Ophthalmol Ther (2025) 14:1999–2020
https://doi.org/10.1007/s40123-025-01175-2

ORIGINAL RESEARCH

Validation of the Eyesi Slit Lamp Simulator for a Simulated Training Curriculum for Residents in Ophthalmology: A National Delphi study

Kasper Ruotsalainen Hansen · Simon Joel Lowater · Kristina Sanne Lyngsø · Anne Helene Køllund Nissen · Anna Stage Vergmann

Received: March 25, 2025 / Accepted: May 21, 2025 / Published online: July 4, 2025

ABSTRACT

***Introduction*:** Simulation-based training is increasingly used in medical education to improve clinical skills. The Eyesi slit lamp simulator offers a platform for training in slit lamp examination and indirect ophthalmoscopy. However, training needs to be evidence-based to ensure reliable evaluation. Therefore, this study aims to identify relevant training modules on the Eyesi slit lamp simulator for ophthalmology residents through a national Delphi study to develop a training curriculum.

***Methods*:** A two-round Delphi study was conducted with chief physicians, consultants, and professors involved in ophthalmology residency training across Denmark. In round one, experts evaluated the relevance of training modules on the Eyesi slit lamp simulator using a binary scale, and modules achieving ≥ 75% in relevance were included in round two. In round two, these modules were rated on a linear scale (1–5) for detailed assessment. Consensus was defined by a standard deviation (SD) < 1.0 and interquartile ranges (IQR) were calculated to assess variability and identify outliers.

***Results*:** A total of 11 experts participated in round one (response rate: 84.6%), and 8 experts participated in round two (response rate: 72.7%). In round one, a strong expert agreement on module relevance was found. In round two, all modules achieved a mean ranking > 3.0, with the lowest being 3.9. Modules with a mean ranking > 3.0 were considered relevant and included in the final analysis. Two modules had an SD > 1.0, indicating a lack of consensus.

***Conclusions*:** This study established expert consensus on 32 essential training modules for the Eyesi slit lamp simulator, supporting their integration into an ophthalmology residency curriculum. Further testing under an evidence-based framework is required to integrate them into a structured curriculum and assess their impact on clinical skills.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40123-025-01175-2.

K. R. Hansen · S. J. Lowater (✉) · K. S. Lyngsø · A. H. K. Nissen · A. S. Vergmann
Research Unit of Ophthalmology, Department of Ophthalmology, Odense University Hospital, J.B. Winsløws Vej 4, 5000 Odense C, Denmark
e-mail: simon.joel.lowater@rsyd.dk

K. R. Hansen · S. J. Lowater · K. S. Lyngsø · A. H. K. Nissen · A. S. Vergmann
Department of Clinical Research, University of Southern Denmark, Odense, Denmark

K. R. Hansen · A. S. Vergmann
Regional Center for Technical Simulation, Region of Southern Denmark, Odense, Denmark

PLAIN LANGUAGE SUMMARY

This study aims to identify which training modules on the Eyesi slit lamp simulator are most useful for eye doctors in training. A slit lamp is a special microscope that enables doctors to look at different parts of the eye in detail, including the retina found in the back of the eye. It is very important for eye doctors to be skilled in using this tool, but it is not usually taught in medical school. The Eyesi slit lamp simulator is a tool that helps them practice eye examinations. To find the best training modules, we asked a group of eye specialists for their opinions by using a two-step survey called a Delphi study. In the first step, the eye specialists decided which modules were useful. In the second step, they rated the useful modules on a scale from 1 to 5, with 1 meaning "not useful" and 5 meaning "very useful." We then measured the agreement among the specialists to decide which modules should be included in the training program. The results showed that the eye specialists agreed on 32 modules for training. By identifying the most useful modules on the Eyesi slit lamp simulator, we hope to create a structured training program. This program will help future eye doctors to develop their skills for patient examinations and improve patient safety.

Keywords: Delphi study; Eyesi slit lamp simulator; Ophthalmology; Residency; Training curriculum

Key Summary Points

Simulation-based training is increasingly applied in ophthalmology to teach relevant skills, but there is a lack of curricula specifically designed for ophthalmology residents.

Therefore, we conducted this study, which aims to identify relevant training modules on the Eyesi slit lamp simulator for ophthalmology residents through a national Delphi study to develop a training curriculum.

A total of 32 training modules were identified as relevant for ophthalmology residents.

The findings contribute to the development of a structured, evidence-based training program. However, further investigation is needed to assess the impact of these modules on the clinical skills of residents and patient safety.

INTRODUCTION

Simulation-based training has become an increasingly popular and established part of various fields within medicine [1–4]. The objective of utilizing simulation-based training is to prepare and provide physicians with the opportunity to practice and hone their skills relevant to their field. Simulation-based training is increasingly used to acquire basic ophthalmological skills in eye surgery using the Eyesi Surgical Simulator (VRMagic GmbH, Haag-Streit Simulation, Mannheim, Germany) [5–13]. However, eye surgery is only performed once the resident has completed or nearly completed specialist training in ophthalmology. In medical school, training in ophthalmology tools and examination techniques is not part of the standard curriculum. As a result, new ophthalmology residents must learn to use specialized instruments, such as the slit lamp, during their residency. The slit lamp is an essential tool for examining the eye in detail but mastering it can be challenging. While residents receive some supervised training, they are largely responsible for developing proficiency on their own. Simulation-based

training can help bridge this gap by providing a safe and structured way for residents to build confidence and improve their skills before independently examining patients. This may reduce the risk of missing sight-threatening conditions due to limited experience.

Simulation-based training is superior to traditional training in medical education, particularly when doctors are trained in specific clinical goals [14, 15]. The Eyesi slit lamp (VRMagic GmbH, Haag-Streit Simulation, Mannheim, Germany) is a new simulator designed for training in slit lamp techniques, and indirect ophthalmoscopy (Fig. 1), thus mirroring clinical practices. However, before implementation of the simulator in a training curriculum for new ophthalmologists, testing under a proper evidence-based framework is needed to ensure efficient training in the required curriculum.

The objective of this study is to identify relevant training modules on the Eyesi slit lamp simulator for ophthalmology residents through a national Delphi study to develop a training curriculum.

