scientific reports

OPEN

Validity evidence of a new virtual reality simulator for phacoemulsification training in cataract surgery

Rémi Yaïci[1,2,3,7 ✉], Jérémie Poirot[1,2,3], Lea Dormegny[1,2,3], Nicole Neumann[1], Eugénie Bazarya[1,2,3], Lauriana Solecki[1,2,3], Arnaud Sauer[1,2,3], David Gaucher[1,2,3], Anne Lejay[1,3,4], Ann Sofia Thomsen[5,6], Nabil Chakfe[1,3,4] & Tristan Bourcier[1,2,3]

This study aims to assess validity evidence of the new phacoemulsification module of the HelpMeSee [HMS] virtual reality simulator. Conducted at the Ophthalmology Department of Strasbourg University Hospital and Gepromed Education Department, Strasbourg, France, this cross-sectional study divided 20 surgeons into two groups based on their experience over or under 300 cataract surgeries. Surgeons filled out a background survey covering their phacoemulsification experience and prior simulator use before undergoing single-session simulations on the EyeSi [EYS] and HMS simulators. Handgrip strength was measured pre- and post-simulation to evaluate grip fatigue. Afterwards, surgeons rated the perceived realism on a seven-point Likert scale. Participants were predominantly right-handed males, with expert surgeons averaging 44 years and intermediate surgeons 29 years of age. Expert surgeons had completed around 2000 phacoemulsification surgeries compared to 150 by intermediates. Primary outcome was to assess the construct validity of HMS simulator based on the difference in total and modules scores between both groups. Significant performance differences were observed between the two groups, with experts scoring higher. HMS scores were 35.8 ± 1.5 out of 46 points for experts and 27.2 ± 2.3 for intermediates ($p = 0.006$). For EYS, scores were 405.2 ± 20.3 out of 500 points for experts and 327.8 ± 25.2 for intermediates ($p = 0.028$). Experts experienced significantly less grip fatigue post-simulation on HMS compared to intermediates. This research evaluates validity evidence of HMS's phacoemulsification modules for the first time. It emphasizes the potential to broaden simulation-based training by targeting diverse populations.

Keywords Virtual reality simulation, Phacoemulsification training, Construct validity, HelpMeSee simulator, EyeSi, Handgrip fatigue

Cataract, which is defined as lens opacification, is the primary cause of reversible blindness in the world, affecting 94 million people worldwide[1]. The cataract treatment is surgical with phacoemulsification (Phaco) remaining the standard procedure for over three decades in developed countries. With increasing life expectancy contributing to a continuous rise in cataract prevalence, there is a growing demand for cataract surgeries[1]. Currently, cataract surgery ranks as the most frequently performed surgical procedure in the world, with an estimated 20 million procedures performed every year[2], which represents 85% of all ophthalmic surgeries[3]. Therefore, training enough competent surgeons represents a crucial challenge in the years to come.

In the traditional surgical training model of apprenticeship, famously summarized by Halsted[4] as "see one, do one, teach one," residents are trained under the supervision of senior surgeons. This methodology has some limitations. It results in extended training periods[5] for residents, requiring them to conduct a specific quantity of full surgeries in order to become certified. A study from 2013 estimated that residents need to complete around

[1]GEPROMED, Strasbourg, France. [2]Department of Ophthalmology, Strasbourg University Hospital, Strasbourg, France. [3]Faculty of Medicine, UNISIMES (UNIté de SIMulation Européenne en Santé), Strasbourg University Hospital, Strasbourg, France. [4]Department of Vascular Surgery and Kidney Transplantation, Strasbourg University Hospital, Strasbourg, France. [5]Department of Ophthalmology, Rigshospitalet-Glostrup, Glostrup, Denmark. [6]Copenhagen Academy for Medical Education and Simulation, Centre for HR, Capital Region of Denmark, Denmark. [7]Ophthalmology Department, Nouvel Hôpital Civil Strasbourg University Hospitals, BP426, 67091 Strasbourg, France. ✉email: remi.yaici@chru-strasbourg.fr

80 surgeries to become competent and 250 to become proficient[6,7]. Besides, the "do one" aspect of the teaching technique generates higher costs[8] and a higher incidence of post-operative complications[9]. Furthermore, this raises ethical concerns regarding the potential loss of chance for the patient. Another crucial point of the surgical training is to develop an objective method for measuring the progress of a resident's learning curve as part of the training. Scoring systems like the Objective Structured Assessment of Cataract Surgical Skill (OSACSS) was developed to assess residents' performance in cataract surgery, using a 5-point Likert scale, totalling up to 100 points[10]. Scoring systems, often seen as intricate, time-consuming, and reliant on the user, may suffer from the Halo effect, where initial success biases subsequent assessments. Thus, the need to find a supplement to this traditional model was deemed essential.

Simulation based medical education (SBME) is an essential tool that mimics real life patient scenarios to create a controlled simulated environment. These simulated experiences have long been confined to two different types of laboratory settings: wet labs, which involve animal eyes, and dry labs, which use synthetic eye models. The wet lab model provides a hands-on experience using tissues similar to those found in humans. It is specifically designed to allow for repetition and variation of surgical techniques. By training in these labs, residents can benefit from a controlled environment devoid of any patient risk. However, the model has also its limitations including limited lab access (high costs, the necessity to use a precise controlled environment with advanced equipment). Furthermore, the tissues used although similar differ from those found in humans (e.g., porcine corneas are thicker, and lenses are clear). These limitations result in most surgeries performed in wet labs being very straightforward, making it challenging to adapt procedures to more complex scenarios. Then there is the ethical issue regarding the use of animals for surgery training. The advantages and disadvantages of the dry lab model are similar to those of wet labs. Despite significant improvements in recent years, dry labs still lack the realism offered by animal tissues.

For over a decade now, the development of virtual reality ophthalmic surgical simulators constitutes a revolution. It allows for an accurate reproduction of cataract surgery[11]. With the help of virtual reality, we can now address the underlying issues of the traditional training methods. VR allows for an objective assessment of a trainee's performance. Construct validity, based on the Messick's unified validity framework, the predominant approach in medical education literature, ensures that there is a correlation between real life cataract surgery experience and trainee scores[12]. Similar to wet and dry labs, virtual reality simulator enhances real life surgical skills, boosts the learning curves[13,14], decreases the rate of post-operative complications[15–17], and reduces total operating time[18]. Moreover, using VR boosts residents' confidence[18]. Another important aspect of the learning curve lies in the operative ergonomics, which increase with experience and its corollary, post-operative muscle fatigue, which decreases correspondingly[19]. For cataract surgery, a portable device, whose validity has been demonstrated, facilitates an evaluation of the muscle fatigue of the gripping muscles: the Handgrip dynamometer (Takei, TK200, Takei Scientific Instruments Japan)[20,21].

