

Democratizing Vitreoretinal Surgery Training With a Portable and Affordable Virtual Reality Simulator in the Metaverse

Fares Antaki ¹⁻³, Cedryk Doucet⁴, Daniel Milad ^{2,3}, Charles-Édouard Giguère⁵, Benoît Ozell ⁴, and Karim Hammamji ^{2,3}

¹ The CHUM School of Artificial Intelligence in Healthcare, Montreal, Quebec, Canada

² Department of Ophthalmology, Université de Montréal, Montreal, Quebec, Canada

³ Department of Ophthalmology, Centre Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada

⁴ Department of Computer Engineering and Software Engineering, Polytechnique Montréal, Montreal, Canada

⁵ Institut universitaire en santé mentale de Montréal (IUSMM), Montreal, Quebec, Canada

Correspondence: Karim Hammamji, Department of Ophthalmology, Centre Hospitalier de l'Université de Montréal (CHUM), 1051 Rue Sanguinet, Montreal, QC H2X 3E4, Canada. e-mail: karim.hammamji@gmail.com

Received: January 25, 2024

Accepted: March 6, 2024

Published: April 2, 2024

Keywords: vitreoretinal surgery; surgical simulation; virtual reality (VR); metaverse

Citation: Antaki F, Doucet C, Milad D, Giguère CÉ, Ozell B, Hammamji K. Democratizing vitreoretinal surgery training with a portable and affordable virtual reality simulator in the metaverse. *Transl Vis Sci Technol.* 2024;13(4):5. <https://doi.org/10.1167/tvst.13.4.5>

Purpose: The purpose of this study was to develop and validate RetinaVR, an affordable, portable, and fully immersive virtual reality (VR) simulator for vitreoretinal surgery training.

Methods: We built RetinaVR as a standalone app on the Meta Quest 2 VR headset. It simulates core vitrectomy, peripheral shaving, membrane peeling, and endolaser application. In a validation study ($n = 20$ novices and experts), we measured: efficiency, safety, and module-specific performance. We first explored unadjusted performance differences through an effect size analysis. Then, a linear mixed-effects model was used to isolate the impact of age, sex, expertise, and experimental run on performance.

Results: Experts were significantly safer in membrane peeling but not when controlling for other factors. Experts were significantly better in core vitrectomy, even when controlling for other factors ($P = 0.014$). Heatmap analysis of endolaser applications showed more consistent retinopexy among experts. Age had no impact on performance, but male subjects were faster in peripheral shaving ($P = 0.036$) and membrane peeling ($P = 0.004$). A learning curve was demonstrated with improving efficiency at each experimental run for all modules. Repetition also led to improved safety during membrane peeling ($P = 0.003$), and better task-specific performance during core vitrectomy ($P = 0.038$), peripheral shaving ($P = 0.011$), and endolaser application ($P = 0.043$). User experience was favorable to excellent in all spheres.

Conclusions: RetinaVR demonstrates potential as an affordable, portable training tool for vitreoretinal surgery. Its construct validity is established, showing varying performance in a way that correlates with experimental runs, age, sex, and level of expertise.

Translational Relevance: Fully immersive VR technology could revolutionize surgical training, making it more accessible, especially in developing nations.

Introduction

Virtual reality (VR) simulation in health care has made significant progress over the past 5 decades and is now considered a cornerstone of medical education.¹ In surgery, it enables trainees to acquire skills

in an immersive learning environment that mitigates patient harm. The digital nature of VR also alleviates the ethical and logistic challenges tied to wet laboratory training, while offering an interactive, high-fidelity experience.² In ophthalmology, VR simulation has been shown to improve the performance of novice cataract surgeons and to decrease their complication

rate.^{3,4} Similar trends have been observed in vitreoretinal surgery training, but without definite evidence on skill transfer to the operating room.^{5,6}

The most frequently studied VR simulator in ophthalmology is the EyeSi Surgical Simulator (Haag-Streit Simulation). It comprises a mannequin head, surgical instruments, foot pedals, and a VR interface, accessible through the operating microscope.⁷ Despite its high cost of acquisition (approximately USD \$200,000) and its annual running costs, the use of EyeSi has been shown to be cost-effective for cataract surgery training when considering the reduction of complications.^{4,8,9} However, in developing nations and under-resourced communities, the simulator's cost could pose a significant acquisition barrier. This may disproportionately affect these already vulnerable groups, further exacerbating their risk of adverse health outcomes.¹⁰

Since the 1970s, head-mounted displays (VR headsets) have steadily decreased in weight and improved in computing capacity. VR headsets have moved beyond academic laboratories and are commercially available with prices starting from USD \$299.¹¹ VR headsets offer several benefits over traditional stationary simulators, including portability, improved immersiveness, and multiplayer capabilities through “the metaverse”.¹² This allows multiple users to concurrently use the system and interact together in a virtual environment. By leveraging their existing hardware and software capabilities, VR headsets can democratize access to surgical simulation, making the metaverse a particularly useful space for global ophthalmic education and collaboration.

In this work, we developed a VR simulation application software for vitreoretinal surgery training that is compatible with commercially available VR headsets. RetinaVR is fully immersive, affordable, and portable, as it leverages the powerful processors, cameras, and sensors of the headset without the need for external haptic devices. We focus on four fundamental skills: core vitrectomy, peripheral shaving, membrane peeling, and endolaser application. To our knowledge, this is the first vitreoretinal surgery simulator of its kind.

Methods

We provide an overview of RetinaVR in [Figure 1](#). RetinaVR was developed as a simulation app that is compatible with off-the-shelf hardware. We focused our work on the affordable Meta Quest 2 VR headset (Meta Platforms Inc., Menlo Park, CA, USA), the best-selling VR headset available at the time.¹³ Four

training modules were built to simulate fundamental skills in vitrectomy surgery.

Virtual Reality Hardware

We carried out all development experiments on the wired HP Reverb, attached to an AMD Ryzen 5 computer with 2600x CPU, 16 GB of RAM, and an AMD Radeon RX 5700 XT graphics card. After each version iteration, we adapted the app for the wireless Meta Quest 2 to allow our domain experts to test the software remotely and to provide iterative feedback. To ensure broad applicability, we utilized the standard controllers packaged with the Meta Quest 2 only, rather than exploring add-on external haptic devices.

The Meta Quest 2 is a general-purpose VR headset that allows for a standalone experience, eliminating the need for wiring or a computer connection. This feature renders it apt for surgical simulation training, providing an unencumbered environment conducive to learning. It comes with two light-weight plastic controllers, each weighing approximately 150 grams, that are tracked by the headset's integrated cameras. The controllers are designed to rest within the curve of the user's palm, allowing the user's fingers to engage with the capacitive face, grip, and trigger buttons, as well as the joystick.