Fig. 1 The Eyesi Slit Lamp Simulator by Haag-Streit, designed for simulation-based training in ophthalmology

METHODS

The Eyesi Slit Lamp Simulator

In the present study, we use the Eyesi slit lamp simulator (VRMagic GmbH, Haag-Streit Simulation, Mannheim, Germany). This simulator is integrated into the hardware of a Haag-Streit BQ 900 slit lamp, and the system replicates all real slit lamp functions including slit width, beam rotation, filters, and illumination adjustments, which provides realistic and real-time visualization of intraocular structures. Users can examine virtual patients with a biomicroscope, a fundoscopy lens, or a gonioscopy lens. A four-tier structured curriculum is built into the Eyesi slit lamp, offering a self-guided learning experience that includes both learning and exam modules and focuses on different key aspects of a slit lamp examination. The learning modules provide guided instructions, while the exam modules require users to apply what they learned independently from the corresponding learning module (Figs. 2 and 3). The beginner courses introduce slit lamp handling and illumination techniques, while advanced courses cover a wide range of clinical cases and standard grading systems. Users will progress through the tiers sequentially with supplementary online educational content integrated to complement practical training. A built-in scoring measurement system evaluates performance objectively and provides training reports [16].

Module Testing

A comprehensive in-depth testing of each module on the Eyesi slit lamp focusing on identifying the content and testing parameters of each module was conducted. Subsequently, a thorough description of each module was prepared. Emphasis was particularly placed on modules that address the official standards of competence concerning medical expertise, and clinical

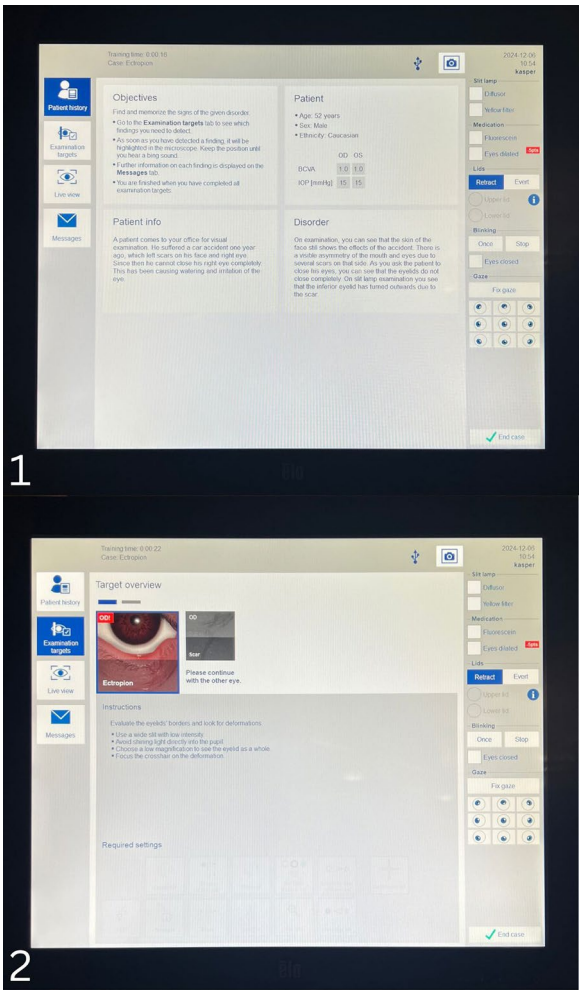


Fig. 2 Still images of the learning module for Ectropion on the Eyesi slit lamp simulator. Picture 1 displays patient information, and picture 2 displays examination targets that can guide users in identifying the correct diagnosis

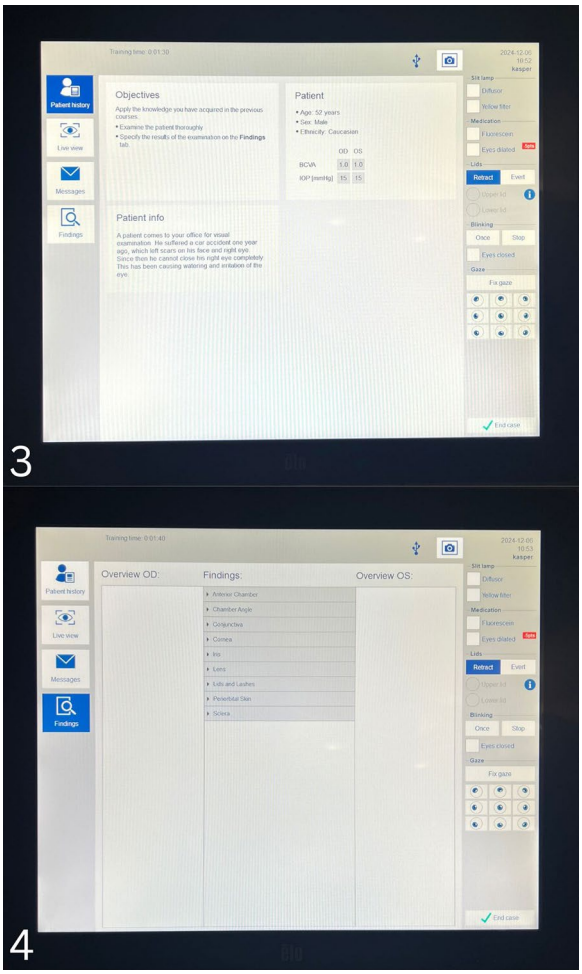


Fig. 3 Still images of the exam module for Ectropion on the Eyesi slit lamp simulator. Picture 3 displays patient information, and picture 4 displays the findings window for selecting the diagnosis

competencies in ophthalmology residency training [17].

Delphi Methodology

Following the detailed description of each module we conducted a national Delphi study.

A Delphi study is a systematic process of consulting, gathering, evaluating, and organizing relevant information, and opinions from a group of experts within a field without the experts meeting each other. The Delphi study benefits from being able to obtain the viewpoint of a panel of experts without yielding to undue sway from specific individuals. Moreover, it can be managed through suitable feedback and adjustments to steer conclusions toward a consensus within a group of experts [18, 19].

Chief physicians, consultants responsible for education, and professors from all ophthalmology departments across Denmark were invited to participate via email. Active involvement in the current ophthalmology training of ophthalmology residents was a requirement for participation. The questionnaires were designed in Google Forms (Google

LLC, Mountain View, CA, USA), and distributed via email. The experts were given 3 weeks to answer in round one and 5 weeks in round two. We sent a reminder email to the participants a week before the deadline. The study proceeded in two rounds:

Round 1

In round one, experts were asked which of the modules available on the Eyesi slit lamp simulator were relevant for a resident in ophthalmology. They received a questionnaire with a small and precise description of each module and had to rank the module as "relevant" or "not relevant." Consensus rates in Delphi studies are not well defined and vary across studies [20, 21]. Therefore, we assumed a consensus threshold of 75% agreement on relevance in the first round would be sufficient to reflect a strong majority opinion among the experts. See Appendix 1 in Supplementary Material for a description of each module on the Eyesi slit lamp simulator.