EyeSi VR Magic simulator (Haag-Streit, Mannheim, Germany) [EYS] has been launched in 2004. It is the most widespread simulator (15 out of 27 French university hospitals currently owned one in 2023 and the total number of EYS is approximately 1000 worldwide). Validity of the simulator has been widely studied[22–26]. A review article from 2020 [27] found that 9 out of 11 studies conducted on the subject demonstrated a correlation between real-life surgical experience and performance in VR modules. In 2017, Thomsen et al. successfully confirmed construct validity in a pre-established program[28], building upon an earlier 2015 study[29] where they determined a threshold assessment score that indicates when a surgical skill is acquired. **HelpMeSee** simulator [HMS] is conceived by a non-profit organisation based in the United States. Their global presence differs from that of EYS. Until 2023, they focused on establishing their presence in Low-to middle-income countries (LMICs), initially concentrating on manual small-incision cataract surgery (MSICS) as part of the Vision 2020 initiative[30]. Construct validity[31] and efficacy, based on the Kirkpatrick model, for the procedure of MSICS have been already demonstrated[32]. In 2023, HMS introduced the phacoemulsification and suturing modules, possible operative complications scenarios and more challenging surgical cases. To our knowledge, there are currently no studies on construct validity for these new phacoemulsification modules.

The primary objective of this study is to assess the construct validity of HMS simulator. Intrinsic validity will be evaluated to distinguish between expert-level surgeons (> 300 surgeries) and intermediate-level surgeons (< 300 surgeries). Extrinsic validity will focus on correlating the overall HMS score with the score previously validated for EYS. The secondary objectives include conducting a comparative analysis of scores and identifying influencing factors between the expert and intermediate groups. Additionally, the study will assess the post-simulation session muscular fatigue on dynamometers.

Materials and methods
Research design and setting
This is a cross-sectional, non-randomised cohort study conducted within the ophthalmology department at the University Hospital of Strasbourg and Gepromed Education Department, Strasbourg, France. The members of the ethics committee of the Faculty of Medicine at Strasbourg have given their approval on the trial (CE-2024-6). The research and all used methods related to human has been complied with all the relevant national regulations, institutional policies and in accordance the tenets of the Helsinki Declaration. The study was carried out between January and February 2024, with statistical analysis completed in March 2024.

Subjects
The inclusion criteria included qualified active ophthalmic surgeons with prior experience performing phacoemulsification. Written and informed consent was obtained from all participants before the beginning of the study. Exclusion criteria included surgeons who have ceased practice and those whose surgical experience

could not be determined. Participants who did not consent or withdrew their consent were excluded from the study.

Two groups of participants based on their level of experience were included: a group who performed more than 300 cataract surgeries ($N = 10$), and a group who performed less than 300 surgeries ($N = 10$). The threshold of 300 cataract surgeries was selected based on the findings of Ho et al.[6] and the estimation by Lansigh et al.[7] who suggested that competence is mostly achieved after approximately 250 surgeries. This approach ensures that one group is considered proficient in this surgery, while another group is still in the process of gaining proficiency. Surgeons were recruited from the ophthalmology department at university hospital in Strasbourg and from private practice in Strasbourg area.

Study design

Prior to the first session, surgeons were asked to fill out a personal survey consisting of general data (gender, age, handedness, TNO score in arcseconds), their past experience with phaco surgeries (total number of surgeries performed, surgeries within the last 12 months, time since last surgery in days), their previous exposure to surgical simulators (phaco and other surgeries), potential factors influencing their practice on the simulator (such as playing musical instruments, and spending more than 3 h per week on video games), and factors affecting muscle fatigue (such as sports). Two virtual reality simulators were used. The **EyeSi** VR Magic simulator [EYS] (Fig. 1a) version 3.10.6 and the **HelpMeSee** simulator [HMS] (Fig. 1b) version 8.228.2. During the simulation sessions, the first simulator that was used for training and the time of day (divided into morning, noon, afternoon, evening) were recorded. Standardised oral explanations were given to each participant at the beginning of the session, and no warm-up period on the simulator was allowed.

All planned steps on each simulator were executed within a single session, without the option to repeat. Objective measurements using the Handgrip Dynamometer (Takei, TK200, Takei Scientific Instruments Japan) were conducted prior to the initial simulation session to assess average grip strength and after each simulation session to evaluate grip fatigue. After the sessions, a seven-point Likert scale was used to determine if the surgeons found the session to be overall realistic of a real surgery (from very realistic to not at all realistic). A five-point Likert scale was also used (ranging from strongly agree to strongly disagree) to assess the following aspects of realism: hand/microscope coordination, multitasking, instruments movement, depth of field in binocular vision, interaction between instruments and the instruments interaction with simulated anatomical structures. A last question assessed the user-friendliness of the computer interface.

Primary outcome measures

The primary outcome measures of this study to assess the intrinsic validity of HMS is the comparison of total and modules scores of HMS between both groups, using the number of surgeries as a proxy measure for surgical experience and therefore surgical proficiency. The primary outcome measures of this study to assess the extrinsic validity of HMS is the correlation of the HMS total score with the score previously validated for EYS.

The two virtual reality simulators use different scoring systems based on different metrics and on different scales. When both simulators make it possible to practice each step of the surgery individually, with sequences subdivided to train each surgical gesture separately, HMS provides users with tactile feedback, allowing surgeons to practice incisions and experience the sensations of instrument entry and exit, as well as hydrosuturing. On the other hand, EYS allows users to customise a programme according to their preferences. To ensure maximum comparability, we decided to only configure the steps that were reproducible on both simulators. Both simulators share several common metrics: they use a cumulative negative point system for errors deemed significant, which could result in a loss of all possible attainable points and the potential for a negative final score. These metrics are adapted to the simulators' respective scales.

The modules for the EYS were selected based on their different difficulty levels and conducted in the university hospital in Strasbourg. Five specific modules were used, each contributing to a total score of 500. These modules and their respective difficulty levels were as follows: Capsulorhexis (Level 3, configured) score/100; Hydro Maneuvers (Level 3) score/100; Phaco Divide and Conquer (Level 6) score/100; Irrigation and Aspiration (Bimanual, Level 4) score/100; and IOL Insertion (Level 3) score/100. The session with HMS was performed at Gepromed in Strasbourg, using the following modules: Capsulorhexis score/6; Hydrodissection score/5; Nucleus rotation score/3; Grooving score/8; Cracking score/5; Quarter Removal score/7; Irrigation and aspiration for cortex removal (bimanual) score/6; IOL insertion score/3; Replacement IOL score/3; with the total score out of 46. All the metrics and their respective scores from the two simulators are summarized in the Fig. 2.

