitate interaction



Fig. 4 Small quiz virtual reality scenes featuring a link game (top left panel) and multiple-choice questions (remaining panels)

within the virtual environment and were not intended to have thematic relevance to the training content.

The quizzes were designed to be completed within 60 s, focusing on evaluating participants' ability to interact with the virtual panels and their knowledge of specific operational skills. After completing the 10-minute learning module, participants retook the same quiz, resulting in four pretest and four posttest quizzes for each training module, with scores ranging from 0 to 4. To ensure alignment with the learning objectives, two staff members reviewed and verified the content validity and relevance of the quiz questions.

Milestone

Since 2004, Milestones have been used to assess the competency development of resident physicians across key domains in ORL-HNS [3]. Two investigators evaluated the participant's performance using a 5-level Milestone scale, specifically tailored to ORL-HNS conditions, including otologic, rhinologic, laryngologic, and head and neck neoplastic diseases. The introduction of Milestones 2.0 has enhanced this framework by adding harmonized milestones across all specialties and incorporating a supplemental guide with examples and resources to improve user proficiency with the tool [40]. The Milestone levels ranged from Level 1 (novice resident: new to the specialty), Level 2 (advanced beginner: performs some tasks with limited supervision), Level 3 (competent resident: completes common tasks independently), Level 4 (proficient resident: expected level at graduation), to Level 5 (expert resident: exceeds peer performance).

DOPS

During target procedures, the same investigators assessed their procedural skills using a validated 10-item DOPS form for ORL-HNS surgeries [34, 41]. Each behavior was rated on a scale from 1 (below expectations) to 10 (above expectations) [42]. For a focused assessment, four items were specifically chosen to evaluate instrumental preparation: preparation pre-procedure, determination of operation areas, technical ability to perform safely, and aseptic technique. In addition to those four items, seeking help and post-procedure management were further chosen to assess surgical skills. These items effectively capture both instrumental preparation and essential surgical skills.

Estimation of cognitive load

We selected the Paas CLS, the NASA TLX, and the CLC questionnaires because they are widely validated tools in educational research and are particularly suited for assessing cognitive load in VR-based learning environments [22, 23, 43]. These measures capture different dimensions of cognitive load: intrinsic, extraneous, and

germane, providing a comprehensive evaluation of the cognitive demands placed on learners during VR training. Their relevance to VR-based learning lies in their ability to quantify the mental effort required for interacting with complex virtual environments, which directly impacts learning outcomes [31–33].

CLS

Immediately after the intervention, we employed the Paas CLS [31] to estimate the total cognitive load. This scale is a single-item measure that asks participants to rate the intensity of mental effort on a 9-point scale (1 = very, very low mental effort; 9 = very, very high mental effort). The CLS has demonstrated good reliability (Cronbach $\alpha = 0.82-0.90$) in instructional research [44].

TLX

This subjective assessment tool measures cognitive load across six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration [32]. Participants rate each dimension on a visual analogue scale (0–20) immediately after the intervention. The TLX is known for its reliability (Cronbach $\alpha \ge 0.80$) in evaluating cognitive load [45].

CLC

This instrument assesses intrinsic (task difficulty and complexity), extraneous (instructional clarity and relevance), and germane (practical focus and amount of learning) cognitive loads [33]. Participants rate each of the six items on a five-point Likert scale (1–5), allowing for a detailed cognitive load score (2–10 per type, 6–30 total). The CLC has acceptable reliability (Pearson correlation coefficient = 0.40-0.62) and validated scores for use in workshop design and evaluation [46].

Measurement of HRV

Participants sat quietly and peacefully for 20 min to ensure physical and emotional stabilization before recording baseline HRV metrics and preparing for the study, reducing the risks of simulator sickness and negative affectivity [47, 48]. Following this, participants donned the HMD and breathed normally before engaging in the intervention. HRV was measured using a Nexus-4 amplifier and recording system (MindMedia BV, Herten, The Netherlands), which recorded electrocardiogram (ECG) signals from a single lead (lead I). HRV parameters were analyzed in accordance with the guidelines set by the European Society of Cardiology and the North American Society of Pacing and Electrophysiology [49].

Electrocardiogram data were acquired at a sample rate of 1024 Hz and saved as raw data for analysis. HRV analysis focused on 5-minute epochs of ECG signals, which were processed using custom-developed MATLAB scripts (The MathWorks, Inc., Natick, MA, USA). ECG signals were continuously recorded for 12 min: during the 10-minute intervention and the subsequent two minutes. HRV metrics were captured once every 30 s from the 5th to the 12th minute to monitor dynamic changes during and immediately after the session.

Sequences of normal-to-normal (NN) intervals were selected, excluding artifacts, ventricular, or supraventricular excitations [50]. Time-domain HRV measures included the standard deviation of NN intervals (SDNN) and the root mean square of successive differences between normal heartbeats (RMSSD). Frequency-domain measures included spectral power in the low-frequency (LF, 0.04–0.15 Hz) and high-frequency (HF, 0.15–0.40 Hz) bands, along with the LF/HF ratio.

The selection of HRV metrics (SDNN, RMSSD, and LF/ HF ratio) was based on their established roles as physiological indicators of stress and cognitive performance [51–54]. HRV provides a valuable assessment of autonomic nervous system activity, enabling objective evaluation of participants' stress levels and engagement during training. Studies have demonstrated that HRV metrics are sensitive to mental workload and stress induced by immersive technologies, validating their relevance for evaluating VR-based educational interventions [22, 55, 56].

Satisfaction score and learning acceptance

GSS: Post-intervention, we assessed "learning satisfaction" using the GSS, which ranges from 0 (very dissatisfied) to 10 (very satisfied), measured on a visual analogue scale [35].

AttrakDiff2

Developed to evaluate the acceptance of technical innovations, the AttrakDiff2 questionnaire [36] uses 28 items to assess four qualities: pragmatic quality, hedonic stimulation, hedonic identification, and attractiveness. Participants respond on a 7-point Likert-like scale, ranging from -3 to 3, in a semantic differential format. The mean values for each quality are calculated to create a scale value for each category. The AttrakDiff2 has been used extensively to assess learner experiences [57].

Qualitative feedback

At the conclusion of the intervention, semi-structured group interviews were conducted to gather comprehensive feedback and reflections from all participants. Figure 5 illustrates the sequential structure and content of the interview questions as well as the thematic analysis process. The interviews were guided by a predefined set of open-ended questions to ensure consistency while allowing flexibility for participants to share unique perspectives. These questions explored various aspects of the IBVR and TBVR modules, including.