Virtual Reality Software

We developed RetinaVR in the Unity 3D game engine. To represent the eye, a virtual sphere was created, and a custom-made fundus illustration was fitted on its inner surface. The virtual instruments (light pipe and vitrector/endolaser) were controlled using standard controllers without the use of a physical eye model. The fulcrum effect was challenging to reproduce due to the lack of haptic feedback from the virtual eye and the disconnect between the two controllers. As such, only one controller could be used to move the eye. For all tasks, the left controller was used as a light pipe, whereas the right controller served as a vitrector or an endolaser probe, and controlled eye movements. The virtual instruments' position and their movements were rotated 45 degrees on the x-axis to allow for ergonomic holding of the controllers. To enhance the realism of the simulation, we added the characteristic sound emission produced by the pneumatic guillotine cutter (recorded at 7500 cuts per minute) during core vitrectomy and peripheral shaving.¹⁴ We also added a laser sound to the endolaser application module. The software development methodology is detailed in

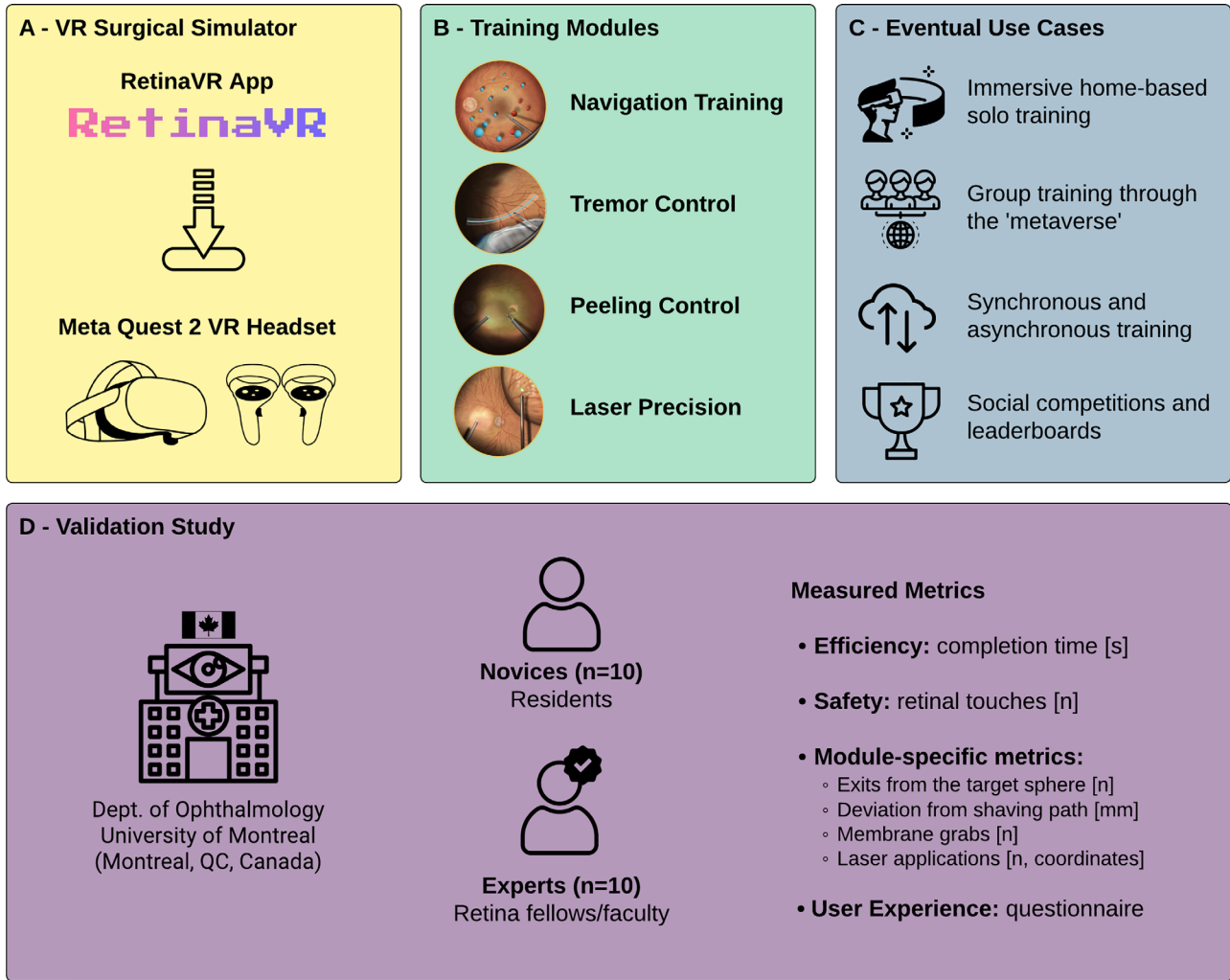


Figure 1. Overview of the RetinaVR development and validation framework. (A) RetinaVR was developed in the Unity 3D game engine and deployed as an “app” on the Meta Quest 2 VR headset. (B) Four training modules simulating fundamental skills in vitrectomy surgery were developed: core vitrectomy (Navigation Training), peripheral shaving (Tremor Control), membrane peeling (Peeling Control), and endolaser application (Laser Precision). (C) Multiple potential use cases were considered as rationale for selecting the app format and the standalone VR headset. Those included the possibility for home-based solo training, synchronous and asynchronous group training through the metaverse, and social competitions and score leaderboards. (D) To determine construct validity, we designed a prospective validation study comparing the performance of novice ($n = 10$) and expert users ($n = 10$) recruited from the University of Montreal in Montreal, Quebec, Canada. We analyzed numerous metrics including efficiency, safety, and module-specific performance, in relation to their level of expertise and demographic factors.

Supplementary A1 and the simulation ergonomics are shown in Supplementary Figure S1.

an expert are shown head-to-head in Supplementary Video S1.

Training Modules

We focused on four fundamental vitreoretinal surgery skills to devise four corresponding training tasks: core vitrectomy (Navigation Training), peripheral shaving (Tremor Control), membrane peeling (Peeling Control), and endolaser application (Laser Precision). Screenshots from each of the modules are shown in Figure 2. Sample runs from a novice and

1. **Navigation Training:** To assess navigation skills in the vitreous body, a sphere collection exercise was designed using the “Collision detection” module in Unity. Initially red, the spheres turn green when collected. To collect a sphere, the tip of the vitrector must maintain contact with it for 2 consecutive seconds (determined heuristically) for it to disappear. The exercise concludes once