We used the Handgrip dynamometer to measure the average grip strength of the participants. Before the simulation, participants underwent a protocol that included three successive three-second contractions of maximum voluntary effort accompanied by verbal encouragement, and 30-second rest periods between each contraction. Right after the surgical simulation, we evaluated grip fatigue through ten three-second contractions of maximum voluntary effort, with a two-second rest interval between each contraction. To assess muscle fatigue, we calculated the percentage difference between the first and last contractions.

Statistical analysis

In line with the general data protection regulation (GDPR), these data were pseudonymously processed. Power and sample size were determined by using G*Power (Heinrich-Heine Universität, Version 3.1.9.7) to detect a statistically significant difference between the total scores of each group on the same simulator. A power of 80% was estimated. We used descriptive statistics for the analysis of numerical and nominal data and inferential statistics to evaluate whether there was a significant difference between the expert and intermediate groups. We assessed this difference using both global scores and scores for each individual step of the procedure. Next, we explored correlations between global scores and various participant characteristics, influencing factors, and

Fig. 1. Virtual reality simulators. (**A**) EyeSi© (EYS) is a computer system with a microscope-like display for 3D visuals. It features a table setup with a mannequin head, a simulated eye, two control pedals for microscope and phacoemulsification operations, and a tactile interface with three ophthalmic probes. (**B**) HelpMeSee© (HMS) is a virtual reality simulator with tactile feedback, resembling a reclining patient connected to a microscope-shaped screen for 3D visuals. The setup includes a table, a mannequin head with tactile feedback, a single pedal for phacoemulsification control, and a tactile interface with two ophthalmic probes that mimic the surgical instruments.

dynamometer scores. This helped us identify potential factors that may have influenced surgical performance in the simulated environment. For continuous and normally distributed variables, we used the Pearson correlation coefficient to assess linear relationships. Before exploring these correlations, we checked for the assumption of normality using the Shapiro-Wilk test and Q-Q plot. We used the Spearman correlation coefficient for variables that were either ordinal or not-normally distributed. To evaluate relationships between variables, we used the following statistical tests: the chi-square test to evaluate the relationships between categorical variables with nominal scales and the Phi correlation coefficient for naturally dichotomous variables. We used Student's t-test for continuous variables that were normally distributed, and the Wilcoxon Mann-Whitney test for non-normally distributed variables. We conducted multiple linear regressions to evaluate the relationships between predictive factors and total scores. This allowed us to determine which factors were significant predictors of performance in the simulated surgical setting. In our analysis, a p-value lower than 0.05 was considered statistically significant. All statistical analyses were obtained using Prism Software version 9.0.0 from GraphPad, located in La Jolla, California, USA.
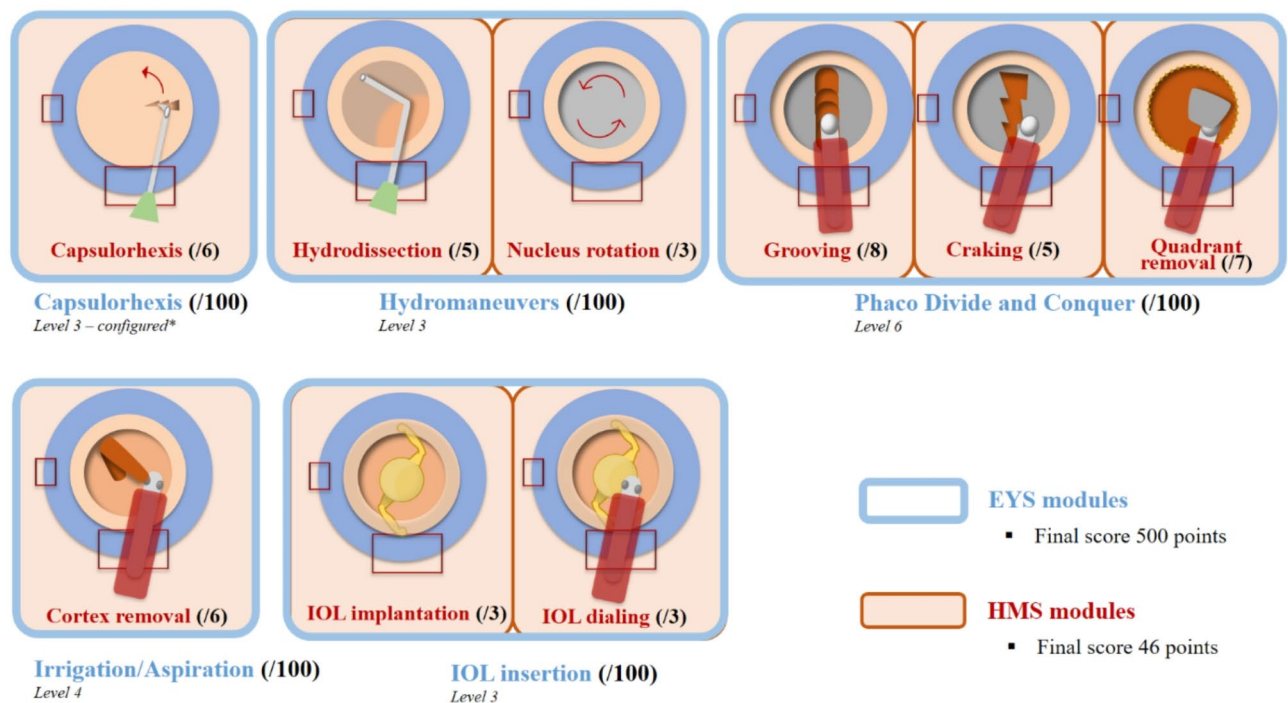
Fig. 2. Common phacoemulsification modules in both simulators with associated scores. The two virtual reality simulators differ in the modules available. The HelpMeSee (HMS) offers the possibility of practising each stage of phacoemulsification, divided into sub-sequences to practise each gesture separately. The EYS can be used to configure a personalised programme for surgeons wishing to evaluate a complete surgical procedure (excluding incisions and hydrosuture). For this study, we used all the steps common to both simulators, i.e. the complete surgery programme (excluding incisions and hydrosuture) on the HMS and a personalised programme on the EYS. The total scores generated were used as the primary endpoint for statistical evaluation. EYS: EyeSi® simulator (Haag-Streit, Mannheim, Germany). HMS: HelpMeSee® simulator (HelpMeSee foundation, New York, United States). *configured: no guiding elements, no weak zonula fibers, initially stained, capsular tension medium.*

Results
Participant characteristics
Between January and February 2024, in total, ten surgeons were included the expert-level group and ten surgeons in the intermediate-level group. The participants were mostly right-handed and male. The average age of participants was 44 in the expert group and 29 in the intermediate group. Although the TNO score had an identical median, there was greater variability within the expert group. Regarding surgical experience, the expert group had a median of 2000 phaco procedures, with 250 performed within the last year, while the intermediate group had a median of 150 procedures, with 75 performed within the last year. Table 1 presents the demographic and characteristic details of the participants involved in the study. A positive correlation was found between the total number of phaco surgeries carried out by surgeons in their career and the number of phaco surgeries done within the past year (Pearson, rho $= 0.484$, $p = 0.031$).