- 1. Engagement: "How engaging did you find the IBVR and TBVR modules? Please elaborate."
- 2. Usability: "Were there any challenges or barriers to using the VR modules?"
- 3. Instructional Content: "Did the content provided in IBVR and TBVR meet your learning needs?"
- 4. Cognitive Load: "How would you describe your mental effort while using IBVR and TBVR?"
- 5. Suggestions for Improvement: "What improvements would you suggest for these modules?"

The qualitative data collected from the interviews were analyzed using the thematic analysis method described by Braun and Clarke [58]. This approach followed six systematic steps: (1) familiarization, (2) generating initial codes, (3) searching for themes, (3) reviewing themes, (4) defining and naming themes, and (5) writing up.

Main outcome measurements

The primary outcome of this study was the Milestone score measured post-intervention. Secondary outcomes included DOPS scores, cognitive load scores, HRV indices, GSS, AttrakDiff2 scores, and qualitative feedback.

Sample size estimation

Sample size was estimated using primary outcome data collected at the beginning of the study (IBVR: 3.1 ± 0.7 ; TBVR: 2.2 ± 0.5). A two-tailed Wilcoxon signed-rank test with the Lehmann method was applied. Based on the observed effect size (Cohen's d = 1.44), a type I error rate of 0.05, and a desired power of 80%, the required minimum sample size was calculated to be seven participants per group.

Statistical analysis

The D'Agostino and Pearson test was used to assess normality. As most variables were not normally distributed, continuous variables are reported as medians with interquartile ranges (IQR). Between-group comparisons of continuous variables were conducted using the Mann-Whitney U test, while within-group comparisons were analyzed using the Wilcoxon signed-rank test. Categorical variables were compared using Fisher's exact test.

HRV data across multiple timepoints were analyzed using generalized estimating equations with a firstorder autoregressive correlation structure to account for within-subject correlation. A two-tailed *p*-value of < 0.05was considered statistically significant.

All statistical analyses were performed using G*Power 3.1.9.7 (Heinrich Heine University, Düsseldorf, Germany), GraphPad Prism 9.0 (GraphPad Software Inc.,



Fig. 5 Flow diagram illustrating the thematic analysis process, including the inductive analysis of responses to a predefined set of open-ended questions

San Diego, CA, USA), and IBM SPSS Statistics version 29.0 (IBM Corp., Armonk, NY, USA).

Results

Participant characteristics

Ten volunteers (four males [40%] and six females [60%]) with a median age of 28 years (IQR, 26–31) showed high interest, with none declining to participate (Fig. 1). Five participants (50%) were first-year ORL-HNS residents, and five (50%) were second-year residents. Nine participants (90%) demonstrated a field-independent cognitive style, and one (10%) was field-dependent. Baseline characteristics were comparable between the IBVR-TBVR and TBVR-IBVR groups, with no statistically significant differences (all p > 0.05; Table 1).

Comparison of IBVR and TBVR modules

Regarding learning outcomes and surgical performance, there were no statistically significant differences in the number of correct answers for small quizzes, whether in pretests, posttests, or percentage change, between the IBVR and TBVR modules (all p > 0.05; Table 2). However, the IBVR module showed significantly higher Milestone levels (8 [IQR, 6–9] vs. 6 [IQR, 6–7]; p = 0.04) and total DOPS scores (140 [IQR, 134–147] vs. 129 [IQR, 124–134]; p < 0.01) compared to the TBVR module (Fig. 6).

When comparing cognitive load, the distributions of the total CLS score, the six TLX subscales, and four CLC type score between the IBVR and TBVR modules were comparable (all p > 0.05; Table 2).

Dynamic changes in HRV metrics across time and VR module subgroups are depicted in Fig. 7. Using generalized estimating equations while controlling for age, sex, residence level, and time, HR, SDNN, RMSSD, and LF/ HR ratio showed no relationship with VR module type or training sequence (all p > 0.05; Table 3). However, SDNN (β , -75.8 [95% CI, -128.2–-23.4], p < 0.01) and RMSSD (β , -75.6 [95% CI, -137.8–-13.5], p = 0.02) were significantly associated with the interaction between VR

Characteristics	Overall	IBVR-TBVR	TBVR-IBVR	P-val-	
	group	group	group	ue ¹	
Participants	n=10	n=5	n=5		
Demographics					
Age (years)	28 (26–31)	26 (26–30)	28 (28–34)	0.08	
Men (<i>n</i> [%])	4 (40)	2 (40)	2 (40)	>0.99	
Cognitive style					
Field-indepen-	9 (90)	5 (100)	4 (80)	> 0.99	
dence (<i>n</i> [%])					
Residency level					
R1 (<i>n</i> [%])	5 (50)	4 (80)	1 (20)	0.21	
Heart rate					
variability					
HR (bpm)	75 (70–80)	78 (72–90)	78.0 (69–84)	0.58	
SDNN-before	59.2	60.2	60.02	0.68	
(ms)	(41.4–79.4)	(49.7–87.9)	(41.3–92.6)		
RMSSD-before	49.0	54.6	39.09	0.53	
(ms)	(23.0–75.5)	(25.8–76.2)	(22.5–82.9)		
LF/HF ratio	1.34	1.32	1.75	0.70	
	(0.80-2.74)	(0.80-2.51)	(0.75-3.11)		

 Table 1
 Comparisons of baseline participants characteristics

Continuous variables are summarized as median (interquartile range) and categorize variables are summarized as n (%), as appropriate. Abbreviations: CI: confidence interval; HR: heart rate; IBVR: image-based virtual reality; LF/HF ratio: low frequency/high frequency ratio; RMSSD: root mean square of successive heartbeat interval difference; SDNN: standard deviation of normal-to-normal intervals; TBVR: textbook-based virtual reality. ¹Between-group differences were analyzed using the Mann-Whitney U test for continuous variables and Fisher's exact test for categorical variables, as appropriate

module and training sequence. Additionally, changes in HRV metrics were not influenced by time (all p > 0.05).