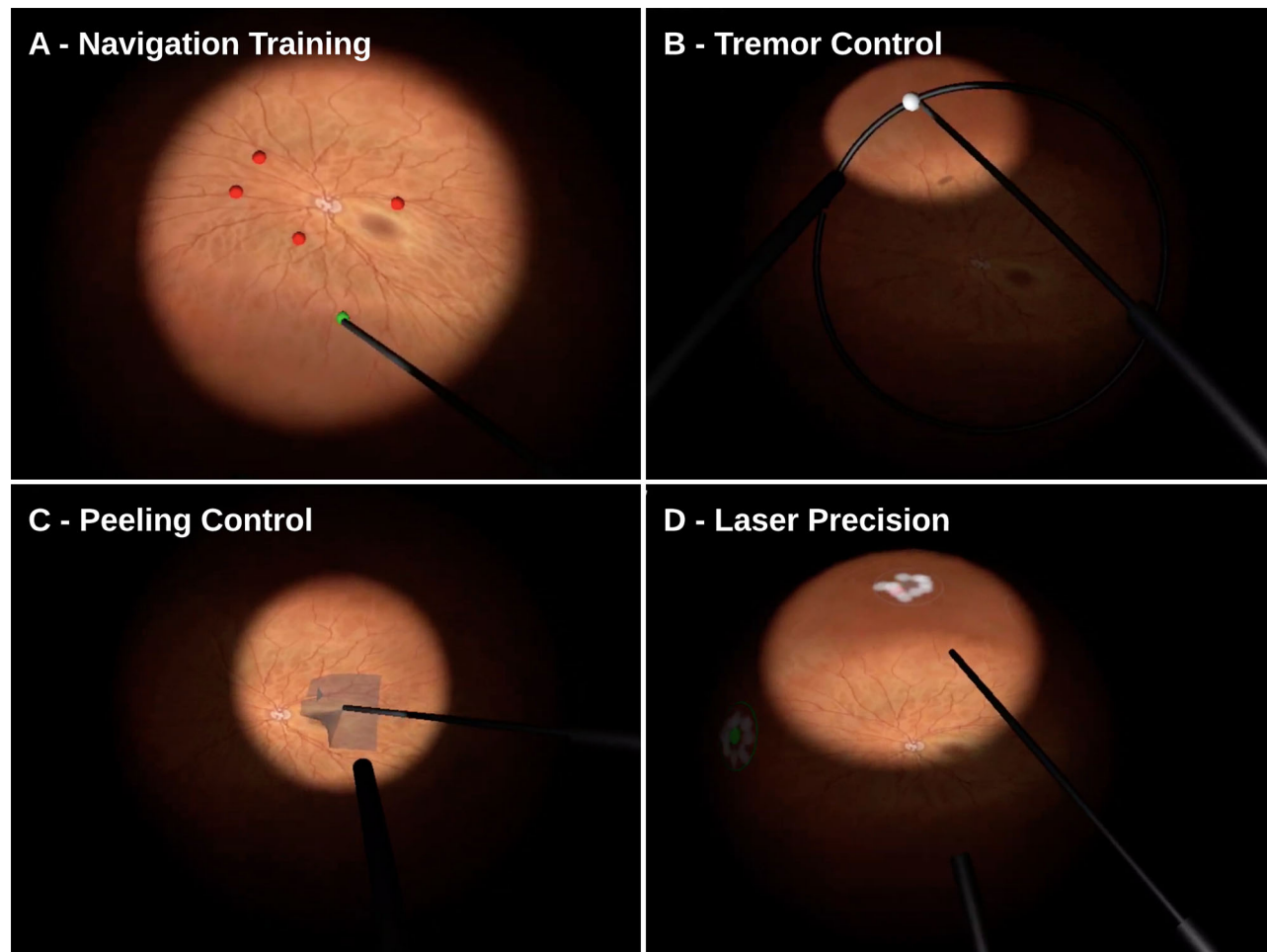


Figure 2. In-game screenshots from the RetinaVR modules. (A) Navigation Training simulates core vitrectomy. The goal of the user is to collide with all red spheres, maintain the vitrector in the sphere, and turn them green. (B) Tremor Control simulates peripheral shaving. The user will engage the tip of the vitrector with a sphere, allowing it to move along a predetermined path. (C) Peeling Control simulates membrane peeling. The user will grasp the membrane by pressing the grip button on the controller before peeling it away from the macula. (D) Laser Precision simulates endolaser application. The user is asked to treat five retinal breaks by applying laser spots to a surrounding donut. A green marker will indicate a fully treated tear.

all 10 spheres at varying depths within the vitreous body are collected.

2. **Tremor Control:** The user's ability to control the vitrector during peripheral shaving is assessed by moving a target sphere along a predetermined path. When the tip of the vitrector collides with the sphere, it causes the sphere to move along the path until the instrument loses contact with the sphere. The goal is to move the sphere along the path, without deviating, as smoothly as possible without touching the retina.
3. **Peeling Control:** This exercise simulated peeling epiretinal membranes using a cutter-based approach (rather than forceps).¹⁵ The objective was to peel the membrane completely from the retina without iatrogenic touch. Users could

enlarge their view by pressing the "X" button on the left controller, simulating a magnifying lens. To grab the membrane, users were required to press the right grip button. The membrane could only be peeled if a neighboring border was detached, requiring multiple grasps.

4. **Laser Precision:** This exercise focused on applying endolaser around five retinal breaks in the periphery. The laser probe had a traditional red spot that varied in size based on its distance to the retina. As in real life, this affected the laser uptake, with larger spots being less intense. The laser was applied by pressing the grip button. Repeat mode was available by holding the button, with an interval of 200 ms. When a tear was considered fully treated, it turned green, signaling to the user to

move on to the next break. During development, to ensure that tears were fully treated, we used a raycasting approach in Unity and heuristically adjusted the threshold for “fully treated” until we achieved our desired goal of two rows of laser spots 360 degrees around each break.

Validation Study

After 2 years of development, we locked RetinaVR in March 2023 to prepare it for human validation. Novices and experts were recruited from the Department of Ophthalmology of the University of Montreal in Quebec, Canada, from April 2023 through October 2023. The “Novice” group included ophthalmology residents in their first, second, or third years of residency and who have not had any hands-on exposure to intraocular surgery. Exposure to oculoplastic and strabismus surgery, and previous VR exposure (other than RetinaVR) were not exclusionary. The “Expert” group included experienced fellowship-trained vitreoretinal surgeons and vitreoretinal surgery fellows. Determining an a priori sample size was challenging due to the lack of existing data on the expected performance differences between novices and experts for our newly developed modules. To address this, we recruited all available retina surgeons at our institution, matching them with an equal number of novices, leading to a total sample size of 20. We felt that this pragmatic approach was reasonable considering the exploratory nature of this work. Our sample size was on par with most studies looking at the validation of existing VR simulation tools in vitreoretinal surgery.⁵ We excluded participants if they had any contraindications for VR gaming, including seizure disorder, vertigo, motion sickness, and known VR cybersickness. This research was conducted in full compliance with the ethical principles outlined in the Declaration of Helsinki. We obtained ethics approval by the Institutional Review Boards of the CHUM Hospital (IRB # 2023-10479-22.035). Informed consent was obtained from all participants after detailing the nature of the study.

Novices and experts were scheduled to test the portable RetinaVR simulator on a Meta Quest 2 headset at their convenience. We often tested in a conference room, requiring only a flat surface. Our lead technical and clinical experts were available during testing, casting the user’s view to a connected computer. Before each recorded test, users received a brief explanation of the tasks while wearing the VR headset. They were also instructed on how to position their hands and calibrate the instruments, and they were allowed a single trial run of each module. A life-

size soft silicone doll head simulated the patient’s head, allowing users to rest their wrists. All users sat superiorly relative to the eye.