The two groups have benefited for the most part from training sessions on phaco virtual reality simulator, with much fewer sessions centred on simulator training for other types of surgeries. The most frequently used phaco virtual reality simulator was EYS ($n = 14$, 70%), followed by HMS ($n = 3$, 15%). Four participants had previous surgical training using the DaVinci robotic simulator (Intuitive Surgical, Sunnyvale, CA, USA).

Few potential influencing factors were reported, except for sports in the intermediate experience group (90%). Overall, participants engaged in a vast array of sports including golf, running, kayaking, tennis, cycling, crossfit, badminton, gymnastics, and climbing. The most played musical instrument was the piano, played by six surgeons. Four participants reported playing video games for 3 or more hours per week.

Total scores
The average global score on the HMS for the expert group was 35.8 ± 1.5, while that of intermediate group was 27.2 ± 2.3. The difference was statistically significant ($p = 0.006$). The average global score on the EYS for the expert group was 405.2 ± 20.3, while that of intermediate group was 327.8 ± 25.2. The difference was also statistically significant ($p = 0.028$). The total scores from both groups on each simulator are represented in Fig. 3.

Additionally, there was a strong correlation between the global scores of the HMS and the EYS (Pearson, rho $= 0.667$, $p < 0.001$). A linear regression model for the HMS global score was significant ($N = 20$, F $(4,15) = 6.11$, $p = 0.004$, $R^2 = 0.620$, adjusted $R^2 = 0.518$). It includes the surgical experience (total number of phaco surgeries

		SENIOR	INTERMEDIATE
Participants	N	10	10
Gender	Men N (%)	8 (80)	7 (70)
Age	Mean ± SD	44.3 ± 11.5	29.1 ± 2.1
Dominant hand	% right-handed	80	70
TNO score (seconds of arc)	Median (min, max)	30 (15, 240)	30 (15, 60)
Previous experience			
Average number of Phaco	Median (Q_{25}–Q_{75})	2000 (1400–6000)	150 (60–185)
Number of Phaco in the last 12 months	Median (Q_{25}–Q_{75})	250 (150–350)	75 (10–80)
Number of days since last PE	Median (Q_{25}–Q_{75})	2 (1–3)	10 (3–15)
Targeted training on VRS de Phaco	N (%)	7 (70)	8 (80)
Targeted training on VRS from another surgery	N (%)	3 (30)	2 (20)
Influencing factors			
Music instrument	N (%)	2 (20)	4 (40)
Video games	N (%)	1 (10)	3 (30)
Sport	N (%)	4 (40)	9 (90)

Table 1. Participant demographics and characteristics. The table includes two groups, categorized by experience level (senior and intermediate), each of equal size. The typical participant in our study is a right-handed male with good binocular vision (median TNO score of 30 arcseconds) who has had one or more virtual reality sessions prior to the study, mainly on the EyeSi simulator. The participants in the senior group were on average fifteen years older than those in the intermediate group. They were more experienced, having performed a median of thirteen times more cataracts than the youngest candidates over the course of their professional careers. They had a more regular activity over the last year, resulting in around three times as many operations and a shorter interval between operations. Few potential influencing factors were found, except for sport, which was massively practised in the intermediate experience group. Phaco: *phacoemulsification*, VRS: *virtual reality simulator*. Senior: experience of more than 300 PE procedures, intermediate: experience of less than 300 PE procedures. None of the participants indicated that they were ambidextrous.
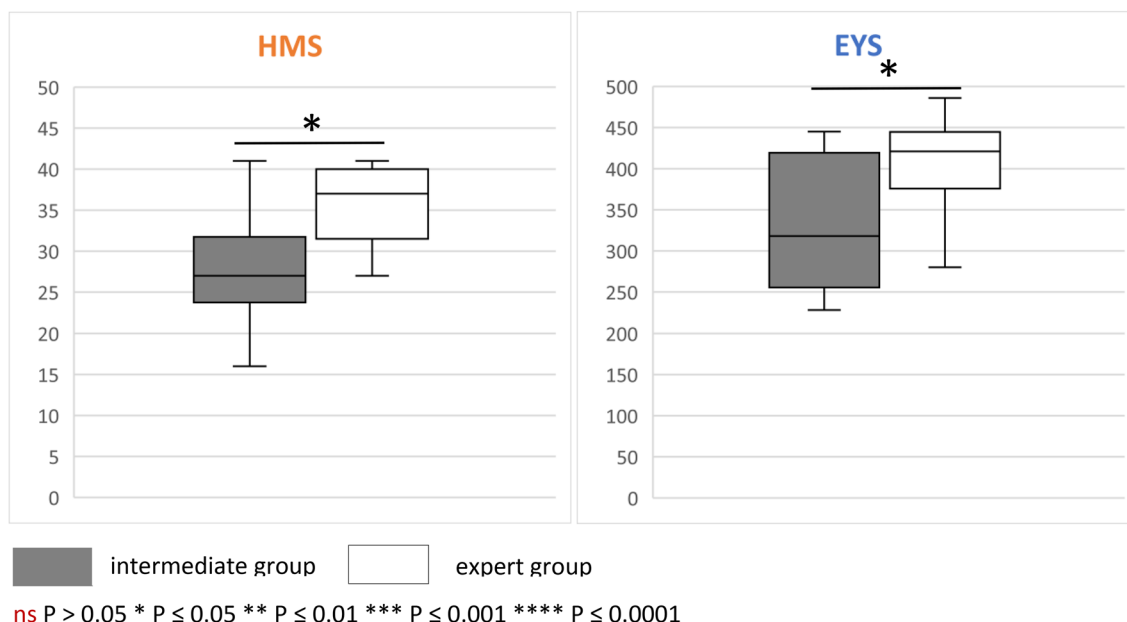

ns P > 0.05 * P ≤ 0.05 ** P ≤ 0.01 *** P ≤ 0.001 **** P ≤ 0.0001

Fig. 3. Total scores of the participants. As the metrics for the two simulators have different scales, that is, maximum score 46 for HelpMeSee© (HMS) and 500 for EyeSi© (EYS), results are represented on two different graphs. On both simulators, the expert group scored significantly higher than the intermediate group, demonstrating the validity (*construct validity*) of the latter.

carried out by surgeons in their career) and the global score on the EYS. The global score on the EYS emerges as the only significant predictor in this model (ß = 0.06, SE = 0.0019, $p = 0.006$).