The GSS score for the IBVR module was significantly lower than that for the TBVR module (8 [IQR, 7–9] vs. 8 [IQR, 8–10]; p = 0.03; Table 2). While there were no significant differences in pragmatic quality or attractiveness, the IBVR module demonstrated significantly lower scores in hedonic stimulation (1.5 [IQR, 0.9–1.8] vs. 1.9 [IQR, 1.5–2.4]; p = 0.01) and hedonic identification (1.6 [IQR,

Page 11 of 19

Table 2 Comparisons of leaning outcomes, cognitive loadestimates, satisfaction, and learning acceptance between theIBVR and TBVR modules

Variables	IBVR module	TBVR module	<i>P</i> -val- ue ¹
Participants	n=10	n=10	
Learning outcomes			
Milestone (overall levels)	8 (6–9)	6 (6–7)	0.04
Small quiz (number of correct answers)			
Pretest	3 (3–3)	3 (3–3)	0.56
Posttest	3 (3–3)	3 (3–3)	0.41
Percentage small quiz change (%)	0 (0–38)	0 (-27-38)	0.49
DOPS (overall scores)	140 (134–147)	129 (124–134)	< 0.01
CLS	6 (5–7)	5 (5–6)	0.91
TLX	53 (41–65)	48 (40–61)	0.39
Mental demand	13 (10–15)	12 (7–14)	0.26
Physical demand	6 (4–12)	6 (4–13)	0.78
Temporal demand	8 (5–13)	8 (4–10)	0.59
Performance	10 (5–15)	12 (4–15)	0.77
Effort	11 (8–12)	9 (7–13)	> 0.99
Frustration	5 (3–7)	3 (2–5)	0.38
CLC	19 (17–21)	19 (18–21)	0.86
Intrinsic cognitive load	4 (2–5)	4 (3–4)	0.48
Extraneous cognitive load	8 (7–9)	8 (7–8)	0.26
Germane cognitive load	8 (6–9)	8 (7–9)	0.08
GSS	8 (7–9)	8 (8–10)	0.03
AttrakDiff2			
Pragmatic quality	1.7 (1.1–2.5)	2.1 (1.3–2.6)	0.09
Hedonic stimulation	1.5 (0.9–1.8)	1.9 (1.5–2.4)	0.01
Hedonic identification	1.6 (1.2–1.8)	2.1 (1.7–2.7)	0.03
Attractiveness	1.6 (1.3–1.9)	2.0 (1.4–2.6)	0.11

Continuous variables are summarized as median (interquartile range). Abbreviations: Cl: confidence interval; CLC: Cognitive Load Component; CLS: Cognitive Load Scale; DOPS: Direct Observation of Procedural Skills; GSS: global satisfaction scale; IBVR: image-based virtual reality; MCQ: multiple-choice question; TBVR: textbook-based virtual reality; TLX: NASA Task Load Index. ¹Within-group differences were analyzed using the Wilcoxson signed rank test



Fig. 6 Milestone levels (left panel) and Direct Observation of Procedural Skills (DOPS) scores (right panel) compared between image-based virtual reality (IBVR) and textbook-based virtual reality (TBVR) subgroups. Boxes indicate medians and interquartile ranges, with whiskers representing minimum and maximum values



Fig. 7 Alterations in heart rate variability metrics over time and across virtual reality (VR) module subgroups. Bars represent medians, and whiskers indicate interquartile ranges. Abbreviations: HR, heart rate; IBVR, image-based VR; LF/HF ratio, low frequency/high frequency ratio; RMSSD, root mean square of successive heartbeat interval differences; SDNN, standard deviation of normal-to-normal intervals; TBVR, textbook-based VR

1.2–1.8] vs. 2.1 [IQR, 1.7–2.7]; p = 0.03) compared to the TBVR module.

Qualitative feedback

At the conclusion of the study, all participants participated in group interviews to provide detailed feedback and reflections (Table 4). The qualitative analysis followed Braun and Clarke's thematic analysis framework, identifying initial codes and progressing through themes to final insights.

Key initial codes emerging from the data included instructional content (100%), cognitive load (70%), selfefficacy (60%), learning experience (30%), technology acceptance (30%), evaluation (10%), and simulator sickness (10%). The majority (70%) of participants reported a positive experience with IBVR, while 50% expressed positive feedback regarding TBVR.

Through the grouping of codes, initial themes were identified, including engagement (70%), usability (90%), instructional content (100%), cognitive load (70%), and suggestions for improvement (60%).

The refinement of initial themes led to the identification of two final themes:

- 1. Usability of instructional content: This theme emphasized the value of detailed and practical instructional materials in enhancing the modules' effectiveness. For example, Participant No. 10 stated, *"The IBVR training displayed more detailed aspects of the actual operation and could potentially replace some hands-on training."*
- 2. Easy engagement: This theme reflected participants' experiences with the modules' design, focusing on how cognitive load and engagement influenced their learning. For instance, Participant No. 5 remarked, "*The monotone voice used in the IBVR training became tiring and caused me to feel sleepy after listening for extended periods.*"

Parameters	Module		Training sequen	e	Interaction ¹		Age		Sex		Residence level		Time	
Reference value	IBVR		first-training				youngest		female				baseline	
	β (95% CI)	P-value	β (95% Cl)	P-value	β (95% CI)	P-value	β (95% CI)	P-value	β (95% CI)	P-value	β (95% CI)	P-value	β (95% Cl)	P-value
HR	13.3 (-11.4-38.0)	0.29	17.1 (-7.4-41.5)	0.17	-8.5 (-49.9-32.9)	0.69	-1.8 (-7.9-4.3)	0.56	24.2 (-13.2-61.6)	0.21	-12.5 (-30.7-5.7)	0.18	0.30 (-0.35-0.94)	0.37
SDNN	38.2 (-3.6-80.0)	0.07	52.5 (11.1–94.0)	0.13	-75.8 (- 128.223.4)	< 0.01	-12.0 (-18.35.	7) <0.01	59.1 (18.8–99.3)	< 0.01	-22.4 (-44.70.1)	0.049	1.3 (-0.6-3.2)	0.18
RMSSD	49.0 (-4.3-102.2)	0.07	47.3 (-4.3-98.9)	0.07	-75.6 (- 137.813.5)	0.02	-12.6 (-20.05	2) <0.01	67.0 (17.4-116.7)	< 0.01	-35.6 (-57.214.1)	< 0.01	0.4 (-1.2-2.1)	0.62
LF/HF ratio	-1.2 (-3.8-1.4)	0.36	0.3 (-2.3-3.0)	0.80	0.1 (-4.3-4.4)	0.98	0.5 (0.2–0.8)	< 0.01	0.1 (-1.5-1.6)	0.96	-1.0 (-3.1-1.2)	0.38	0.1 (-0.1-0.1)	0.99
Parameters	Group		1	Age			Sex		Resider	level acc		F	ime	
Reference value	IBVR-TBVR		^	oungest		-	female		R1			q	aseline	
	β (95% CI)	4	P-value f	3 (95% CI)	<i>P</i> -value	1	β (95% CI)	P-value	β (95%	C)	P-value	g	(95% CI) /	9-value
HR	5.1 (-18.2-28.5)	0		-2.0 (-8.4-4.4)	0.54		25.3 (- 13.7-64.4)	0.20	-12.1 (-	· 30.9–6.6)	0.21	I	0.1 (-0.6-0.4), (0.70
SDNN	38.1 (11.7–64.6)	v	< 0.01	-12.0 (-18.45	5.6) < 0.01		59.3 (18.4–100.3)	< 0.01	-22.3 (-	.44.5-0.02)	0.05	1	3 (-0.5-3.1), (2.17
RMSSD	38.1 (6.8–69.5)	0	- 20.0	-12.6 (-20.1	5.2) < 0.01		67.4 (17.0–117.8)	< 0.01	-35.6 (-	57.114.1)	< 0.01	0	4 (-1.3-2.0), ().66
LF/HF ratio	-0.1 (-2.2-2.1)	0	0.96	15 (0.8-11.1)	< 0.01)	0.1 (-1.5-1.5)	0.98	-1.0 (-3	3.1-1.2)	0.38	0	1 (-0.1-0.1), (0.92