Round 2

In round two, the experts received a detailed description of the modules for which at least 75% of the experts had ranked the module as "relevant" in round one. The experts were asked to assess the relevance of each module and rank the relevance on a linear scale from 1 to 5, where 1 represented "not relevant" and 5 represented "highly relevant." See Appendix 1 in Supplementary Material for a description of each module on the Eyesi slit lamp simulator.

Statistics

The data from round one were analyzed based on the percentage of the panel agreeing on relevance. For round two, analysis included a combination of measures; mean, standard deviation (SD), interquartile range (IQR), and an assessment of outliers.

The data were analyzed using StataCorp. 2023. *Stata Statistical Software: Release 18*. College Station, TX, USA: StataCorp LLC.

Measurement of Consensus

Based on the results from round one, relevant modules were identified for inclusion in round two. In round two, these modules had to meet the criteria of (1) a minimum mean score of 3 (based on the mean score of the experts' ratings), (2) no scores of 1, and (3) an assessment of consensus among experts' ratings to be included in subsequent evaluation. Consensus was determined as $SD < 1.0$, as described in guidelines for expert consensus evaluations [22]. SD measures the overall spread or variability of the expert ratings. A higher SD indicates more variability, signaling lower agreement among experts, while a low SD indicates high agreement among experts [23].

The interquartile range evaluates the middle 50% of the ratings, providing insight into how consistent the expert ratings were. A smaller IQR suggests that most experts gave similar ratings, indicating higher agreement, while a larger IQR indicates greater variability in their ratings. Outliers are ratings that differ significantly from others. These were identified using the standard boxplot method, where any expert rating that was below the lower fence ($Q1 - (1.5 \times IQR)$) or above the upper fence ($Q3 + (1.5 \times IQR)$) was considered an outlier [24].

Ethics

The participants agreed to participate before we initiated the study and were aware that it was a research study and that the results would be published.

The study does not involve sensitive or confidential information; therefore, no notification to the Data Protection Agency or need of approval from the Research Ethics Committee was required. According to Sect. 14(2) of the Danish Act on Research Ethics Review of Health Research Projects, health research studies based solely on questionnaires and interviews that do not involve human biological material are not subject to notification to the committee.

RESULTS

Based on the expert feedback and following the exclusion of two modules due to high variability in ratings, consensus was reached on a total of 32 modules, which were identified as relevant for residents in ophthalmology.

A total of 28 ophthalmologists across all ophthalmology departments in Denmark involved with ophthalmology training were invited to participate. A total of 13 agreed to participate (46.4%).

In round one, 11 out of 13 participants responded, yielding a response rate of 84.6%. The regional distribution of participants was 18.2% from the Capital Region of Denmark, 36.4% from the Region of Southern Denmark, 18.2% from Region Zealand, 18.2% from the Region of Central Denmark, and 9.1% from the Region of Northern Denmark. In round two, 8 out of 11 participants responded, resulting in a response rate of 72.7%. The regional distribution of participants in the second round was 25% from the Capital Region of Denmark, 37.5% from the Region of Southern Denmark, 12.5% from Region Zealand, 12.5% from the Region of Central Denmark, and 12.5% from the Region of Northern Denmark.

Learning and exam modules associated with each other were compiled into a single description. This resulted in a total of 27 modules in round one. Of these, 21 modules were ranked as relevant by at least 75% of the participants (Table 1). The remaining six modules were excluded from further evaluation. For round two, the 21 relevant modules were split into individual learning and exam modules, leading to a total of 34 modules to be ranked. A more detailed description of each module was included in the questionnaire for round two. In this round, all 34 modules received a mean score of at least 3, with the lowest mean score being 3.9 (Table 2). No modules were ranked with a score of 1. The progression through the two rounds of the evaluation process is illustrated in the flowchart in Fig. 4.

The SD of the expert rankings was calculated for each module (Table 2). Overall, the experts agreed during the ranking of the modules with an SD < 1. Two of the modules "A1—background and diffuse illumination" and "B1—free training" had an SD > 1 and lacked consensus between the experts. IQR was also calculated for all modules to compare with the SD. The two modules with an SD > 1 both had an IQR of 2, indicating a higher variability in expert ratings compared with the modules, where the IQR was 1 or lower. One exception was module "A2—specular reflections," which scored an IQR of 1.5. This led to the exclusion of "A1—background and diffuse illumination" and ">B1—free training" from further validation, resulting in the identification of 32 modules.

We calculated the upper and lower fences of a boxplot (Table 2) for each module to adjust for outliers, to check if they affected the SD and thus impacted the consensus. In the outlier analysis, we found one expert had nine outliers whereas the maximum number of outliers among the other experts was two.

To evaluate the impact of this discrepancy, we performed a post-hoc analysis by recalculating the mean SD and IQR excluding the expert with nine outliers to assess the impact of the outliers. Exclusion of the expert with nine outliers resulted in a reduction of the mean SD for all modules from 0.55 to 0.47, and a reduction in the mean IQR from 0.63 to 0.53, indicating higher consensus among the remaining experts. The overall mean ranking for all modules increased from 4.61 to 4.70 when excluding the expert with nine outliers (Table 3). For the modules "A1—background and diffuse illumination" and "B1—free training" the IQR value remained the same after adjusting for ratings of the excluded expert. However, we observed a slight increase in SD for these two modules: "A1—Background and diffuse illumination" had an SD increase from 1.05 to 1.07, and "B1—free training" had an SD increase from 1.17 to 1.20, indicating a marginal decrease in agreement among the remaining experts.