Module scores

An analysis of each HMS module scores showed a significant difference between the expert group and the intermediate group for the module *capsulorhexis* ($p = 0.029$) and *quadrant removal* ($p = 0.019$). This analysis of each EYS module scores showed a significant difference between the expert group and the intermediate group for the module *divide and conquer* ($p = 0.023$). The different module scores from both groups on each simulator are represented in the Fig. 4.

Additionally, there was a correlation between the module scores *capsulorhexis* and *grooving* (Pearson, rho = 0.500, $p < 0.025$). A linear regression model for the module score *capsulorhexis* was significant ($N = 20$, $F_{(4, 15)} = 6.243$, $p = 0.004$, $R^2 = 0.525$, adjusted $R^2 = 0.625$). The module score *grooving* emerges as the only significant predictor in this model (ß = 0.874, SE = 0.355, $p = 0.026$).
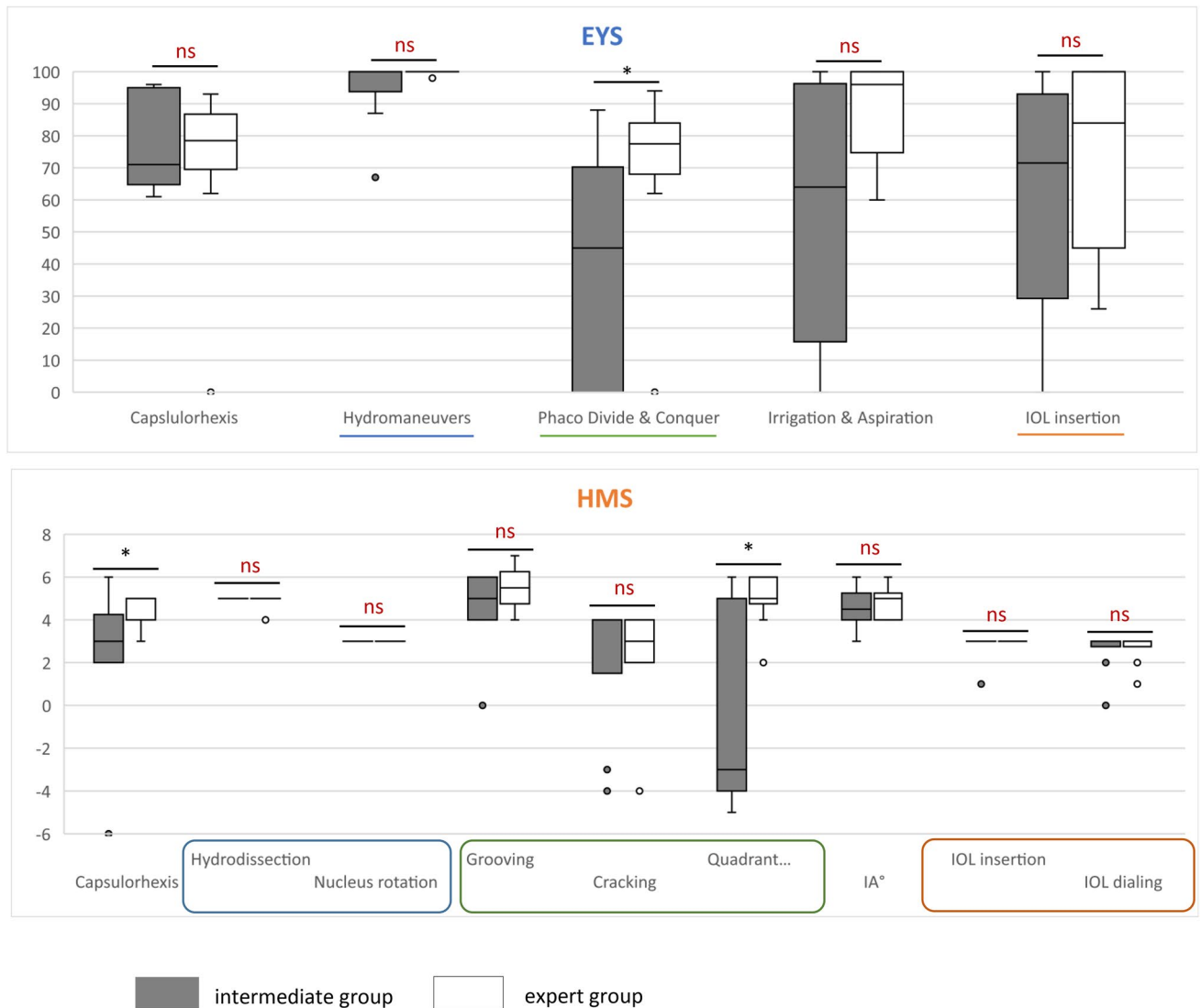

Fig. 4. Scores of the participants by module. A comparison was made between the expert and intermediate groups on each of the modules to assess which modules were the most discriminating between the two groups. As the metrics of the two simulators have different scales, they are represented on two different graphs. (**A**) **Eyesi** simulator (EYS). The expert group showed a better average for each module. However, only the *Phaco Divide & Conquer* module showed a significant difference between the scores of the two groups. (**B**) **HelpMeSee** simulator (HMS). Unlike the EYS, some modules: *hydrodissection*, *nucleus rotation* and *IOL insertion* showed no difference between the two groups. For other modules: *grooving*, *cracking*, *cortex removal* and *IOL dialing*, the average of the expert group was higher than that of the intermediate group, although the difference was not statistically significant. Finally, two modules proved to be significantly discriminating: *capsulorhexis* and *quadrant removal*. ns P > 0.05 * P ≤ 0.05 ** P ≤ 0.01 *** P ≤ 0.001 **** P ≤ 0.0001. IA: *irrigation & aspiration for cortex removal*

Satisfaction survey results

Both the intermediate and expert-level groups shared comparable opinions regarding both simulators across all surveyed elements, with one notable exception: the hand-microscope coordination, where the senior group showed more reservations compared to the intermediate group. In terms of positive feedback, both simulators received favourable responses primarily for movement realism and user-friendly computer interface. Alternatively, the interaction between instruments and anatomical structures received the least positive feedback. The expert group's responses regarding multitasking were inconsistent, with three participants fully endorsing its realism while one participant disagreed entirely. The responses to the surveys from both groups for each simulator are graphically summarized in the Fig. 5. The comments left by participants provided a valuable insight into interpreting the satisfaction survey responses, revealing shortcomings in realism, particularly in capsulorhexis.