Discussion The findings of this study suggest that both IBVR and TBVR modules contribute to enhancing the surgical skills and knowledge of junior residents. IBVR may

TBVR modules contribute to enhancing the surgical skills and knowledge of junior residents. IBVR may offer advantages in surgical competency, as indicated by trends toward higher Milestone and DOPS scores, without significantly increasing cognitive load or affecting HRV metrics. However, these benefits may come at the expense of user experience, as IBVR was associated with lower global satisfaction, hedonic stimulation, and hedonic identification. In contrast, TBVR was perceived as more accessible, with its structured and engaging format making it particularly effective for procedural reviews and foundational learning. Meanwhile, IBVR was valued for its detailed instructional content and potential to enhance self-efficacy, particularly in complex surgical scenarios.

While IBVR showed significant benefits in competency development, the observed lower global satisfaction, hedonic stimulation, and hedonic identification highlight areas for improvement in novice learner experience. Based on participant feedback, a phased approach to surgical training may be beneficial: starting with TBVR for establishing foundational knowledge and procedural familiarity and gradually incorporating IBVR to enhance readiness for complex scenarios and real-world applications. Training modules should be tailored to accommodate learners at different levels of expertise, with TBVR serving as a stepping stone to the more immersive and cognitively demanding IBVR.

Future iterative improvements to IBVR modules could enhance user engagement and reduce cognitive strain by incorporating clearer, more vivid voice narrations, minimizing background noise, and offering shorter, focused training sessions. To mitigate the lower hedonic stimulation associated with IBVR, integrating gamification elements, dynamic feedback, and personalized content could make the experience more enjoyable and engaging [59]. Additionally, eliminating repetitive demonstrations and ensuring that instructional content is concise and targeted would further improve the learner experience [60]. These refinements aim to enhance hedonic stimulation and identification, thereby striking an optimal balance between instructional effectiveness and learner satisfaction.

Comparisons with previous research

Our findings align with prior research, including metaanalyses demonstrating that immersive VR generally has a positive, albeit modest, effect on learning outcomes across various disciplines, particularly at junior levels [61]. These effects are attributed to immersive VR's ability to promote active engagement and enhance spatial reasoning. Our study builds on this by showing that

Table 4 Qualitative feedback of the IBVR and TBVR modules

No.	Age (years)	Sex	Cog- ni-	Resi- dence	IBVR module	TBVR module	Initial themes
			tive style	level			
1	26	F	FI	R1	Having participated in ventilation tube insertion procedures allows me to observe them more atten- tively and receive more targeted feedback. In con- trast, since I have not been involved in any da Vinci operations, I find it challenging to retain the details after viewing VR simulations of these procedures	The sensor location of the facial nerve detector and the position of the ground line could be illustrated in greater detail.	engagement, us- ability, instructional content, cognitive load, suggestions for improvement
2	26	F	FI	R1	The IBVR training videos are meticulously produced, and the accompanying quizzes are notably engaging.	I can control my reading speed more effectively with the textbook com- pared to watching the videos.	engagement, us- ability, instructional content, cognitive load
3	26	F	FI	R1	l gained a lot from the IBVR training, although the narration was often disrupted by background noise.	TBVR instructions appear to be more straightforward to comprehend than image-based VR instructions.	engagement, us- ability, instructional content, cognitive load
4	28	Μ	FI	R1	The IBVR training captures the feel of actual surgical operations well. Enhancing the simulations to allow for operating instruments in VR could make the experience even more realistic. Nevertheless, the training provides valuable feedback and has significantly improved my practice in the operating room.	The text should provide more detailed descriptions of the procedures, and the illustrations could be clearer. Some terms were unclear to me.	engagement, us- ability, instructional content, cognitive load, suggestions for improvement
5	31	Μ	FI	R2	The monotone voice used in the IBVR training became tiring and caused me to feel sleepy after listening for extended periods.	The TBVR training contained too many slides on each page. I found the learning effectiveness of the IBVR to be superior to that of the TBVR.	engagement, us- ability, instructional content, cognitive load
6	28	F	FI	R1	The pace of the IBVR training was somewhat too slow.	There were too many tests, and the font size of the text was quite small.	instructional content
7	32	Μ	FD	R2	The effectiveness of the IBVR training could be enhanced by integrating it with actual practice sessions.	Very stylish.	usability, instructional content, suggestions for improvement
8	28	F	FI	R2	The titles were overly wordy. While the narration was clear, it lacked intonation, and the video quality was high.	It was difficult to navigate the pages of the VR textbooks. I suggest reducing the number of words in the dialog boxes.	usability, instructional content, suggestions for improvement
9	28	F	FI	R2	The IBVR training provided clearer instructions than the TBVR training, as the steps were depicted more vividly and directly. I recommend reducing the time spent on demonstrating sensor placements for the facial nerve detector. Additionally, could IBVR training be extended to include neck dissection procedures?	The pretest and posttest small quizzes aligned well with the TBVR instruc- tional content and were more engag- ing than reading the actual textbook. However, the head-mounted display made me feel a bit dizzy. Additionally, some sentences related to the da Vinci system were somewhat difficult to understand.	engagement, us- ability, instructional content, cognitive load, suggestions for improvement
10	36	Μ	FI	R2	The IBVR training displayed more detailed aspects of the actual operation and could potentially replace some hands-on training. It would be beneficial to eliminate repetitive demonstrations, like sensor placements, to maintain attention.	TBVR training was more suitable for reviewing procedures. The VR provided an excellent visual experience and captured my attention.	engagement, us- ability, instructional content, cognitive load, suggestions for improvement

Abbreviations: F: female; FD: field dependence; FI: field independence; IBVR: image-based virtual reality; M: male; R1, the first-year resident; R2, the second-year resident; TBVR: textbook-based virtual reality

IBVR is more effective than TBVR in improving Milestone and DOPS scores, highlighting its potential for skill acquisition in surgical training. However, consistent with findings by Kavanagh et al. (2017) [62], the impact of immersive VR on advanced psychomotor skills, such as surgery, appears less pronounced, potentially due to increased cognitive load and technical, ergonomic, and cost-related challenges. Despite these limitations, immersive VR remains valuable in hands-on learning contexts by fostering spatial knowledge and procedural memory [63, 64]. Our study provides unique insights into the comparative effectiveness of semi-immersive (TBVR) and fully immersive (IBVR) VR systems, particularly when applied sequentially. While the small sample size limits generalizability, the observed benefits of a TBVR-IBVR sequence in optimizing skill acquisition and user experience are noteworthy. By incorporating physiological metrics such as HRV and cognitive load assessments, our research offers a multidimensional evaluation of VR-based training—an aspect underexplored in previous literature. Participant feedback further enriches these findings, emphasizing the usability of instructional content and ease of engagement as critical factors influencing the outcomes of VR training modules.