Statistical analyses were conducted to assess the significance of these observations. The Mann–Whitney U test yielded a Z-statistic of 1.64 and a p-value of 0.10 for the two-tailed test, with an exact p-value of 0.25, indicating insufficient statistical evidence to conclude that the expert with nine outliers had significantly more

Table 1 Overview of expert ranking of module relevance in round one

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	Expert 11	Percent R	Percent NR
A1 basic light microscopy	R	R	R	R	R	R	R	R	R	R	R	100	0
A1 navigation training	R	R	R	R	R	R	R	NR	R	R	R	90.90	9.10
A1 slit light adjustment	R	R	R	R	R	R	R	R	R	R	R	100	0
A1 illumination and viewing direction	R	R	R	R	R	R	R	R	R	R	R	100	0
A1 background and diffuse illumination	R	R	R	R	R	R	R	NR	NR	R	R	81.82	18.18
A1 fundoscopy lens handling	R	R	R	R	R	R	R	R	R	R	R	100	0
A1 gonioscopy lens handling	R	R	R	R	R	R	R	R	R	R	R	100	0
A2 diffuse and focal illumination	R	R	R	R	R	R	R	NR	R	R	R	90.90	9.10
A2 illumination techniques in transparent media	NR	R	R	NR	R	R	R	NR	R	R	R	72.70	27.30
A2 the narrow slit	R	R	R	R	R	R	R	NR	R	R	R	90.90	9.10
A2 specular reflections	R	R	R	R	R	R	NR	R	NR	R	R	81.82	18.18
A2 gonioscopic examination of corneal wedge	R	R	R	R	R	R	R	NR	R	R	R	90.90	9.10

Table 1 continued

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	Expert 11	Percent R	Percent NR
B1 examination of the anterior segment	R	R	R	R	R	R	R	R	R	R	R	100	0
B1 free training	R	R	R	R	R	R	R	R	R	R	R	100	0
C1 the anterior segment exam and corresponding learning modules (lids and lashes, conjunctiva, cornea, lens, uveitis, and glaucoma)	R	R	R	R	R	R	R	R	R	R	R	100	0
C1 retina: optic disc learning and exam modules	R	R	R	R	R	R	R	R	R	R	R	100	0
C1 retina: AMD learning and exam modules	R	R	R	R	R	R	R	R	R	R	R	100	0
C1 retina: diabetic retinopathy learning and exam modules	R	R	R	R	R	R	R	R	R	R	R	100	0
C1 gonioscopy learning and exam modules	NR	R	R	R	R	R	R	R	R	R	R	90.90	9.10
C2 cataract learning and exam modules	R	R	R	R	R	NR	NR	NR	R	R	NR	63.60	36.40

Table 1 continued

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	Expert 11	Percent R	Percent NR
C2 corneal neovascularization learning and exam modules	NR	R	NR	R	R	NR	NR	NR	R	R	NR	45.50	54.50
C2 van Herick angle width learning and exam modules	R	R	R	R	R	R	R	R	R	R	R	100	0
C2 cells and flare learning and exam modules	R	R	R	R	R	R	R	R	R	R	R	100	0
C2 trachoma learning module	NR	NR	NR	NR	R	NR	NR	NR	NR	R	R	27.30	72.70
C2 Shaffer–Kanski angle width learning and exam module	NR	R	R	R	R	NR	NR	NR	R	R	R	63.60	36.40
C2 Spaeth angle classification learning and exam module	NR	NR	NR	R	R	NR	NR	NR	NR	R	NR	27.30	73.70
D retina learning and exam modules	R	R	R	R	R	R	R	R	R	R	R	100	0

The table displays the rankings and percentage of responses for relevant (R) and not relevant (NR) provided by experts for each module

AMD Age-related macular degeneration

Table 2 Overview of expert rankings and consensus measures in round two

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Mean	SD	Q1	Q3	IQR	Upper fence	Lower fence	Mean no expert with nine outliers	SD no expert with nine outliers	IQR no expert with nine outliers
A1 basic light microscopy	5	4	5	4	5	5	5	5	4.75	0.43	4.50	5	0.50	3.75	5.75	4.86	0.37	0
A1 navigation training	5	4	5	5	5	5	5	5	4.88	0.33	5	5	0	5	5	5	0.13	0
A1 slit light adjustment	5	3	5	5	5	4	5	5	4.63	0.70	4.50	5	0.5	3.75	5.75	4.86	0.42	0
A1 illumination and viewing direction	5	4	5	5	5	5	5	5	4.88	0.33	5	5	0	5	5	5	0.13	0
A1 background and diffuse illumination	5	3	4	3	4	2	5	5	3.88	1.05	3	5	2	0	8	4	1.08	2
A1 fundoscopy lens handling	5	4	5	5	5	5	5	5	4.88	0.33	5	5	0	5	5	5	0.13	0

Table 2 continued

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Mean	SD	Q1	Q3	IQR	Upper fence	Lower fence	Mean no expert with nine outliers	SD no expert with nine outliers	IQR no expert with nine outliers
A1 gonioscopy lens handling	5	4	5	5	5	5	5	5	4.88	0.33	5	5	0	5	5	5	0.13	0
A2 diffuse and focal illumination	4	4	5	4	5	3	5	5	4.38	0.70	4	5	1	2.50	6.50	4.43	0.73	1
A2 the narrow slit	5	4	5	4	5	5	5	5	4.75	0.43	4.50	5	0.50	3.75	5.75	4.86	0.37	0
A2 specular reflections	4	3	4	3	5	5	5	5	4.25	0.83	3.50	5	1.50	1.25	7.25	4.43	0.75	1
A2 gonioscopic examination of corneal wedge	5	3	5	5	5	5	5	5	4.75	0.66	5	5	0	5	5	5	0.25	0
B1 examination of the anterior segment	5	5	5	5	5	4	5	5	4.86	0.33	5	5	0	5	5	4.86	0.35	0
B1 free training	3	5	3	2	5	5	5	5	4.13	1.17	3	5	2	0	8	4	1.20	2

Table 2 continued

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Mean	SD	Q1	Q3	IQR	Upper fence	Lower fence	Mean no expert with nine outliers	SD no expert with nine outliers	IQR no expert with nine outliers
C1 the anterior segment. Lids and lashes learning module	5	3	5	5	5	5	4	5	4.63	0.70	4.50	5	0.50	3.75	5.75	4.86	0.42	0
C1 the anterior segment. Conjunctiva learning module	5	3	5	4	5	5	4	5	4.5	0.71	4	5	1	2.50	6.50	4.17	0.5	1
C1 the anterior segment. Cornea learning module	5	4	5	5	5	5	4	5	4.75	0.43	4.50	5	0.50	3.75	5.75	4.86	0.37	0
C1 the anterior segment. Lens learning module	5	4	5	5	5	4	5	5	4.75	0.43	4.50	5	0.50	3.75	5.75	4.86	0.37	0