Muscle fatigue

The average resting grip strength was measured for experts at 35.12 ± 2.7 kg and intermediates at 35.95 ± 2.8 kg with no significant difference between the two groups ($p = 0.832$). Throughout the study, the average fatigue level showed no correlation with the time of day ($p = 0.573$).

After using the HMS simulator, mean grip fatigue was 5.41 ± 1.97 kg for expert surgeons and 16.95 ± 3.05 kg for intermediate level surgeons. The difference was statistically significant ($p = 0.006$). The average fatigue levels after using the EYS were 8.9 for experts and 12.85 for intermediate level surgeons with no significant difference ($p = 0.579$). The strength and fatigue scores from both groups for each simulator are summarised in the Table 2.

Discussion

This study assessed for the first time the extrinsic and intrinsic construct validity of the Phaco modules of HMS. The construct validity of a virtual reality simulator can be defined by the extent to which the simulator is able to differentiate between those of different experience levels[33]. Our two groups are representative of two career phases of an ophthalmologic surgeon. The median age of the intermediate group was 29 years, aligning with the average age of ophthalmologists at the end of their residency. This group consisted primarily of residents and clinical associate professors similar in age and surgical experience. On the other hand, the expert group was very heterogeneous with a wide age range, spanning 33 to 68 years, a wide range in terms of Phaco experience, from a minimum of 300 to a maximum of 17,500 procedures, mostly depending of practice location. Despite differences in experience and age, participants had the same male-to-female ratio and an equal distribution of right and left-handed individuals. Additionally, they had identical TNO medians. Most surgeons in our cohort had prior experience with the EYS virtual reality simulator (70%), which is the most widely used simulator in France. There were very few influencing factors among participants, with the main one being sports practice. However, due to the diverse range of sports practiced by participants, the low number of participants for each individual sport resulted in statistically insignificant analysis outcomes. Few sports could have had a significant impact on simulator performance[34].

Both simulators found a significant difference between the scores of the expert and intermediate groups. The HMS score showed correlations with the EYS score in both univariate and multivariate analyses. This is, to our knowledge the first study that confirms the validity of phaco modules of the HMS simulator, while relying on the demonstrated validity of EYS. Training on a virtual reality simulator improves several aspects of surgical training, reducing complications and total operating time. It also has the benefit of improving user confidence and reducing stress levels[28]. Simulated training is beneficial to patients, learners, and instructors alike. This result presents a complementary opportunity to improve the use of SBME. Both simulators are different and complement each other: HMS offers tactile feedback and two probes but lacks a simulated eye, while EYS features three probes and a simulated eye. The simulated training for surgical steps like incision and hydrosuture is more effective because of the tactile feedback provided by HMS. While HMS is non-profit-oriented and primarily targeted towards LMICs, EYS is designed for commercial distribution. Different populations due to their varying distributions will use the two simulators, yet it is known that they will provide equivalent benefits. It is important to note that few centres are equipped with both at the same time, thus making their comparison useful.

Both simulators use distinct scoring systems but on both, the modules that significantly distinguished performances between the two groups were the nucleus fragmentation. On the HMS, the module *capsulorhexis* was also discriminating. This observation is particularly interesting, as these steps are recognised as the most challenging to master[35,36]. Additionally, on the HMS, the module *capsulorhexis* correlates with the module *grooving* in both univariate and multivariate analyses. These validity findings prompt us to contemplate giving greater importance to these steps or even setting a minimum score, contributing to discussions surrounding the creation of a "surgical licence". Indeed, we can rely on these validity results to consider a greater weighting of these stages, or even a minimum score.

The satisfaction survey indicated that the EYS received more favourable feedback than the HMS. Crucial elements were considered to be lacking in the HMS simulation, for instance the absence of a microscope and its corresponding pedal, the inability to adjust settings on the phaco machine, and the limited manoeuvrability of the instrument with no lateral adjustability. This lack of flexibility forces the surgeon to conform to the expectations of the virtual reality simulator rather than focus on the surgery. It is important to note that a new version was released at the time this manuscript was being written, introducing greater flexibility.

The results should be interpreted according to the surgeons' comments, the study timeline (Q1 2024), and focused only on the evaluated modules. It is important to consider the operational history of each simulator (EYS being 20 years old), and the ongoing development of the phacoemulsification module in HMS. The participants used indeed the first version (released in august 2023) of the phacoemulsification module. Throughout this study, HMS has demonstrated construct validity, proving it a highly promising training module. Small incision cataract
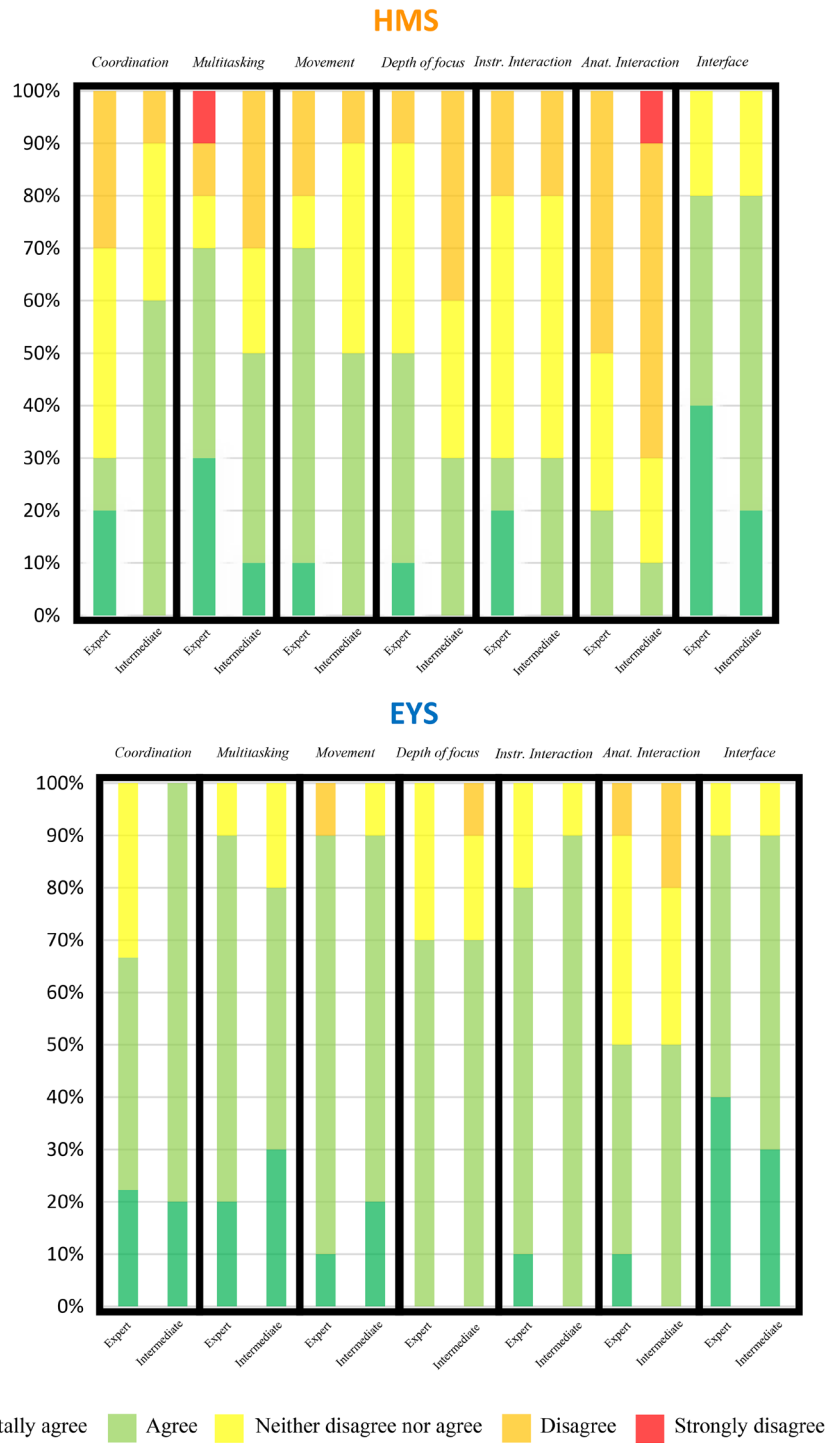
Fig. 5. Results of the satisfaction questionnaires for HelpMeSee® (HMS) and Eyesi® (EYS) simulators. All the participants were asked to complete a satisfaction questionnaire on the realism of the simulators after each evaluation session. The two groups gave a broadly similar assessment of the two simulators for all the questions asked, except for hand-microscope co-ordination, where the senior group had more reservations (> 30% difference in satisfaction). The two simulators also elicited the most mixed opinions on the interaction of the instruments with the simulated anatomical elements. On all the questions asked, the opinions given on the EYS simulator were more favourable at the time the study was conducted (Q1 2024). Coordination: hand-microscope coordination for real surgery. Multitasking: multitasking (microscope, hand, pedals, phacoemulsification machine settings). Movement: Movement of the surgical instruments. Depth of focus: Depth of field in binocular vision. Instr. Interaction: Interaction between surgical instruments. Anat. Interaction: Interaction of surgical instruments with simulated anatomical elements. Interface: User-friendly computer interface.