Collected Data

To determine the construct validity of our simulator, we needed to study the impact of user factors, like age, self-reported sex, and level of expertise on simulation performance. We collected all possible measurable performance metrics directly from RetinaVR, using built-in code. All modules were evaluated based on three criteria: Efficiency, Safety, and Module-specific performance. For all modules, Efficiency was assessed by measuring completion time in seconds, whereas Safety was assessed by counting the number of iatrogenic retinal touches. Module-specific performance metrics varied depending on the module. In Navigation Training, the number of exits from the target sphere was counted. For Tremor Control, the number of exits from the target sphere was counted, along with the mean and maximum deviation from the shaving path in millimeters. In Membrane Peeling, the number of membrane grasps was counted, with the hypothesis that the number of grasps would vary with experience and technique. For Laser Precision, the number of laser spots was recorded, with the hypothesis that a parsimonious use of laser was better as long as the tears were treated.¹⁶ The precise coordinates of the laser spots around tears were also recorded to determine the treatment pattern.

User Experience

To measure user experience (UX), we administered a French abbreviated version of the validated Immersive Virtual Environments Questionnaire (IVEQ) version 2.¹⁷ The questionnaire consisted of 26 questions: 2 to gauge the user’s prior experience with VR, 21 that were gradable using a 10-point Likert scale, and 3 open-ended questions for general comments and feedback. The gradable questions assessed a broad range of UX factors, including *Presence* ($n = 3$), *Engagement* ($n = 2$), *Immersion* ($n = 2$), *Flow* ($n = 2$), *Emotion* ($n = 2$), *Skill* ($n = 2$), *Judgment* ($n = 3$), *Experience Consequence* ($n = 2$), and *Technology Adoption* ($n = 3$). The questionnaire is available in Supplementary A2. The three open-ended questions aimed to gather positive feedback, negative feedback, and suggestions for improvement. To analyze the free text responses, we elucidated the prevalent themes from each response and then consolidated them into broad categories. Once a coherent representation of data

across all participants was achieved, the frequency of each theme was recorded and summarized.

Statistical Methods

We hypothesized that expert performance would significantly differ from novice performance, and that demographic factors, such as age and sex, along with the experimental run, may influence these performance outcomes. We first explored the unadjusted differences in performance between novices and experts by calculating the standardized mean difference for each performance metric. This effect size analysis was useful to contextualize our findings, given the disparate units (count-, time-, and distance-based) and varying scales of the performance metrics stemming from the differing difficulty levels of the four training modules. Additionally, given the novelty of our experimental design, the lack of existing normative data to define “good or bad” or “fast or slow” performance also necessitated this scaled analysis. We interpreted the effects as follows: 0.01 to 0.19 (minimal), 0.20 to 0.49 (small/mild), 0.50 to 0.79 (medium/moderate), 0.80 to 0.99 (large), and >1.0 (very large).¹⁸

We then carried out an adjusted analysis and explored differences between novices and experts while controlling for age, sex, and experimental run – factors that can influence the VR gaming experience.^{19–21} We used a linear mixed-effect model, which allowed us to isolate the effect of each factor while controlling for all others. All analyses were performed in R version 4.3.1 at a 5% alpha level.

Results

Baseline Demographics

We recruited 20 participants, including 10 novices and 10 experts. Their baseline and demographic characteristics are detailed in Table 1. Novices were

significantly younger and predominantly female subjects. Novices had no prior surgical experience (a selection criterion), whereas the experts, on average, had 16.6 years (10.71) of post-residency surgical experience. Novices reported more hours of VR gaming than experts, but this difference was not statistically significant. They also reported more hours of training on VR-based surgical simulators than experts, with this difference being statistically significant. We provide descriptive statistics on the performance of novices and experts across all runs and modules in Supplementary Tables S1 to S4.

Impact of the Expertise

We first compared the performance of novices and experts using an unadjusted model. The results are summarized in Figure 3. The detailed effect size analyses are shown in Supplementary Table S5. The linear mixed-effects model results are summarized in Table 2.

Regarding efficiency, we found trends that novices were slower than experts, except in membrane peeling. None of those effects were statistically significant when all experimental runs were combined. When examining the runs individually, we found that novices were faster than experts in the first membrane peeling run (very large effect = -1.10, 95% confidence interval [CI] = -2.03 to -0.14). In the linear mixed-effects model, when controlling for age, sex, and experimental run, the trends were maintained, but we found no statistically significant difference in efficiency between novices and experts in any of the modules.

Regarding safety, we found that experts were safer than novices in the membrane peeling module when all experimental runs were combined (very large effect = 1.06, 95% CI = 0.52 to 1.60). This effect was also present in the first (very large effect = 1.34, 95% CI = 0.35 to 2.31) and second run (very large effect = 1.12, 95% CI = 0.16 to 2.05), but not the third run. We also found trends that the experts were safer in all other modules, but those differences were not statis-

Table 1. Demographic and Baseline Characteristics of the Novice and Expert Users

Characteristic	Novices (n = 10)	Experts (n = 10)	Difference
Age, y	28.2 (3.61)	47.1 (12.3)	P = 0.001
Female sex – n	7 (70%)	3 (30%)	P = 0.0736
Surgical expertise – y	0 (0)	16.6 (10.71)	P = 0.001
Previous VR gaming – h	6.7 (12.82)	2.1 (3.6)	P = 0.299
Previous VR surgical training – h	22.6 (23.29)	3.8 (6.53)	P = 0.033

Values are mean (SD) unless otherwise specified.
VR, virtual reality.