Table 2 continued

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Mean	SD	Q1	Q3	IQR	Upper fence	Lower fence	Mean no expert with nine outliers	SD no expert with nine outliers	IQR no expert with nine outliers
C1 the anterior segment. Uveitis learning module	5	5	5	5	5	5	5	5	5	0	5	5	0	5	5	5	0	0
C1 the anterior segment. Glaucoma learning module	5	4	5	5	5	5	5	5	4.88	0.33	5	5	0	5	5	5	0.13	0
C1 the anterior segment exam module	4	4	5	4	3	4	4	5	4.13	0.60	4	4.50	0.50	3.25	5.25	4.14	0.64	1
C1 retina: optic disc learning modules	5	4	5	5	5	5	5	5	4.88	0.33	5	5	0	5	5	5	0.13	0
C1 retina: optic disc exam module	5	4	5	4	3	5	5	5	4.5	0.71	4	5	1	2.50	6.50	4.57	0.73	1

Table 2 continued

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Mean	SD	Q1	Q3	IQR	Upper fence	Lower fence	Mean no expert with nine outliers	SD no expert with nine outliers	IQR no expert with nine outliers
C1 retina: AMD learning modules	5	4	5	5	3	5	4	5	4.5	0.71	4	5	1	2.50	6.50	4.57	0.73	1
C1 retina: AMD exam module	5	4	5	5	3	5	4	5	4.5	0.71	4	5	1	2.50	6.50	4.57	0.73	1
C1 retina: diabetic retinopathy learning modules	5	4	5	5	5	5	4	5	4.75	0.43	4.50	5	0.50	3.75	5.75	4.86	0.37	0
C1 retina: diabetic retinopathy exam module	5	5	5	4	4	5	4	5	4.63	0.48	4	5	1	2.50	6.50	4.57	0.50	1
C1 gonioscopy learning modules	5	4	5	5	4	5	4	5	4.63	0.48	4	5	1	2.50	6.50	4.71	0.46	1
C1 gonioscopy exam module	5	4	5	4	4	3	4	5	4.25	0.66	4	5	1	2.50	6.50	4.29	0.70	1

Table 2 continued

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Mean	SD	Q1	Q3	IQR	Upper fence	Lower fence	Mean no expert with nine outliers	SD no expert with nine outliers	IQR no expert with nine outliers
C2 van Herick angle width learning modules	5	4	5	5	5	5	4	5	4.75	0.43	4.50	5	0.50	3.75	5.75	4.86	0.37	0
C2 van Herick angle width exam module	5	4	5	4	4	2	4	5	4.13	0.93	4	5	1	2.50	6.50	4.14	0.99	1
C2 cells and flare learning modules	5	3	5	5	4	5	4	5	4.50	0.71	4	5	1	2.50	6.50	4.14	0.50	1
C2 cells and flare exam module	5	4	5	5	4	5	4	5	4.63	0.48	4	5	1	2.50	6.50	4.71	0.46	1
D retina learning modules	5	5	5	5	5	5	4	5	4.88	0.33	5	5	0	5	5	4.86	0.35	0

Table 2 continued

Module	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Mean	SD	Q1	Q3	IQR	Upper fence	Lower fence	Mean no expert with nine outliers	SD no expert with nine outliers	IQR no expert with nine outliers
D retina exam module	5	5	5	5	4	5	4	5	4.75	0.43	4.50	5	0.50	3.75	5.75	4.17	0.45	1

The table presents experts' rankings for each module in round two alongside measures used to interpret consensus, including the mean, standard deviation (SD), interquartile range (IQR), and adjusted values excluding the expert with nine outliers

outliers than the expert with two outliers. The Wilcoxon signed-rank test, however, revealed a significant difference in the mean SD when excluding the expert with nine outliers ($Z = 3.03$, $p = 0.0024$, exact $p = 0.0018$), confirming that the exclusion led to increased agreement among the remaining experts.

DISCUSSION

This study aims to identify relevant training modules on the Eyesi slit lamp simulator for ophthalmology residents through a national Delphi study to develop a training curriculum. The experts mostly agreed on which modules were relevant to include in the training curriculum, resulting in a consensus on 32 modules.

This study builds on current research in simulation-based training in medical education, which supports the use of simulation-based training to enhance proficiency in a variety of medical procedures [25–28]. Contrary to these studies, the focus of this study was to establish expert consensus regarding the essential components on the Eyesi slit lamp simulator for an ophthalmology residency training curriculum. By identifying the relevant training modules, we aim to shape a structured, evidence-based training program specifically for slit lamp examination and indirect ophthalmoscopy. Simulation-based training has been shown to improve proficiency across multiple specialties and its increasing use in ophthalmology, especially through the use of simulators such as the Eyesi platform, reflects broader trends in medical education [29]. For instance, Wang et al., showed specific requests for increased training in ophthalmic skills for emergency medicine residents [30]. An improvement in proficiency and reduction in errors may lead to greater cost-effectiveness compared with traditional medical training. This has already been shown in other specialties [31, 32]. This study contributes to the existing body of evidence by outlining a training program designed to enhance the proficiency, and skills of residents in slit lamp examination and indirect ophthalmoscopy through simulation-based training. Further research directly