Despite advancements in VR-based educational tools, a critical gap remains in the provision of adaptive learning content tailored to individual needs [65]. Our findings suggest that a TBVR-IBVR sequence facilitates a smoother transition for VR novices, reducing learning barriers while enhancing surgical skill acquisition. This structured approach may help learners progressively adapt to the more immersive and cognitively demanding IBVR modules.

As learners build foundational skills in TBVR's structured environment, they develop the confidence necessary to navigate the challenges of IBVR, which requires higher interaction and cognitive engagement. While cognitive load reports were similar across groups, the IBVR-TBVR group exhibited lower SDNN and RMSSD metrics, indicative of higher stress levels and reduced cognitive performance [52]. Conversely, the TBVR-IBVR group scored higher in pragmatic quality and hedonic identification during the initial training phase.

This phased approach aligns with self-determination theory, which posits that meeting the psychological needs of competence, autonomy, and relatedness fosters motivation and goal achievement [66]. Gradually increasing the complexity of learning environments allows learners to develop effectively without becoming overwhelmed, enhancing their readiness for more advanced VR-based scenarios [67]. A learner-centered approach combining TBVR's simplicity with IBVR's immersive experience optimizes the educational trajectory, improves surgical training outcomes, and better prepares learners for realworld applications while fully leveraging VR technology.

Thematic analysis identified key strengths and areas for improvement in both IBVR and TBVR modules. IBVR's rich instructional content and immersive design were highlighted as significant advantages, particularly in enhancing usability for complex surgical training. However, challenges related to cognitive load and engagement underscored the need for iterative refinement to further improve the learner experience.

Participant feedback highlights the value of IBVR:

- Participant 1 appreciated receiving targeted feedback, especially with prior ventilation tube insertion experience.
- Participant 3 noted significant benefits from IBVR.
- Participant 4 commented on the realistic feel of the surgical operations captured by IBVR.
- Participant 9 found IBVR clearer and more vivid than TBVR.
- Participant 10 observed that IBVR offered detailed insights into actual operations, suggesting its potential to supplement or even replace certain aspects of hands-on training.

Overall, our results suggest that while IBVR improves skill acquisition, its impact on knowledge retention is less evident. The high fidelity of IBVR likely enhances spatial reasoning and procedural memory, key for complex surgical training. Higher fidelity VR systems offer a more realistic experience and better retention of spatial knowledge [68]. Thus, the use of IBVR in surgical education not only aids in mastering complex procedures but also helps solidify spatial and procedural understanding, which is essential for effective surgical training.

However, IBVR may have lower hedonic qualities compared to TBVR for several reasons:

- Cognitive demands: IBVR immerses learners in a rich, interactive 3D environment, which can be more cognitively demanding than TBVR. Although a recent study indicated that redundant formats in an immersive VR environment did not increase cognitive load compared to solitary formats, IBVR requires active engagement, which can be tiring or stressful for some learners, leading to reduced enjoyment.
- Technological limitations: Hardware issues, such as motion sickness or discomfort, can arise in VR conditions [69]. However, the higher level of immersion in IBVR may amplify these effects due to the fully immersive nature of the experience, which often involves more sensory input and prolonged exposure.
- Expectations vs. reality: High expectations for immersion may not be fully met, affecting user satisfaction [70]. As noted by participants 5, 6, and 8, this gap between expectation and reality can negatively affect their hedonic experience.

While less immersive, TBVR is effective during early learning phases by offering a structured, user-friendly environment that prevents cognitive overload [71]. Its benefits include:

- Enhanced focus: Clear, distraction-free content presentation aids in foundational learning [72].
- Ease of use: Simple interfaces make TBVR more accessible, especially for less tech-savvy learners [73].
- Structured learning: TBVR's step-by-step format is ideal for building foundational knowledge [74].
- Technological limitations: Motion sickness or discomfort can occur in both IBVR and TBVR conditions. However, TBVR's semi-immersive nature and structured presentation may reduce the likelihood and intensity of such issues compared to the fully immersive IBVR. This advantage is attributed to TBVR's simpler interaction design and lower sensory input, which can create a more stable and predictable learning environment [61]. Participant feedback also highlighted TBVR's user-friendly design as a key factor in minimizing discomfort and maintaining engagement during training.
- Integration with traditional learning: TBVR can supplement textbooks or lectures with interactive sessions that bridge the gap between theoretical knowledge and practical application [75].

Study limitations

This proof-of-concept study offers valuable insights but has several limitations. First, participants' prior VR experience was not considered, which may have affected the learning curve, especially for novices engaging with the immersive IBVR content. Second, the use of a machine gun and a crossbow as placeholders in the small quizzes, which were default assets from the VR development software, may have caused confusion, detracted from the educational focus of the application, or potentially influenced HRV metrics in participants sensitive to weapon imagery. Future iterations will address this issue by replacing these placeholders with anatomically accurate hand models or neutral tools more appropriate for medical training. Third, operational skills were assessed through direct observation of interactions and surgical performance. However, variability in residents' prior experience with the assessed procedures could have influenced the results. Although common assessment criteria were used, a comprehensive evaluation of all clinical competencies was not feasible. Fourth, the study involved only ten junior residents, which limits the generalizability of the findings. As a proof-of-concept pilot study, the primary goal was to explore feasibility and trends rather than draw definitive conclusions. Future larger-scale studies with more diverse cohorts are necessary to validate these results and assess the long-term effectiveness of IBVR and TBVR training modules in broader clinical settings. Additionally, this study only evaluated the immediate outcomes of VR training over a short-term, 8-session program. The lack of long-term follow-up prevents the assessment of skill retention and transferability acquired through IBVR and TBVR. Future studies should include extended follow-up periods to evaluate the durability of learning outcomes and their applicability in realworld clinical settings. Lastly, the VR training modules were conducted in a controlled, artificial environment that may not fully replicate the complexities and unpredictability of real-world surgical settings. The study also focused on a limited set of surgical procedures, which may restrict the generalizability of the findings to other contexts or specialties. Future research should explore a broader range of surgical scenarios and incorporate hybrid training environments that combine virtual and real-world elements to bridge the gap between VR-based

Conclusions

training and clinical practice.