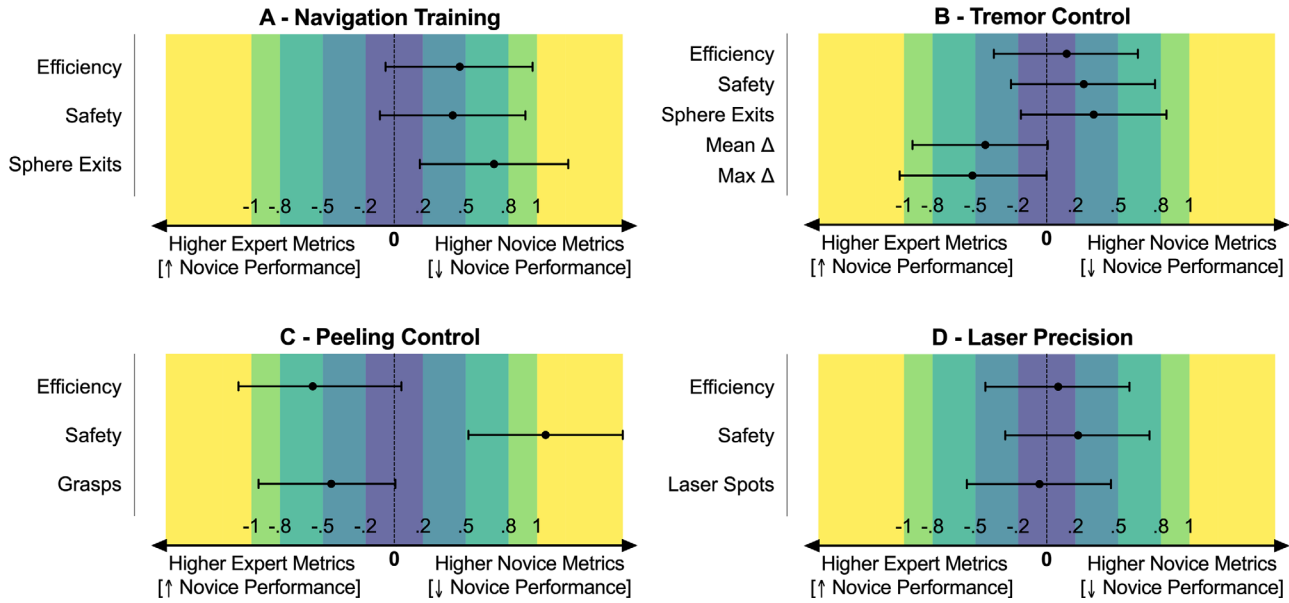


Figure 3. Forest plot showing the unadjusted effect sizes of efficiency, safety and task-specific performance between experts and novices. The point effect estimate is Cohen's D and represents the standardized mean difference between novice and expert performance, along with 95% confidence intervals. Positive effects, represented by values to the right of the y-axis, indicate higher novice metrics (e.g. longer novice completion times, and more novice retinal touches), suggesting lower novice performance. Conversely, negative effects, represented by values to the left of the y-axis, indicate higher expert metrics (e.g. longer expert completion times, and more expert retinal touches), implying better novice performance. The viridis color palette is used to interpret the effect sizes, representing a minimal effect (0.01–0.19), a small or mild effect (0.20–0.49), a medium or moderate effect (0.50–0.79), a large effect (0.80–0.99), and a very large effect (>1.0).

tically significant. In the linear mixed-effects model, when controlling for age, sex, and experimental run, the trends were maintained, but we found no statistically significant difference in safety between novices and experts in any of the modules.

Regarding task-specific performance, we found that experts performed better in the core vitrectomy module, demonstrating significantly fewer exits from the target spheres (moderate effect = 0.7, 95% CI = 0.18 to 1.22). This effect was mostly driven by the second experimental run (very large effect = 1.11, 95% CI = 0.15 to 2.04). In the linear mixed-effects model, that difference was maintained while controlling for all other user factors, with novices exiting the spheres an excess of 21.46 times ($P = 0.014$). In peripheral shaving, we found trends that novices had more sphere exits than experts while demonstrating less deviation from the shaving path, but those differences were not statistically significant. In the linear mixed-effects model, when controlling for other factors, those differences were also not statistically significant.

In membrane peeling, we found trends that experts grasped the membrane more times than novices, but that difference was not statistically significant in the unadjusted model. The trend was maintained in the linear mixed-effects model, but the difference was not statistically significant. In endolaser application,

we found no difference in the amount of laser used between novices and experts in the adjusted and unadjusted models. However, a heatmap analysis of the laser spot distribution showed clinically significant differences in treatment patterns among novices and experts, as shown in Figure 4.

Impact of Participant Age and Sex

We evaluated the impact of participant age and sex on their performance, while controlling for experimental run and expertise. As shown in Table 3, in the linear mixed-effects model, age had no impact on performance in any of the modules. Male participants were 12.35 seconds faster in peripheral shaving ($P = 0.036$) and 32.21 seconds faster in membrane peeling ($P = 0.004$) compared to women. We observed trends of male participants being more efficient, safer, and performing better in most task-specific metrics, but none of those effects were statistically significant.

Impact of the Learning Curve

We also evaluated the learning curve by repeating the experiments three times for each participant. As shown in Table 4, in the linear mixed-effects model, efficiency improved with each experimental run during

Table 2. Linear Mixed-Effects Model (Adjusted) for the Impact of Expertise on Performance

Module and Metric	Expertise	
	Estimate	P Value
Navigation training		
Efficiency	9.30	0.50
Safety	0.63	0.50
Sphere exits	21.46	0.014
Tremor control		
Efficiency	7.05	0.29
Safety	0.10	0.94
Sphere exits	38.14	0.29
Mean Δ	0.00	0.81
Max Δ	−0.02	0.71
Membrane peeling		
Efficiency	−16.42	0.17
Safety	2.90	0.11
Grasps	−1.03	0.25
Laser precision		
Efficiency	16.68	0.29
Safety	0.63	0.29
Laser spots	10.15	0.64

This model controls for experimental run, user age and sex. The only significant effect is the difference in performance during Navigation Training. Novices had an excess of 21.46 sphere exists compared to experts ($P = 0.014$). Efficiency estimates are in seconds, and Safety estimates are in number of iatrogenic retinal touches. Sphere exits, number of grasps, and laser spots are count data. Mean and maximal deviation metrics are provided in meters in this table.

all modules. At each run, completion time decreased by 7.67 seconds for core vitrectomy ($P = 0.005$), 12.02 seconds for peripheral shaving ($P < 0.001$), 17.92 seconds for membrane peeling ($P < 0.001$), and 25.68 seconds for endolaser application ($P < 0.001$). We found that repetition improved safety scores during membrane peeling, with 1.37 fewer iatrogenic retinal touches with each run ($P = 0.003$). Similar trends were observed for all modules, but the effects were not statistically significant. However, it did reduce the number of laser spots used by the participants. At each run, the number of sphere exits decreased by 5.42 times ($P = 0.038$) in core vitrectomy and by 17.00 times during peripheral shaving ($P = 0.011$). In endolaser application, participants used 11.20 less laser shots at each run to treat the tears ($P = 0.043$).

User Experience

Overall, the users rated the experience from favorable to excellent in all 8 spheres of UX, as shown in

Supplementary Table S6. Positive feedback predominantly centered on three themes: the realistic 3D environment ($n = 18$), the ability to practice in a low-risk environment ($n = 9$), and the authentic representation of the vitrectomy experience ($n = 5$). Other sporadic comments praised the innovation, immersion, and portability of the experience. Negative feedback mentioned the fulcrum effect and controller-simulation movement translation ($n = 8$), the controller size and ergonomics ($n = 6$), and difficulty with visualization and depth perception ($n = 6$). Other comments included the lack of progress indicators, headset fit, and unrealistic shaving module. Suggestions for improvement suggested improving the controllers and ergonomics ($n = 10$), providing better instructions and real-time feedback ($n = 5$), and improving movement translation ($n = 5$). It was also recommended to attempt to improve headset fit, build more complete case-based modules, improve graphics, and gamify the experience.