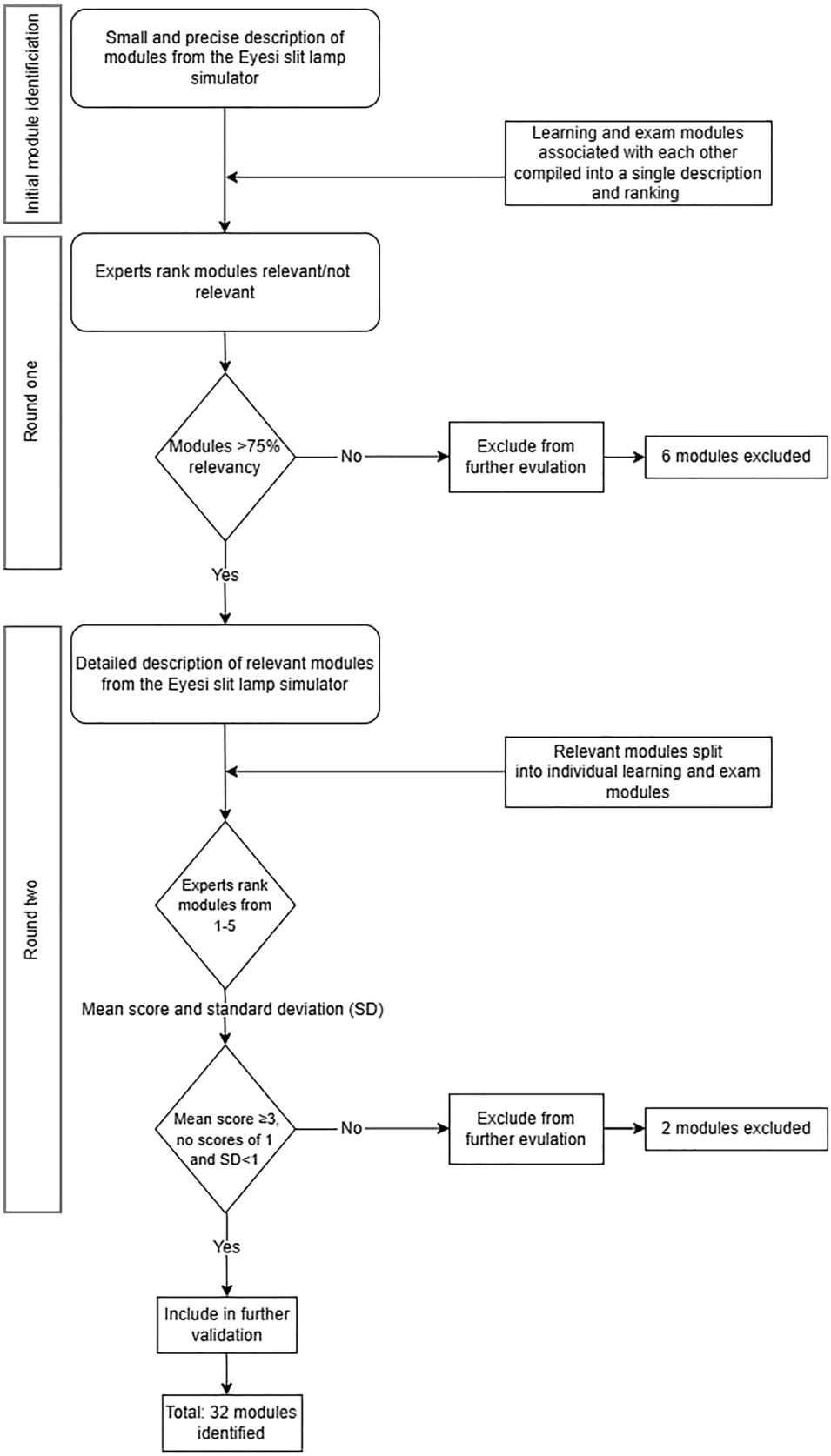
comparing simulation-based training to traditional training methods is needed to assess if tools such as the Eyesi slit lamp simulator improves learning efficiency, diagnostic accuracy, and overall clinical performance.

The use of the Delphi methodology enabled the gathering of expert opinions in a structured manner, contributing to the reliability of the results [33]. The anonymity of the Delphi process helps to reduce potential influence of individuals who might be more outspoken and control the direction of a discussion in a real-life meeting. The absence of full feedback between rounds and the limited number of rounds may result in less reduction in bias. Delphi studies typically involve multiple rounds with iterative feedback, which allows experts to adjust their opinions in response to the groups developing consensus, thus leading to a more refined agreement. The fixed structure of the simulator limited further discussion as the format of the simulator was predetermined and offered no opportunities for adjustments. Yet, even with only two rounds and less feedback, we reached a strong agreement on most relevant training modules. The national extent of this study comprising experts from ophthalmology departments across Denmark strengthens the generalizability and relevance of the findings to ophthalmology residency training on a national level. Additionally, the consensus-driven approach ensures an alignment of the selected modules with expert recommendations.

In our study, we received a low response rate on the initial invitation to participate in the study. A total of 11 out of 28 invited experts answered the questionnaire in the first round, providing a response rate of only 39%. The low response rate may have been a consequence of the fact that we only sent out one email for the initial participation. However, we did send out reminders during round one, which yielded a response from 11 out of 13 experts, who agreed to participate. A similar methodology was used in round two. Sending out a reminder doing the invitation period may have helped us gather a larger expert panel. Another explanation could be that the invited experts are in senior positions with tight and busy schedules potentially hindering participation. Panel sizes in Delphi studies vary widely and there are no defined standard panel sizes [33–35]. Our panel of 11 experts, although on the lower end, represented at least 1 expert from each of the five Danish regions in both rounds, thus improving the generalizability of our findings. There are no established rules regarding acceptable response or participation rates in Delphi studies, but a commonly suggested benchmark for each round is a rate of at least 70% [36–38]. Although there is no minimally acceptable response rate, a response rate of 60% has been used as the threshold of acceptability [39]. While the current study reached a participation rate during the initial round of only 39% we did, however, receive a response rate of 84.6% and 61.5% during round one and two, respectively, from the experts who agreed to participate.

The decision to use binary questions to assert consensus in the first round of the Delphi study was due to the fixed and unchangeable design of the Eyesi slit lamp simulator. Given the structure of the simulator not being modifiable, there was limited opportunity for further adjustments based on expert input. This made structured binary responses more suitable for the first round of the Delphi study. In round one, we chose to provide small and precise descriptions of each module, as the purpose of this round was for the experts to rank the main modules according to their overall relevance to residents. In round two, we provided a more detailed description of each relevant module to allow the experts to grade the submodules under each main module. Inclusion of detailed descriptions of the modules in round one might have increased transparency, but our intention was to maintain a broader focus in the initial round and not overwhelm the experts with information.

In the second round, we decided to use a linear scale instead of a Likert scale. A linear scale is intuitive for most respondents, and the idea of the center is more natural [22]. The data were analyzed using parametric methods because they fit well with the use of a linear scale. Additionally, the mean, and SD are easily understandable for many practitioners making them rational for communicating the results. To define consensus


Initial module identification
Small and precise description of modules from the Eyesi slit lamp simulator
Learning and exam modules associated with each other compiled into a single description and ranking
Round one
Experts rank modules relevant/not relevant
Modules >75% relevancy
No
Exclude from further evulation
6 modules excluded
Yes
Round two
Detailed description of relevant modules from the Eyesi slit lamp simulator
Relevant modules split into individual learning and exam modules
Experts rank modules from 1-5
Mean score and standard deviation (SD)
Mean score ≥3, no scores of 1 and SD<1
No
Exclude from further evulation
2 modules excluded
Yes
Include in further validation
Total: 32 modules identified

◂**Fig. 4** Flowchart outlining the selection process of relevant training modules for ophthalmology residents, based on expert evaluation in two rounds with criteria for inclusion and exclusion based on relevance and consensus

we used SD rather than IQR [22], as SD incorporates the full distribution of responses, thus capturing a wider spectrum of expert opinion. This is particularly useful in understanding consensus where variation in responses might be significant. We defined consensus as an SD threshold of < 1.0, which minimizes the risk of false consensus to less than 10% for sample sizes above 10. Even with only eight experts in round two, the risk remained under 15% as shown in fig. 5 by Franc et al. [22].