This proof-of-concept study suggests that IBVR modules may enhance Milestone and DOPS scores more effectively, while TBVR provides an accessible starting point, facilitating clarity and engagement without increasing cognitive load or stress. A structured progression from TBVR to IBVR could optimize training efficiency, supporting skill acquisition while managing cognitive demands. Given the study's exploratory nature and relatively small sample size, future large-scale research should further refine VR-based training strategies, investigate learner variability, and assess long-term skill retention to strengthen the evidence base for integrating VR in surgical education.

Abbreviations

CI	Confidence Interval
CLC	Cognitive Load Component
CLS	Cognitive Load Scale
DOPS	Direct Observation of Procedural Skills
GEFT	Group Embedded Figures Test
GSS	Global Satisfaction Score
HF	High-Frequency
HMD	Head-mounted Display
HRV	Heart Rate Variability
IBVR	Image-Based VR
IQR	Interquartile Range
LF	Low-Frequency
ORL-HNS	Otorhinolaryngology-Head and Neck Surgery
RMSSD	Root Mean Square of Successive Differences Between Normal
	Heartbeats
SDNN	Standard Deviation of NN Intervals
TBVR	Textbook-Based VR
TLX	Task Load Index
VR	Virtual Reality

Acknowledgements

We are deeply grateful to all the volunteers who participated in this study. We also extend our sincere thanks to our dedicated study staff, particularly Ping-Yun Lin, Yun-Kai Chang, Ming-Yen Chung, and Chung-Fang Hsiao from the Department of Otorhinolaryngology-Head and Neck Surgery, Linkou Medical Center, Chang Gung Memorial Hospital, Taoyuan City, Taiwan, ROC, for their invaluable technical assistance.

Author contributions

WNL - conceptualization, methodology, investigation, and led the writing of the original draft. HHC - conceptualization, formal analysis, methodology, validation, visualization, and co-led the writing of the original draft. YPC - conceptualization, data curation, formal analysis, methodology, software, validation, visualization, and resources. LJH - methodology, investigation, and validation. CJK - methodology, investigation, supervision, and validation. TJF - investigation, supervision, and validation. HYL - methodology, supervision, and validation. LAL - conceptualization, methodology, formal analysis and writing results, visualization, resources, led the writing of the original draft, project administration, and led funding acquisition as PI. All authors - contributed to the review and editing of paper drafts and approved the final version submitted.

Funding

This work was supported by grants from the National Science and Technology Council, Taiwan (grant number 108-2511-H-182 A-001), and the Chang Gung Medical Foundation, Taiwan (grant number CMRPG3L0811–2). The funding sources had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to institutional ownership of the data, privacy considerations, and ongoing data analysis but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of Chang Gung Medical Foundation (No: 201601821B0), and all procedures involving human participants were conducted in accordance with the ethical standards of the institutional and/or national research committee and with the 1975 Helsinki Declaration [26] and the CONSORT guidelines [27]. All participants provided written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Otorhinolaryngology-Head and Neck Surgery, Linkou Medical Center, Chang Gung Memorial Hospital, No. 5, Fu-Hsing Street, Gueishan District, Taoyuan 33305, Taiwan

²School of Medicine, Graduate Institute of Clinical Medicine Sciences, College of Medicine, Chang Gung University, Taoyuan, Taiwan

³Department of Community Medicine, Cathay General Hospital, Taipei, Taiwan

⁴School of Medicine, College of Life Science and Medicine, National Tsing Hua University, Hsinchu, Taiwan

⁵Department of Computer Science and Information Engineering, Chang Gung University, Taoyuan, Taiwan

Received: 16 October 2024 / Accepted: 25 April 2025 Published online: 07 May 2025

References

- Alkire BC, Raykar NP, Shrime MG, Weiser TG, Bickler SW, Rose JA, Nutt CT, Greenberg SLM, Kotagal M, Riesel JN, et al. Global access to surgical care: a modelling study. Lancet Global Health. 2015;3(6):e316–23.
- Hanna JS, Herrera-Almario GE, Pinilla-Roncancio M, Tulloch D, Valencia SA, Sabatino ME, Hamilton C, Rehman SU, Mendoza AK, Gómez Bernal LC, et al. Use of the six core surgical indicators from the lancet commission on global surgery in Colombia: a situational analysis. Lancet Global Health. 2020;8(5):e699–710.