Discussion

We built a RetinaVR, a fully immersive, affordable, and portable VR simulator for vitreoretinal surgery training. RetinaVR is a standalone app that leverages the powerful processors and cameras of commercially available VR headsets and controllers, without relying on external touch haptic devices. RetinaVR is a proof of concept for a new way of approaching surgical simulation in the metaverse, at a fraction of the cost of traditional VR simulators. It democratizes access to surgical simulation, and has the potential to spur innovation in global ophthalmology.

To ensure RetinaVR’s affordability and accessibility, we designed it to require only a quick app download. The app is a mere 100 megabytes, taking approximately 20 seconds to download on average global broadband speeds and less than 2 minutes in Sub-Saharan Africa.^{22,23} To simulate surgical instruments, we used the standard built-in controllers, rather than integrating custom hardware. Using pen-like haptic feedback devices could have provided a more faithful simulation of instruments, but it would have come at a high cost.²⁴ Because our simulator did not require instruments to be anchored to a physical eye model, the fulcrum effect was difficult to simulate. This effect, encountered when using the vitrector and light pipe through a trocar, necessitates unique skills to move the instrument tips. Tactile feedback could not be provided without haptic devices, thereby limiting the surgeon-eye interactions to visual cues only. Despite that, we feel

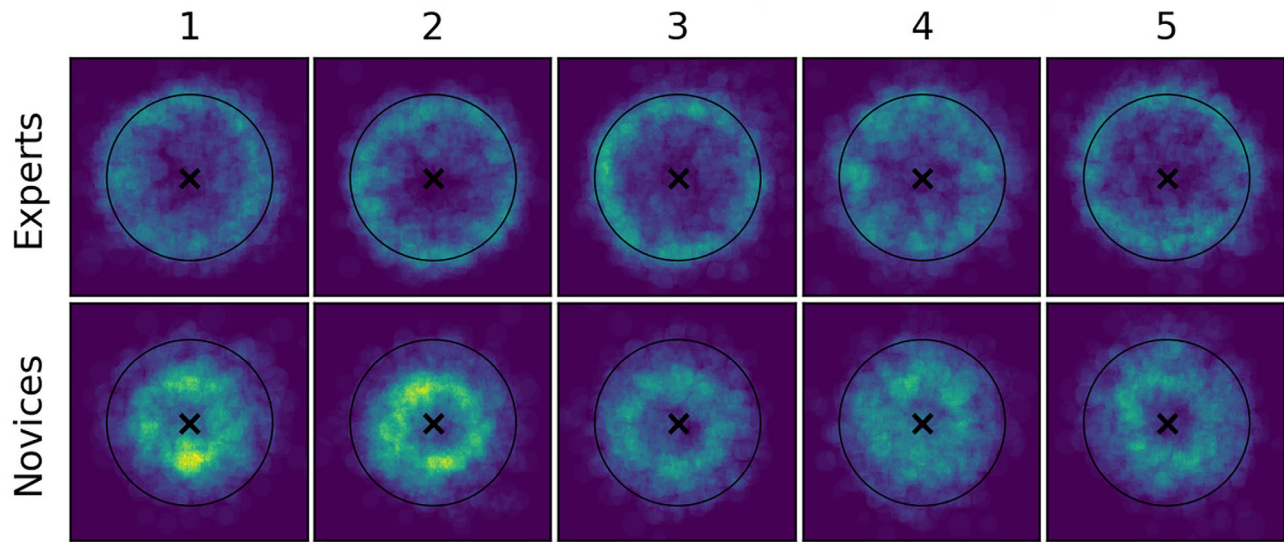


Figure 4. Heatmap of laser shot distribution in the endolaser application (Laser Precision) module. Heatmap illustrating laser treatment patterns for all five retinal tears, differentiated by novices and experts. Each square represents a unique tear, with color intensity corresponding to the number of laser spots applied, using the viridis color palette. The color gradient ranges from *purple* (least density) to *bright yellow* (highest density). The central target represents the center-point of the retinal break, as shown in the Laser Precision module. Experts showed a uniform distribution of laser spots, characterized by a consistent spread around each tear, maintaining a uniform distance from the central point. There is a ring-like pattern with minimal laser applications directly on the tears. In contrast, novices showed a more erratic pattern (particularly in tears 1 and 2), with a concentration of laser spots towards the center-point of each tear. This indicates a less controlled application, resulting in a scattered distribution with variable intensity and less discernible uniformity.

that we accurately replicated the motion inversion and scaled motion required to move the vitrector tip, allowing the users to successfully complete the modules and improve at each run. This is supported by the demonstration of the learning curve and the high scores for the *Flow* theme in the UX questionnaire. The users did suggest, however, improvements in instrumentation. Although our plastic controllers were lightweight, they were still considerably heavier than conventional surgical instruments. Their weight was 4 times that of a typical 23G vitrector. For comparison, the Bi-Blade vitrectomy cutter weighs approximately 37 grams with the tubing (personal communication with Bausch + Lomb).

To capture user performance during simulation, we were faced with two options: either collect as many metrics as possible and analyze them post hoc, or develop a scoring system by assigning weights to measurable metrics based on our subjective assessment of their importance. The latter approach raised concerns about how to objectively measure task efficiency, safety, and good performance, and how to determine the appropriate point deductions for mistakes. Given the potential for heuristic bias, we chose the first option and developed code in RetinaVR to quantify those metrics. We conducted a rigorous analysis of the data through an effect size analysis.

This was crucial for interpreting the significance of observed differences, because these experiments were being conducted for the first time with no normative databases to establish good or poor performance benchmarks. We then built an adjusted model to examine the impact of age, sex, and experimental run on performance, and controlled for those factors when comparing novices and experts.

We believe to have demonstrated construct validity.^{25,26} This refers to the ability of RetinaVR to measure user behaviors and performance in a way that correlates with their inherent factors and level of expertise. We found that participant age had no impact on overall performance when we controlled for sex, expertise, and experimental run. However, we found that male participants performed membrane peeling and peripheral shaving tasks more quickly than female participants, with no significant differences in safety and task-specific performance. Some evidence suggests that gaming proficiency may decline with age and show differences between sexes.^{21,27–29} However, we believe that this phenomenon is more likely attributable to a disparity in prior gaming experience, rather than innate age or sex-related abilities. These effects may be even less pronounced in a surgical simulation context like ours, where older participants typically have more prior surgical experience. In parallel, we found that repetition

Table 3. Linear Mixed-Effects Model (Adjusted) for the Impact of Age and Sex on Performance

Module and Metric	Age		Sex	
	Estimate	P Value	Estimate	P Value
Navigation training				
Efficiency	0.29	0.64	-18.96	0.11
Safety	0.00	0.93	-0.17	0.83
Sphere exits	0.72	0.06	-12.71	0.07
Tremor control				
Efficiency	0.52	0.09	-12.35	0.036
Safety	0.00	0.94	-1.13	0.32
Sphere exits	2.11	0.20	-57.02	0.07
Mean Δ	0.00	0.63	0.00	0.95
Max Δ	0.00	0.54	0.00	0.94
Membrane peeling				
Efficiency	0.70	0.19	-32.21	0.004
Safety	-0.03	0.68	-0.54	0.71
Grasps	0.01	0.79	-1.13	0.14
Laser precision				
Efficiency	0.99	0.17	-13.28	0.32
Safety	0.00	0.91	0.61	0.23
Laser spots	0.78	0.43	-5.90	0.75