		SENIOR	INTERMEDIATE	P
Average resting grip (in kg)	*Mean ± SD*	35.1 ± 2.7	36.0 ± 2.8	0.832
Mean grip fatigue (in kg)				
HMS	*Mean ± SD*	5.4 ± 2.0	17.0 ± 3.1	0.006
EYS	*Mean ± SD*	8.9 ± 9.5	12.9 ± 10.4	0.579

Table 2. Muscle fatigue. The table summarised the average resting strengths and fatigue after stimulation of the two groups (senior and intermediate). Throughout the study, the average fatigue level after using the HMS simulator was significant higher for the intermediate group although no significant difference between the two groups was shown at the baseline. This significant difference was not observed after using the EYS simulator. HMS: *HelpMeSee simulator*, EYS: *Eyesi simulator*. Senior: experience of more than 300 procedures, intermediate: experience of less than 300 procedures.

surgery is currently the basis for designing the HMS phaco module and destined for developing countries. While HMS offers more possibilities with the tactile feedback, such as incision and hydrosuture, these aspects were not evaluated to maintain comparability with EYS. A technical hurdle remains the lack of tactile feedback on EYS. For reliable results, the evaluation of the HMS phacoemulsification module should include a training session, an eBook and assistance from qualified instructors.

Lower levels of fatigue were observed within the expert group training on HMS. There is a strong correlation between muscular fatigue and the participants' overall surgical experience. Research suggests that muscle fatigue tends to increase with operating time[37], while operating time tends to decrease with surgical experience[38]. Consequently, this result provides an additional element regarding the validity assessment of the simulator. One point that can be highlighted is to question whether the tactile feedback present in the HMS and absent in the EYS contributed to demonstrating better ergonomics in the senior group.

Several study limitations were identified. Due to participants' limited availability and logistical limitations (availability of simulation facilities), the randomization of order of appearance could not be maintained. This represents a significant bias (warm-up effect on performance and satisfaction score psychology). HMS offers a diverse array of surgical steps such as incision and hydrosuture, which were not evaluated to maintain comparability with EYS. It is necessary to perform future studies on the full scope of HMS modules.

Conclusion

The confirmed validity of the phacoemulsification modules in this study opens new possibilities, such as the creation of a surgical licence and the expansion of simulation based medical education using **HelpMeSee**. While **HelpMeSee** has shown promise in various applications and has proved to be complementary to the existing offer, it is crucial to consider its current ways of improvement and conduct further developments. These developments will need to be accompanied by validity and effectiveness studies. More generally, given their different modalities and accessibility, it is essential to support the development of virtual reality simulators to ensure they meet educational needs effectively.

Data availability

The datasets used and analyzed during the study reported herein are available from the corresponding author on reasonable request.

References

1. World Health Organization. *World Report on Vision* (World Health Organization, 2019).
2. Gower, E. W. et al. Perioperative antibiotics for prevention of acute endophthalmitis after cataract surgery. *Cochrane Database Syst. Rev.* **2017** (2017).
3. Muttuvelu, D. V. & Andersen, C. U. Cataract surgery education in member countries of the European Board of Ophthalmology. *Can. J. Ophthalmol.* **51**, 207–211 (2016).
4. Osborne, M. P. William Stewart Halsted: his life and contributions to surgery. *Lancet Oncol.* **8**, 256–265 (2007).
5. La Cour, M., Thomsen, A. S. S., 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.* **257**, 877–881 (2019).
6. Ho, J. & Claoué, C. Cataract skills: how do we judge competency? *J. R. Soc. Med.* **106**, 2–4 (2013).
7. Lansingh, V. C., Daher, P. G., Flores, T. D., Star, E. M. L. & Martinez, J. M. How many cataract surgeries does it take to be a good surgeon? *Rev. Mex Oftalmol.* **97**, 33–34 (2023).
8. Hosler, M. R. et al. Impact of resident participation in cataract surgery on operative time and cost. *Ophthalmology*. **119**, 95–98 (2012).
9. Martin, K. R. G. & Burton, R. L. The phacoemulsification learning curve: per-operative complications in the first 3000 cases of an experienced surgeon. *Eye*. **14**, 190–195 (2000).
10. Martin, J. A. et al. Objective structured assessment of technical skill (OSATS) for surgical residents: objective structured assessment of technical skill. *Br. J. Surg.* **84**, 273–278 (1997).
11. Sikder, S., Tuwairqi, K., Al-Kahtani, E., Myers, W. G. & Banerjee, P. Surgical simulators in cataract surgery training. *Br. J. Ophthalmol.* **98**, 154–158 (2014).
12. Messick, S. Validity of test interpretation and use. Princeton, NJ, USA. *Educ. Test. Serv.* (1990).