- Tsue TT. Developing the otolaryngology milestones. J Grad Med Educ. 2014;6(1 Suppl 1):162–5.
- 4. Bergmark RW, Shaye DA, Shrime MG. Surgical care and otolaryngology in global health. Otolaryngol Clin North Am. 2018;51(3):501–13.
- Han SY, Lee DY, Chung J, Kim YH. Comparison of endoscopic and microscopic ear surgery in pediatric patients: A meta-analysis. Laryngoscope. 2019;129(6):1444–52.
- Fujiwara K, Fukuhara T, Niimi K, Sato T, Kitano H. Load evaluation of the Da Vinci surgical system for transoral robotic surgery. J Robot Surg. 2015;9(4):315–9.
- Chiesa-Estomba CM, Larruscain-Sarasola E, Lechien JR, Mouawad F, Calvo-Henriquez C, Diom ES, Ramirez A, Ayad T. Facial nerve monitoring during Parotid gland surgery: a systematic review and meta-analysis. Eur Arch Otorhinolaryngol. 2021;278(4):933–43.
- Curry M, Malpani A, Li R, Tantillo T, Jog A, Blanco R, Ha PK, Califano J, Kumar R, Richmon J. Objective assessment in residency-based training for transoral robotic surgery. Laryngoscope. 2012;122(10):2184–92.
- Khalafallah YM, Bernaiche T, Ranson S, Liu C, Collins DT, Dort J, Hafner G. Residents' views on the impact of robotic surgery on general surgery education. J Surg Educ. 2021;78(3):1007–12.
- Izard SG, Juanes JA, Garcia Penalvo FJ, Estella JMG, Ledesma MJS, Ruisoto P. Virtual reality as an educational and training tool for medicine. J Med Syst. 2018;42(3):50.
- Andersen SA, Mikkelsen PT, Konge L, Caye-Thomasen P, Sorensen MS. Cognitive load in mastoidectomy skills training: virtual reality simulation and traditional dissection compared. J Surg Educ. 2016;73(1):45–50.
- 12. Robertson GG, Card SK, Mackinlay JD. Three views of virtual reality: nonimmersive virtual reality. Computer 1993, 26(2).
- Jiang H, Vimalesvaran S, Wang JK, Lim KB, Mogali SR, Car LT. Virtual reality in medical students' education: scoping review. JMIR Med Educ. 2022;8(1):e34860.
- Mergen M, Graf N, Meyerheim M. Reviewing the current state of virtual reality integration in medical education - a scoping review. BMC Med Educ. 2024;24(1):788.
- Wei T, Xing Y, Wu Y, Kwan HY, Li L. Virtual reality design in reading user experience: 3D data visualization with interaction in digital publication figures. Sci Program. 2022;2022:1–7.
- Zheng Z, Wang B, Wang Y, Yang S, Dong Z, Yi T, Choi C, Chang EJ, Chang EY. Aristo. In: Proceedings of the 25th ACM international conference on Multimedia. 2017: 690–698.
- Macias-Velasquez S, Medellin-Castillo HI, Garcia-Barrientos A. New-user experience evaluation in a semi-immersive and haptic-enabled virtual reality system for assembly operations. Int J Hum Comput Stud 2024, 190.
- McGowin G. SM Fiore 2024 Mind the Gap! Advancing immersion in virtual Reality—Factors, measurement, and research opportunities. Proc Hum Factors Ergon Soc Annual Meeting 68 1 1648–54.
- 19. Snelson C. Y-C Hsu 2019 Educational 360-Degree videos in virtual reality: a scoping review of the emerging research. TechTrends 64 3 404–12.
- Deng S, Wheeler G, Toussaint N, Munroe L, Bhattacharya S, Sajith G, Lin E, Singh E, Chu KYK, Kabir S et al. A virtual reality system for improved Image-Based planning of complex cardiac procedures. J Imaging 2021, 7(8).
- Dubinski D, Won SY, Hardung C, Rafaelian A, Paschke K, Arsenovic M, Behmanesh B, Schneider M, Freiman TM, Gessler F, et al. Enhancing surgical education for medical students through virtual reality: the digital surgical operating theatre tour. World Neurosurg. 2025;194:123523.
- Chao YP, Chuang HH, Hsin LJ, Kang CJ, Fang TJ, Li HY, Huang CG, Kuo TBJ, Yang CCH, Shyu HY, et al. Using a 360 degrees virtual reality or 2D video to learn history taking and physical examination skills for undergraduate medical students: pilot randomized controlled trial. JMIR Serious Games. 2021;9(4):e13124.
- 23. Chao YP, Kang CJ, Chuang HH, Hsieh MJ, Chang YC, Kuo TBJ, Yang CCH, Huang CG, Fang TJ, Li HY, et al. Comparison of the effect of 360 degrees versus two-dimensional virtual reality video on history taking and physical examination skills learning among undergraduate medical students: a randomized controlled trial. Virtual Real. 2023;27(2):637–50.
- Sankaranarayanan G, Odlozil CA, Wells KO, Leeds SG, Chauhan S, Fleshman JW, Jones DB, De S. Training with cognitive load improves performance under similar conditions in a real surgical task. Am J Surg. 2020;220(3):620–9.
- Selzer MN, Gazcon NF, Larrea ML. Effects of virtual presence and learning outcome using low-end virtual reality systems. Displays. 2019;59:9–15.
- Shephard DA. The 1975 declaration of Helsinki and consent. Can Med Assoc J. 1976;115(12):1191–2.

- 27. Schulz KF, Altman DG, Moher D, Fergusson D. CONSORT 2010 changes and testing blindness in RCTs. Lancet. 2010;375(9721):1144–6.
- Lee LA, Chao YP, Huang CG, Fang JT, Wang SL, Chuang CK, Kang CJ, Hsin LJ, Lin WN, Fang TJ, et al. Cognitive style and mobile e-learning in emergent otorhinolaryngology-head and neck surgery disorders for millennial undergraduate medical students: randomized controlled trial. J Med Internet Res. 2018;20(2):e56.
- 29. Witkin HA, Oltman PK, Raskin E, Karp SA. A manual for the embedded figures tests. Palo Alto, CA: Consulting Psychologists; 1971.
- Cai J-Y, Wang R-F, Wang C-Y, Ye X-D, Li X-Z. The influence of learners' cognitive style and testing environment supported by virtual reality on English-speaking learning achievement. Sustainability. 2021;13(21):11751.
- 31. Paas FG. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. J Educ Psychol. 1992;84(4):429–34.
- Leppink J, Paas F, van Gog T, van der Vleuten CPM, van Merriënboer JJG. Effects of pairs of problems and examples on task performance and different types of cognitive load. Learn Instruction. 2014;30:32–42.
- Naismith LM, Cheung JJ, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. Med Educ. 2015;49(8):805–14.
- Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE guide 31. Med Teach. 2007;29(9):855–71.
- Lee LA, Wang SL, Chao YP, Tsai MS, Hsin LJ, Kang CJ, Fu CH, Chao WC, Huang CG, Li HY, et al. Mobile technology in e-learning for undergraduate medical education on emergent otorhinolaryngology-head and neck surgery disorders: pilot randomized controlled trial. JMIR Med Educ. 2018;4(1):e8.
- Hassenzahl M, Burmester M, Koller F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: *Mensch & Computer 2003.* edn. Edited by Ziegler J, Szwillus G. Weisbaden, German: Vieweg +Teubner Verlag; 2003: 187–196.
- Renkl A, Stark R, Gruber H, Mandl H. Learning from worked-out examples: the effects of example variability and elicited self-explanations. Contemp Educ Psychol. 1998;23(1):90–108.
- Chi MTH, Bassok M, Lewis MW, Reimann P, Glaser R. Self-explanations: how students study and use examples in learning to solve problems. Cogn Sci. 1989;13(2):145–82.
- Morrison GR, Ross SM, Kemp JE, Kalman H. Designing effective instruction. 8th ed. Hoboken, NJ, U.S.A.: Wiley Inc.; 2019.
- Cabrera-Muffly C, Cusumano C, Freeman M, Jardine D, Lieu J, Manes RP, Marple B, Puscas L, Svrakic M, Thorne M, et al. Milestones 2.0: otolaryngology resident competency in the postpandemic era. Otolaryngol Head Neck Surg. 2022;166(4):605–7.
- Yang YY, Lee FY, Hsu HC, Huang CC, Chen JW, Cheng HM, Lee WS, Chuang CL, Chang CC, Huang CC. Assessment of first-year post-graduate residents: usefulness of multiple tools. J Chin Med Assoc. 2011;74(12):531–8.
- Kara CO, Mengi E, Tumkaya F, Topuz B, Ardic FN. Direct observation of procedural skills in otorhinolaryngology training. Turk Arch Otorhinolaryngol. 2018;56(1):7–14.
- 43. Bui D, Benavides E, Soki F, Ramaswamy V, Kosecki B, Bonine B, Kim-Berman H. A comparison of virtual reality and three-dimensional multiplanar educational methods for student learning of cone beam computed tomography interpretations. J Dent Educ. 2024;88(11):1572–81.
- Paas FG, Van Merrienboer JJ, Adam JJ. Measurement of cognitive load in instructional research. Percept Mot Skills. 1994;79(1 Pt 2):419–30.
- Xiao YM, Wang ZM, Wang MZ, Lan YJ. [The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index]. Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi. 2005;23(3):178–81.
- Naismith LM, Haji FA, Sibbald M, Cheung JJ, Tavares W, Cavalcanti RB. Practising what we preach: using cognitive load theory for workshop design and evaluation. Perspect Med Educ. 2015;4(6):344–8.
- 47. Hsin LJ, Chao YP, Chuang HH, Kuo TBJ, Yang CCH, Huang CG, Kang CJ, Lin WN, Fang TJ, Li HY et al. Mild simulator sickness can alter heart rate variability, mental workload, and learning outcomes in a 360 degrees virtual reality application for medical education: a post hoc analysis of a randomized controlled trial. Virtual Real 2022:1–17.
- Howell BC, Hamilton DA. Baseline heart rate variability (HRV) and performance during a set-shifting visuospatial learning task: the moderating effect of trait negative affectivity (NA) on behavioral flexibility(). Physiol Behav. 2022;243:113647.
- Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, Schwartz PJ. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;17(3):354–81.