Outliers can have a significant effect on statistical variables, such as mean, SD, and IQR [40]. Therefore, we decided to identify and evaluate outliers within our data in round two. One expert had a total number of nine outliers across the 34 modules, compared with a maximum of two outliers for any other expert. One expert expressed via email skepticism about simulation-based learning and a preference for hands-on practice. This perspective may have contributed to the discrepancies observed in their ratings, thus highlighting the potential influence of individual biases on consensus metrics. In the questionnaire, we did not include an option for the experts to fill in reasons for their respective ratings during the two rounds. This method might have helped us in assessing exact reasons for one expert having nine outliers.

To assess the impact of these discrepancies, we conducted a post hoc analysis by comparing the number of outliers using the Mann–Whitney *U* test. As a nonparametric test, it is particularly suitable for comparing two independent groups without assuming that the data are normally distributed [41]. A significance level of $\alpha = 0.05$ was established for this test. However, the results found no statistical evidence that the expert with nine outliers had significantly more than the others. This is likely due to the small sample size, which may have limited the statistical power of the test [42]. To further evaluate the influence of this experts' ratings, we recalculated the mean SD and IQR after their exclusion. The mean SD decreased from 0.55 to 0.47, while the mean IQR decreased from 0.63 to 0.53, suggesting a higher level of consensus among the remaining experts. To assess the statistical significance of this change, we performed a Wilcoxon signed–rank test, a nonparametric test that does not assume normality and is robust to outliers [43]. A significance level of $\alpha = 0.05$ was also used for this test. The test confirmed a significant difference in the mean SD with and without the expert with nine outliers, indicating improved agreement among the remaining experts. Despite this, the exclusion did not affect the selection of modules included in the training curriculum.

Table 3 Summary statistics for expert rankings in round two

Variable	Mean	SD
Mean ranking for all modules	4.61	0.27
Mean ranking for all modules (excluding expert with nine outliers)	4.70	0.29
Mean SD of rankings for all modules	0.55	0.24
Mean SD of rankings for all modules (excluding expert with nine outliers)	0.47	0.29
Mean IQR of rankings for all modules	0.63	0.55
Mean IQR of rankings for all modules (excluding expert with nine outliers)	0.53	0.61

This table presents the overall mean, standard deviation (SD), and interquartile range (IQR) for all modules ranked in round two, both including and excluding the expert with nine outliers. These values represent the average metrics across all modules combined

Future studies will aim to develop an evidence-based training program, testing the modules from this study for integration into a structured curriculum. This will play a critical role in utilizing evidence-based training for ophthalmology residents, further strengthening the educational role of simulation-based training.

This study acknowledges certain limitations such as the fixed structure of the Eyesi simulator, which restricted adjustments based on expert feedback and the small sample size in the second round, which may have limited the diversity of perspectives. The lack of feedback from the experts during the rounds could also have provided valuable insight for the future planned studies on the simulator. Despite these limitations, the strengths of the current study strongly support its findings. The nationwide scope and strong consensus among experts significantly enhance the applicability of the findings to ophthalmology residency training in Denmark. Moreover, the structured Delphi approach ensured a systematic and robust gathering of expert opinions. In spite of a smaller expert group, the high level of consensus achieved highlights the values of the study in identifying essential training modules.

CONCLUSIONS

This study demonstrates a strong expert consensus on the relevance of the identified modules for residency training in simulated slit lamp examination. While the results highlight the relevance of the identified modules for inclusion in a training curriculum, further research and evidence-based testing are needed to confirm their impact on improving competencies of residents in a real-world clinical scenario. This research could potentially aid in shaping a future ophthalmology training curricula for slit lamp examination.

Medical Writing, Editorial, or Other Assistance. No medical writing, editorial, or other assistance was received during the writing of this article.

Author Contributions. Kasper Ruotsalainen Hansen, Simon Joel Lowater, Kristina Sanne Lyngsø, Anne Helene Køllund Nissen, and Anna Stage Vergmann contributed to the study conception and design. Material preparation, data collection and analysis were performed by Kasper Ruotsalainen Hansen and Anna Stage Vergmann. The first draft of the manuscript was written by Kasper Ruotsalainen Hansen, and Kasper Ruotsalainen Hansen, Simon Joel Lowater, Kristina Sanne Lyngsø, and Anne Helene Køllund Nissen, and Anna Stage Vergmann commented on previous versions of the manuscript. Kasper Ruotsalainen Hansen, Simon Joel Lowater, Kristina Sanne Lyngsø, Anne Helene Køllund Nissen, and Anna Stage Vergmann read and approved the final manuscript.

Funding. The Eyesi slit lamp simulator used in this study was funded by the Toyota Foundation and the Grosserer L. Foght's Foundation. No funding or sponsorship was received for the publication of this article. No funding agents had any role in the conduct, analysis, or presentation of the study.

Data Availability. The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of Interest. Kasper Ruotsalainen Hansen, Simon Joel Lowater, Kristina Sanne Lyngsø, Anne Helene Køllund Nissen, and Anna Stage Vergmann declare that they have no competing interests. Anna Stage Vergmann is an Editorial Board member of Opthamology and Therapy. Anna Stage Vergmann was not involved in the selection of peer reviewers for the manuscript nor any of the subsequent editorial decisions.

Ethical Approval. The participants agreed to participate before we initiated the study and were aware that it was a research study and that the results would be published. The study does not involve sensitive or confidential

information;therefore, no notification to the Data Protection Agency or approval from the Research Ethics Committee was required. According to Sect. 14(2) of the Danish Act on Research Ethics Review of Health Research Projects, health research studies based solely on questionnaires and interviews that do not involve human biological material are not subject to notification to the committee.

REFERENCES

1. Naur TMH, Nilsson PM, Pietersen PI, Clementsen PF, Konge L. Simulation-based training in flexible bronchoscopy and Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration (EBUS-TBNA): a systematic review. Respiration. 2017;93(5):355–62.

2. Lorello GR, Cook DA, Johnson RL, Brydges R. Simulation-based training in anaesthesiology: a systematic review and meta-analysis. Br J Anaesth. 2014;112(2):231–45.

3. Haiser A, Aydin A, Kunduzi B, Ahmed K, Dasgupta P. a systematic review of simulation-based training in vascular surgery. J Surg Res. 2022;279:409–19.