This model controls for all other variables including expertise when examining the effect of age and sex. Age had no impact on performance, but sex did. Male participants were 12.35 seconds faster in peripheral shaving ($P = 0.036$) and 32.21 seconds faster in membrane peeling ($P = 0.004$). Efficiency estimates are in seconds, and Safety estimates are in number of iatrogenic retinal touches. Sphere exits, number of grasps, and laser spots are count data. Mean and maximal deviation metrics are provided in meters in this table.

boosted efficiency in all modules, and enhanced safety in the membrane peeling module. It also improved task-specific performance during core vitrectomy and peripheral shaving. This demonstrates a learning curve across experimental runs – with users getting better with repetition or practice. We feel that this observation reinforces the notion that user performance was not a random occurrence but rather a reflection of genuine learning. This learning curve has also been demonstrated for the vitreoretinal modules of the EyeSi simulator in numerous studies.^{30,31}

A crucial aspect of this project is the demonstration of how user expertise affects performance. We report on several notable findings in our work. First, novices tended to be slower in all modules, except in membrane peeling. Interestingly, in membrane peeling, they tended to be faster, while also being less safe, causing significantly more iatrogenic retinal touches, and grasping the membranes less frequently. These contrasts possibly highlight the influence of real-world surgical experience. Experts demonstrated a more

Table 4. Linear Mixed-Effects Model (Adjusted) for the Impact of Experimental Run On Performance

Module and Metric	Experimental Run	
	Estimate	P Value
Navigation training		
Efficiency	-7.67	0.005
Safety	-0.20	0.26
Sphere exits	-5.42	0.038
Tremor control		
Efficiency	-12.02	<0.001
Safety	-0.47	0.13
Sphere exits	-17.00	0.011
Mean Δ	0.00	1.00
Max Δ	-0.01	0.22
Membrane peeling		
Efficiency	-17.92	<0.001
Safety	-1.37	0.003
Grasps	0.27	0.23
Laser precision		
Efficiency	-25.68	<0.001
Safety	-0.02	0.92
Laser spots	-11.20	0.043

This model controls for age, sex, and expertise when examining the role of the experimental run. At each run, completion time decreased by 7.67 seconds for core vitrectomy ($P = 0.005$), 12.02 seconds for peripheral shaving ($P < 0.001$), 17.92 seconds for membrane peeling ($P < 0.001$), and 25.68 seconds for endolaser application ($P < 0.001$). Furthermore, repetition improved safety scores during membrane peeling, with 1.37 fewer iatrogenic retinal touches with each run ($P = 0.003$). At each run, the number of sphere exits decreased by 5.42 times ($P = 0.038$) in core vitrectomy and by 17.00 times during peripheral shaving ($P = 0.011$). In endolaser application, participants used 11.20 less laser shots at each run to treat the tears ($P = 0.043$). Efficiency estimates are in seconds, and Safety estimates are in number of iatrogenic retinal touches. Sphere exits, number of grasps, and laser spots are count data. Mean and maximal deviation metrics are provided in meters in this table.

cautious and deliberate approach, peeling slowly and carefully to minimize shearing forces on the macula. In contrast, novices, perhaps viewing the simulation as such, exhibited riskier behavior by attempting to complete the module at a faster pace, leading to more iatrogenic damage. Second, experts performed significantly better in the core vitrectomy module, exhibiting fewer target sphere exits – a difference that was maintained when controlling for other user factors. Third, during endolaser application, we found clinically important differences in the treatment patterns between novices and experts. This speaks to the

construct validity of those modules and their ability to faithfully simulate the surgical experience.

RetinaVR marks a proof of concept for a novel type of platform for vitreoretinal surgery training simulation. We believe that RetinaVR can change the scope of surgical simulation in a number of ways. First, trainees can conveniently access RetinaVR using their personal headsets, integrating it alongside their existing VR-based entertainment, gaming, or sports activities. Second, residency programs can effectively train multiple residents simultaneously by investing in multiple affordable VR headsets. The platform's online metaverse integration, relying on Meta's cloud servers, enables multiplayer group training sessions, connecting residents virtually with expert surgeons from around the world, breaking down geographic barriers and fostering a global learning community. Third, the platform allows for both synchronous and asynchronous learning, which enables trainees to obtain real-time feedback from mentors while also catering for individual learning styles and schedules. Finally, gamification elements, such as points, badges, and international leaderboards, can further enhance engagement and encourage healthy competition, spurring innovation and collaboration in the field of vitreoretinal surgery.

Although RetinaVR has demonstrated construct validity to a certain extent, our work has some limitations and further validation is necessary. First, statistical significance in our analyses was limited by the low sample size and high variance among novices. Despite that, most of our effects were congruent with the expected behaviors of novices and experts. Second, we have not yet demonstrated skill transfer to the operating room, a crucial step in validating a surgical simulator. However, to our knowledge, in vitreoretinal surgery, this has not been shown even for popular simulators like the EyeSi.⁵ RetinaVR remains a work in progress: the user interface, including menu appearances, profile creation, login functionality, and leaderboards, require further development before public release. We are also working on incorporating feedback from this study to determine future directions for RetinaVR. Despite these limitations, we are proud of what was achieved with limited resources. RetinaVR serves as a proof of concept for developing affordable VR surgical simulation apps in an academic laboratory setting, fostering innovation in surgical training and medical education. Driven by the relentless innovation of industry titans like Meta and Apple, we are confident that standalone VR headsets will soon reach a high level of maturity.³² These future headsets may offer superior hand tracking capabilities, enabling the use of RetinaVR without traditional controllers. By

integrating inexpensive 3D-printed instruments and a physical eye model, they may replicate the physical interaction between the surgeon and the eye – a crucial element in vitreoretinal surgery. This will pave the way for the widespread availability of an off-the-shelf, affordable, and validated RetinaVR simulator, empowering the trainees worldwide with an immersive surgical training experience.

Acknowledgments

The authors express their gratitude to the Canadian Ophthalmological Society and Bayer Inc. for funding this work.

Supported by the Innovation in Retina Research Award, granted in June 2021 during the Canadian Ophthalmological Society (COS) Annual Meeting. The project was awarded the First Prize (CAD \$35,000) and the Audience Award (CAD \$5000). The award is a joint venture between the COS and Bayer Inc.

Data Sharing Statements: All data produced in the present study are available upon reasonable request to the authors. RetinaVR is not currently in the Oculus store.