13. Pokroy, R. et al. Impact of simulator training on resident cataract surgery. *Graefes Arch. Clin. Exp. Ophthalmol.* **251**, 777–781 (2013).
14. McCannel, C. A., Reed, D. C. & Goldman, D. R. ophthalmic surgery simulator training improves resident performance of capsulorhexis in the operating room. *Ophthalmology.* **120**, 2456–2461 (2013).
15. Dean, W. H. et al. Intense simulation-based surgical education for manual small-incision cataract surgery: the ophthalmic learning and improvement initiative in cataract surgery randomized clinical trial in Kenya, Tanzania, Uganda, and Zimbabwe. *JAMA Ophthalmol.* **139**, 9 (2021).
16. Ferris, J. D. 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.* **104**, 324–329 (2020).
17. Rogers, G. M. et al. Impact of a structured surgical curriculum on ophthalmic resident cataract surgery complication rates. *J. Cataract Refract. Surg.* **35**, 1956–1960 (2009).
18. Thomsen, A. S. S. et al. Operating room performance improves after proficiency-based virtual reality cataract surgery training. *Ophthalmology.* **124**, 524–531 (2017).
19. Morton, J. & Stewart, G. D. The burden of performing minimal access surgery: ergonomics survey results from 462 surgeons across Germany, the UK and the USA. *J. Robot Surg.* **16**, 1347–1354 (2022).
20. Mallard, J. et al. Early skeletal muscle deconditioning and reduced exercise capacity during (neo)adjuvant chemotherapy in patients with breast cancer. *Cancer.* **129**, 215–225 (2023).
21. Cadenas-Sanchez, C. et al. Reliability and validity of different models of TKK hand dynamometers. *Am. J. Occup. Ther.* **70**, 7004300010p1–7004300010p9 (2016).
22. Spiteri, A. V. et al. Development of a virtual reality training curriculum for phacoemulsification surgery. *Eye.* **28**, 78–84 (2014).
23. Bergqvist, J., Person, A., Vestergaard, A. & Grauslund, J. Establishment of a validated training programme on the Eyesi cataract simulator. A prospective randomized study. *Acta Ophthalmol. (Copenh).* **92**, 629–634 (2014).
24. Saleh, G. M. et al. The development of a virtual reality training programme for ophthalmology: repeatability and reproducibility (part of the International Forum for Ophthalmic Simulation studies). *Eye.* **27**, 1269–1274 (2013).
25. Privett, B., Greenlee, E., Rogers, G. & Oetting, T. A. Construct validity of a surgical simulator as a valid model for capsulorhexis training. *J. Cataract Refract. Surg.* **36**, 1835–1838 (2010).
26. Mahr, M. A. & Hodge, D. O. Construct validity of anterior segment anti-tremor and forceps surgical simulator training modules attending versus resident surgeon performance. *J. Cataract Refract. Surg.* **34**, 980–985 (2008).
27. Nayer, Z. H., Murdock, B., Dharia, I. P. & Belyea, D. A. Predictive and construct validity of virtual reality cataract surgery simulators. *J. Cataract Refract. Surg.* **46**, 907–912 (2020).
28. Thomsen, A. S. S. et al. High correlation between performance on a virtual-reality simulator and real-life cataract surgery. *Acta Ophthalmol. (Copenh).* **95**, 307–311 (2017).
29. Thomsen, A. S. S., Kiilgaard, J. F., Kjærbo, H., la Cour, M. & Konge, L. Simulation-based certification for cataract surgery. *Acta Ophthalmol. (Copenh).* **93**, 416–421 (2015).
30. Vision 2020: the cataract challenge. *Community Eye Health* **13**, 17–19 (2000).
31. Hutter, D. E. et al. A validated test has been developed for assessment of manual small incision cataract surgery skills using virtual reality simulation. *Sci. Rep.* **13**, 10655 (2023).
32. Nair, A. G. et al. Effectiveness of simulation-based training for manual small incision cataract surgery among novice surgeons: a randomized controlled trial. *Sci. Rep.* **11**, 10945 (2021).
33. Sjøberg, D. I. K. & Bergersen, G. R. Construct validity in Software Engineering. *IEEE Trans. Softw. Eng.* **49**, 1374–1396 (2023).
34. Gibney, E. J. Performance skills for surgeons: lessons from sport. *Am. J. Surg.* **204**, 543–544 (2012).
35. Dooley, I. J. & O'Brien, P. D. Subjective difficulty of each stage of phacoemulsification cataract surgery performed by basic surgical trainees. *J. Cataract Refract. Surg.* **32**, 604–608 (2006).
36. Al-Jindan, M. et al. Assessment of learning curve in phacoemulsification surgery among the Eastern Province Ophthalmology Program residents. *Clin. Ophthalmol. Auckl. NZ.* **14**, 113–118 (2020).
37. Slack, P., Coulson, C., Ma, X., Webster, K. & Proops, D. The effect of operating time on surgeons' muscular fatigue. *Ann. R Coll. Surg. Engl.* **90**, 651–657 (2008).
38. Wiggins, M. N. & Warner, D. B. Resident physician operative times during cataract surgery. *Ophthalmic Surg. Lasers Imaging Retina.* **41**, 518–522 (2010).

Acknowledgements

Author contributions

R.Y. designed the study, contributed to the acquisition, analysis and interpretation of data and drafted the manuscript. J.P contributed to the acquisition of data and revised the manuscript. L.D. participated in the design of the study, in the conception of the manuscript and revised it. N.N. contributed to the analysis and interpretation of data and revised the manuscript. F.L contributed to the analysis and interpretation of data. E.B contributed in the conception of the manuscript. L.S. participated in the acquisition of data, participated in the conception of the manuscript and revised it. A.L revised the manuscript. A.S participated in the acquisition of data, revised the manuscript. D.G. participated in the acquisition of data, revised the manuscript. A.S.T contributed to the analysis and interpretation of data and revised the manuscript. N.C. revised the manuscript. T.B. designed the study, participated in the acquisition and interpretation of data and revised the manuscript.

Funding

Declarations

Ethics approval and consent to participate

Ethics approval was obtained by the Ethics Committee of Strasbourg University Hospital. Participants were

enrolled after providing written informed consent.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to R.Y.

Reprints and permissions information is available at www.nature.com/reprints.

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