- 50. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health. 2017;5:258.
- Solhjoo S, Haigney MC, McBee E, van Merrienboer JJG, Schuwirth L, Artino AR Jr., Battista A, Ratcliffe TA, Lee HD, Durning SJ. Heart rate and heart rate variability correlate with clinical reasoning performance and self-reported measures of cognitive load. Sci Rep. 2019;9(1):14668.
- 52. Suriya-Prakash M, John-Preetham G, Sharma R. Is heart rate variability related to cognitive performance in visuospatial working memory? Indian J Physiol Pharmacol. 2017;61(1):14–22.
- Durantin G, Gagnon JF, Tremblay S, Dehais F. Using near infrared spectroscopy and heart rate variability to detect mental overload. Behav Brain Res. 2014;259:16–23.
- Thayer JF, Ahs F, Fredrikson M, Sollers JJ 3rd, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev. 2012;36(2):747–56.
- 55. Malinska M, Zuzewicz K, Bugajska J, Grabowski A. Heart rate variability (HRV) during virtual reality immersion. Int J Occup Saf Ergon. 2015;21(1):47–54.
- Tronchot A, Maximen J, Casy T, Common H, Thomazeau H, Jannin P, Huaulme A. The influence of virtual reality simulation on surgical residents' heart rate during an assessment of arthroscopic technical skills: A prospective, paired observational study. Orthop Traumatol Surg Res. 2024;110(8):103915.
- Ingadottir B, Blondal K, Thue D, Zoega S, Thylen I, Jaarsma T. Development, usability, and efficacy of a serious game to help patients learn about pain management after surgery: an evaluation study. JMIR Serious Games. 2017;5(2):e10.
- 58. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Res Psychol. 2006;3(2):77–101.
- Young G, Stehle S, Walsh B, Tiri E. Exploring virtual reality in the higher education classroom: using VR to build knowledge and Understanding. JUCS - J Univers Comput Sci. 2020;26(8):904–28.
- 60. Castro-Alonso JC, de Koning BB, Fiorella L, Paas F. Five strategies for optimizing instructional materials: Instructor- and Learner-Managed cognitive load. Educ Psychol Rev. 2021;33(4):1379–407.
- 61. Coban M, Bolat YI, Goksu I. The potential of immersive virtual reality to enhance learning: A meta-analysis. Educational Res Rev. 2022;36:100452.
- 62. Kavanagh S, Luxton-Reilly A, Wuensche B, Plimmer B. A systematic review of virtual reality in education. Themes Sci Technol Educ. 2017;10(2):85–119.
- 63. Conrad M, Kablitz D, Schumann S. Learning effectiveness of immersive virtual reality in education and training: A systematic review of findings. Computers Education: X Real. 2024;4:100053.
- 64. Buttussi F, Chittaro L. Acquisition and retention of Spatial knowledge through virtual reality experiences: effects of VR setup and locomotion technique. Int J Hum Comput Stud 2023, 177.
- Marougkas A, Troussas C, Krouska A, Sgouropoulou C. How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade. Multimedia Tools Appl. 2023;83(6):18185–233.
- 66. Guay F. Applying self-determination theory to education: regulations types, psychological needs, and autonomy supporting behaviors. Can J School Psychol. 2021;37(1):75–92.
- Ogden K, Kilpatrick S, Elmer S. Examining the nexus between medical education and complexity: a systematic review to inform practice and research. BMC Med Educ. 2023;23(1):494.
- 68. Al-Jundi HA, Tanbour EY. A framework for fidelity evaluation of immersive virtual reality systems. Virtual Reality. 2022;26(3):1103–22.
- Chattha UA, Janjua UI, Anwar F, Madni TM, Cheema MF, Janjua SI. Motion sickness in virtual reality: an empirical evaluation. IEEE Access. 2020;8:130486–99.
- Di Natale AF, Bartolotta S, Gaggioli A, Riva G, Villani D. Exploring students' acceptance and continuance intention in using immersive virtual reality and metaverse integrated learning environments: the case of an Italian university course. Educ Inform Technol. 2024;29:14749–68.
- Rusticus SA, Pashootan T, Mah A. What are the key elements of a positive learning environment? Perspectives from students and faculty. Learn Environ Res. 2023;26(1):161–75.
- 72. Sorqvist P, Dahlstrom O, Karlsson T, Ronnberg J. Concentration: the neural underpinnings of how cognitive load shields against distraction. Front Hum Neurosci. 2016;10:221.
- 73. Wong JT, Chen E, Au-Yeung N, Lerner BS, Richland LE. Fostering engaging online learning experiences: investigating situational interest and mind-wandering as mediators through learning experience design. Educ Inform Technol 2024.

- 74. van Kesteren MTR, Meeter M. How to optimize knowledge construction in the brain. NPJ Sci Learn. 2020;5:5.
- Zhao X, Ren Y, Cheah KSL. Leading virtual reality (VR) and augmented reality (AR) in education: bibliometric and content analysis from the web of science (2018–2022). SAGE Open 2023, 13(3).

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.