Author Contributions: F.A., C.D., B.O., and K.H. conceptualized the study and designed the experiments. F.A. and K.H. obtained the funding. C.D. and B.O. designed RetinaVR software. F.A. and K.H. provided continuous iterative feedback to improve RetinaVR. F.A., C.D., and D.M. carried out the clinical validation study. C.E.G. performed the statistical analyses. F.A. and C.D. drafted the initial manuscript. F.A. and C.D. designed the figures and tables. All authors reviewed and discussed the results. All authors edited and revised the manuscript before approving the final version of this manuscript.

Disclosure: F. Antaki, Bayer (F, funding for this work); C. Doucet, None; D. Milad, None; C.-E. Giguère, None; B. Ozell, None; K. Hammamji, None

References

1. Pottle J. Virtual reality and the transformation of medical education. *Future Healthc J.* 2019;6:181–185.
2. Mao RQ, Lan L, Kay J, et al. Immersive virtual reality for surgical training: a systematic review. *J Surg Res.* 2021;268:40–58.

3. Thomsen ASS, Smith P, Subhi Y, et al. High correlation between performance on a virtual-reality simulator and real-life cataract surgery. *Acta Ophthalmol.* 2017;95:307–311.
4. Ferris JD, Donachie PH, Johnston RL, et al. Royal College of Ophthalmologists' National Ophthalmology Database study of cataract surgery: report 6. The impact of EyeSi virtual reality training on complications rates of cataract surgery performed by first and second year trainees. *Br J Ophthalmol.* 2020;104:324–329.
5. Rasmussen RC, Grauslund J, Vergmann AS. Simulation training in vitreoretinal surgery: a systematic review. *BMC Ophthalmol.* 2019;19:90.
6. Jaud C, Salleron J, Cisse C, et al. EyeSi Surgical Simulator: validation of a proficiency-based test for assessment of vitreoretinal surgical skills. *Acta Ophthalmol.* 2021;99:390–396.
7. Carr L, McKechnie T, Hatamnejad A, Chan J, Beattie A. Effectiveness of the Eyesi Surgical Simulator for ophthalmology trainees: systematic review and meta-analysis [published online ahead of print April 20, 2023]. *Can J Ophthalmol.* 2023. Available at: <http://dx.doi.org/10.1016/j.cjco.2023.03.014>.
8. Kaur S, Shirodkar A-L, Nanavaty MA, Austin M. Cost-effective and adaptable cataract surgery simulation with basic technology. *Eye.* 2022;36:1384–1389.
9. la Cour M, Thomsen ASS, Alberti M, Konge L. Simulators in the training of surgeons: is it worth the investment in money and time? 2018 Jules Gonin lecture of the Retina Research Foundation. *Graefes Arch Clin Exp Ophthalmol.* 2019;257:877–881.
10. Oseni J, Adebayo A, Raval N, et al. National access to EyeSi Simulation: a comparative study among U.S. ophthalmology residency programs. *Nepal J Ophthalmol.* 2023;15:e112–e118.
11. Anon. Meta Quest 2: immersive all-in-one VR headset. Available at: <https://www.meta.com/us/quest/products/quest-2/>. [Accessed November 14, 2023].
12. Chengoden R, Victor N, Huynh-The Thien, et al. Metaverse for healthcare: a survey on potential applications, challenges and future directions. arXiv Preprint [csAI] 2022. Available at: <http://arxiv.org/abs/2209.04160>.
13. Armstrong M. *Meta leads the way in VR headsets.* Statista; 2023. Available at: <https://www.statista.com/chart/29398/vr-headset-kpis/> [Accessed November 17, 2023].
14. Ruparelia S, Orr S, Choudhry N, et al. Risk for surgical team hearing loss with vitrectomy. *J Vitreoretin Dis.* 2023;7:397–403.
15. Anon. No more forceps: a cutter-based approach to ILM peeling - retina today. Available at: <https://retinatoday.com/articles/2023-may-june/no-more-forceps-a-cutter-based-approach-to-ilm-peeling>. [Accessed November 23, 2023].
16. Wang JC, Ryan EH, Ryan C, et al. Factors associated with the use of 360-degree laser retinopexy during primary vitrectomy with or without scleral buckle for rhegmatogenous retinal detachment and impact on surgical outcomes (pro study report number 4). *Retina.* 2020;40:2070–2076.
17. Tcha-Tokey K, Christmann O, Loup-Escande E, Richir S. Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *Int J Virtual Real.* 2016;16:33–48.
18. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge; 2013.
19. Kojić T, Spang R, Vergari M, et al. Effects of user factors on user experience in virtual reality: age, gender, and VR experience as influencing factors for VR exergames. *Quality and User Experience.* 2023;8:3.
20. Stanney K, Fidopiastis C, Foster L. Virtual reality is sexist: but it does not have to be. *Front Robot AI.* 2020;7:4.
21. Lorenz M, Brade J, Klimant P, et al. Age and gender effects on presence, user experience and usability in virtual environments-first insights. *PLoS One.* 2023;18:e0283565.
22. Anon. *Average global mobile and fixed broadband download & upload speed worldwide 2023.* New York, NY: Statista. Available at: <https://www.statista.com/statistics/896779/average-mobile-fixed-broadband-download-upload-speeds/>. [Accessed November 17, 2023].
23. Anon. *Sub-Saharan Africa: average download speed by country 2022.* New York, NY: Statista. Available at: <https://www.statista.com/statistics/1274951/average-download-speed-in-sub-saharan-africa-by-country/>. [Accessed November 17, 2023].
24. Anon. *HapticVR. Fundamental Surgery; 2022.* Available at: <https://fundamentalsurgery.com/platform/hapticvr/>. [Accessed November 14, 2023].
25. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull.* 1955;52:281–302.
26. Gavazzi A, Bahsoun AN, Van Haute W, et al. Face, content and construct validity of a virtual reality simulator for robotic surgery (SEP Robot). *Ann R Coll Surg Engl.* 2011;93:152–156.
27. Erfani M, El-Nasr MS, Milam D, et al. The effect of age, gender, and previous gaming experience

- on game play performance. In: *Human-Computer Interaction*. Berlin Heidelberg: Springer; 2010:293–296.
28. Shen C, Ratan R, Cai YD, Leavitt A. Do men advance faster than women? Debunking the gender performance gap in two massively multiplayer online games. *J Comput Mediat Commun*. 2016;21:312–329. Available at: [Accessed November 14, 2023].
 29. Nelson JA. Are women really more risk-averse than men? A re-analysis of the literature using expanded methods. *J Econ Surv*. 2015;29:566–585.
 30. Vergmann AS, Vestergaard AH, Grauslund J. Virtual vitreoretinal surgery: validation of a training programme. *Acta Ophthalmol*. 2017;95:60–65.
 31. Solverson DJ, Mazzoli RA, Raymond WR, et al. Virtual reality simulation in acquiring and differentiating basic ophthalmic microsurgical skills. *Simul Healthc*. 2009;4:98–103.
 32. Waisberg E, Ong J, Masalkhi M, et al. *The future of ophthalmology and vision science with the Apple Vision Pro*. Eye; 2023. Available at: <http://dx.doi.org/10.1038/s41433-023-02688-5>.