4. Bjerrum F, Thomsen ASS, Nayahangan LJ, Konge L. Surgical simulation: current practices and future perspectives for technical skills training. Med Teach. 2018;40(7):668–75.

5. Forslund Jacobsen M, Konge L, la Cour M, Holm L, Kjaerbo H, Moldow B, et al. Simulation of advanced cataract surgery-validation of a newly developed test. Acta Ophthalmol. 2020;98(7):687–92.

6. Petersen SB, Vestergaard AH, Thomsen ASS, Konge L, Cour M, Grauslund J, et al. Pretraining of basic skills on a virtual reality vitreoretinal simulator: A waste of time. Acta Ophthalmol. 2021;100(5), e1074–e1079. https://doi.org/10.1111/aos.15039

7. Thomsen AS, Bach-Holm D, Kjærbo H, Højgaard-Olsen K, Subhi Y, Saleh GM, et al. Operating room performance improves after proficiency-based virtual reality cataract surgery training. Ophthalmology. 2017;124(4):524–31.

8. Thomsen AS, Kiilgaard JF, Kjaerbo H, la Cour M, Konge L. Simulation-based certification for cataract surgery. Acta Ophthalmol. 2015;93(5):416–21.

9. Thomsen AS, Smith P, Subhi Y, Cour M, Tang L, Saleh GM, et al. High correlation between performance on a virtual-reality simulator and real-life cataract surgery. Acta Ophthalmol. 2017;95(3):307–11.

10. Thomsen ASS, Kiilgaard JF, la Cour M, Brydges R, Konge L. Is there inter-procedural transfer of skills in intraocular surgery? A randomized controlled trial. Acta Ophthalmol. 2017;95(8):845–51.

11. Mellum ML, Vestergaard AH, Grauslund J, Vergmann AS. Virtual vitreoretinal surgery: effect of distracting factors on surgical performance in medical students. Acta Ophthalmol. 2020;98(4):378–83.

12. Vergmann AS, Vestergaard AH, Grauslund J. Virtual vitreoretinal surgery: validation of a training programme. Acta Ophthalmol. 2017;95(1):60–5.

13. Rasmussen RC, Grauslund J, Vergmann AS. Simulation training in vitreoretinal surgery: a systematic review. BMC Ophthalmol. 2019;19(1):90.

14. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Acad Med. 2011;86(6):706–11.

15. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. Acad Med. 2013;88(8):1178–86.

16. Haag-Streit. Eyesi slit lamp: Haag-Streit; 2024 [Available from: https://haag-streit.com/en/products/categories/simulators-training/training-simulators/eyesi-slit-lamp. Accessed 28 Nov 2024.

17. DOS. Specialllægeuddannelsen [Available from: https://dansk-oftalmologisk-selskab.dk/speciallaegeuddannelsen/. Accessed 28 Nov 2024.

18. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess. 1998;2(3):i–iv (**1-88**).

19. Shang Z. Use of Delphi in health sciences research: a narrative review. Medicine (Baltimore). 2023;102(7): e32829.

20. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol. 2014;67(4):401–9.

21. Niederberger M, Spranger J. Delphi technique in health sciences: a map. Front Public Health. 2020;8:457.

22. Franc JM, Hung KKC, Pirisi A, Weinstein ES. Analysis of Delphi study 7-point linear scale data by parametric methods: use of the mean and standard deviation. Methodol Innov. 2023;16(2):226–33.

23. Nisbet R, Elder JF, Miner G. Handbook of statistical analysis and data mining applications. 1st ed. Amsterdam: Academic Press/Elsevier; 2009.

24. Mohr DL, Wilson WJ, Freund RJ. Statistical methods. 4th ed. London: Academic Press; 2021.

25. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003–2009. Med Educ. 2010;44(1):50–63.

26. Herrera-Aliaga E, Estrada LD. Trends and innovations of simulation for twenty first century medical education. Front Public Health. 2022;10: 619769.

27. Elendu C, Amaechi DC, Okatta AU, Amaechi EC, Elendu TC, Ezeh CP, et al. The impact of simulation-based training in medical education: a review. Medicine (Baltimore). 2024;103(27): e38813.

28. Kennedy CC, Cannon EK, Warner DO, Cook DA. Advanced airway management simulation training in medical education: a systematic review and meta-analysis. Crit Care Med. 2014;42(1):169–78.

29. Lee R, Raison N, Lau WY, Aydin A, Dasgupta P, Ahmed K, et al. A systematic review of simulation-based training tools for technical and non-technical skills in ophthalmology. Eye (Lond). 2020;34(10):1737–59.

30. Wang K, Williams SR, Chong I, Fredrick DR. Needs assessment of emergency department residents in examining and managing patients with ophthalmologic complaints. J Acad Ophthalmol. 2019;11:e50–6.

31. Cohen ER, Feinglass J, Barsuk JH, Barnard C, O'Donnell A, McGaghie WC, et al. Cost savings from reduced catheter-related bloodstream infection after simulation-based education for residents in a medical intensive care unit. Simul Healthc. 2010;5(2):98–102.

32. Williams SA, Fitzpatrick K, Chandler NM, Arnold JL, Snyder CW. Financial and safety impact of simulation-based clinical systems testing on pediatric trauma center transitions. Pediatric Qual Saf. 2022;7(5): e578.

33. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: How to decide its appropriateness. World J Methodol. 2021;11(4):116–29.

34. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000;32(4):1008–15.

35. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. Int J Nurs Stud. 2001;38(2):195–200.

36. Keeney S. The Delphi technique in nursing and health research. Chichester: Wiley-Blackwell; 2011.

37. Mao X, Loke AY, Hu X. Developing a tool for measuring the disaster resilience of healthcare rescuers: a modified Delphi study. Scand J Trauma Resusc Emerg Med. 2020;28(1):4.

38. van der Horst N, Backx F, Goedhart EA, Huisstede BM. Return to play after hamstring injuries in football (soccer): a worldwide Delphi procedure regarding definition, medical criteria and decision-making. Br J Sports Med. 2017;51(22):1583–91.

39. Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. JAMA. 2012;307(17):1805–6.

40. Aggarwal CC. Data mining: the textbook. 2015th ed. Cham: Springer Nature; 2015.

41. Rosner B, Grove D. Use of the Mann-Whitney U-test for clustered data. Stat Med. 1999;18(11):1387–400.

42. Kang H. Sample size determination and power analysis using the G*Power software. J Educ Eval Health Prof. 2021;18:17.

43. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. 3rd ed. Boca Raton: Chapman & Hall/CRC; 2003. p. 1–1193.