AMERICAN ACADEMY™
OF OPHTHALMOLOGY

Operating Room Performance Improves after Proficiency-Based Virtual Reality Cataract Surgery Training

Ann Sofia Skou Thomsen, MD,[1,2] Daniella Bach-Holm, PhD, MD,[1] Hadi Kjærbo, MD,[1] Klavs Højgaard-Olsen, MD,[1] Yousif Subhi, MD,[2] George M. Saleh, FRCSEd, FRCOphth,[3] Yoon Soo Park, PhD,[4] Morten la Cour, DMSc, MD,[1] Lars Konge, PhD, MD[2]

Purpose: To investigate the effect of virtual reality proficiency-based training on actual cataract surgery performance. The secondary purpose of the study was to define which surgeons benefit from virtual reality training.

Design: Multicenter masked clinical trial.

Participants: Eighteen cataract surgeons with different levels of experience.

Methods: Cataract surgical training on a virtual reality simulator (EyeSi) until a proficiency-based test was passed.

Main Outcome Measures: Technical performance in the operating room (OR) assessed by 3 independent, masked raters using a previously validated task-specific assessment tool for cataract surgery (Objective Structured Assessment of Cataract Surgical Skill). Three surgeries before and 3 surgeries after the virtual reality training were video-recorded, anonymized, and presented to the raters in random order.

Results: Novices (non−independently operating surgeons) and surgeons having performed fewer than 75 independent cataract surgeries showed significant improvements in the OR—32% and 38%, respectively—after virtual reality training ($P = 0.008$ and $P = 0.018$). More experienced cataract surgeons did not benefit from simulator training. The reliability of the assessments was high with a generalizability coefficient of 0.92 and 0.86 before and after the virtual reality training, respectively.

Conclusions: Clinically relevant cataract surgical skills can be improved by proficiency-based training on a virtual reality simulator. Novices as well as surgeons with an intermediate level of experience showed improvement in OR performance score. *Ophthalmology 2016;■:1–8 © 2016 by the American Academy of Ophthalmology*

Complication rates in operations are affected by the experience and surgical skills of the surgeon.[1–3] Ideally, simulation-based training of surgical skills improves performance in the operating room and thereby diminishes the complication rate related to inexperience.[4,5] Yet, the effect of simulation-based training on operating room performance has never been investigated prospectively for the entire cataract surgical procedure.

By using proficiency-based training, learners train to a predefined, evidence-based benchmark (i.e., proficiency level) measured by valid performance metrics. This approach has proven to be one of the most effective ways to train technical skills[6] and is continually implemented in ophthalmology training programs.[7] In contrast to repetition- and time-based training, proficiency-based training ensures that only surgeons who meet the defined benchmark progress in the training program, and eventually operate on patients.[4]

Different training models exist for the training of cataract surgical skills.[8] One of the advantages of using virtual reality simulators is that performance metrics are embedded in the software, enabling continuous performance feedback and allowing feasible implementation of proficiency-based training.[9] The EyeSi simulator (VRmagic, Mannheim, Germany), is the most commonly used virtual reality simulator in ophthalmic surgery, including cataract surgery, and its performance metrics have previously been investigated and an evidence-based proficiency level has been established.[10]

Nevertheless, our knowledge on transfer of skills from a simulated environment to the operating room is still limited.[8] Previous retrospective studies and 1 case series with 3 trainees have shown an effect of the implementation of standardized cataract surgical training programs, including virtual reality training, on complication rates or time to complete surgery.[11–15] However, the retrospective design—not controlling for other variables that may have influenced trainees' performance—and divergent results call for prospective studies investigating the effect of virtual reality training on operating room performance.[16] Whereas the focus has mainly been on surgical trainees at the beginning of their learning curve, the question of which training level is appropriate to benefit from the training remains unanswered.[17]

http://dx.doi.org/10.1016/j.ophtha.2016.11.015
ISSN 0161-6420/16

The aim of this study was to investigate the effect of proficiency-based virtual reality training on cataract surgical skill in the operating room for surgeons with different levels of experience.

Methods

This study was conducted as a multicenter clinical trial with masking of both raters and outcome assessors. The Ethics Committee of the Capital Region of Denmark ruled that approval was not required for this study (protocol no. H-6-2014-011). The study adheres to the tenets of the Declaration of Helsinki and is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for simulation research.[18]

Participants

From April 1, 2014, to March 11, 2015, 19 cataract surgeons from Denmark were included in the study. We intended to include cataract surgeons representing all experience levels, including surgeons not yet operating independently (i.e., performing only single steps of the cataract surgery). Because of a limited cohort of ophthalmologists undergoing cataract surgical training in Denmark, we invited all surgeons having performed fewer than 1000 surgeries to participate in the study. Eligible participants were identified by contacting chief consultants and/or cataract surgical program directors at all ophthalmology departments in Denmark. From a cohort of 22, 2 did not respond to the invitation and 4 did not find time to participate in the study. Furthermore, 3 expert surgeons (defined as having performed >1000 surgeries) were included in the study. In total, 19 cataract surgeons, employed at 9 different ophthalmology departments and 2 private clinics in Denmark, were enrolled into the study. We divided the surgeons into 4 groups according to experience level: novices, the group of surgeons who did not yet operate independently but performed only steps of the cataract surgical procedure (0 independent operations at time of enrollment in the study); intermediates, defined as surgeons on the steeper part of the learning curve[19] (1−75 independent operations); experienced surgeons (76−999 operations); and expert surgeons (≥1000 operations). All participants gave oral and written consent before inclusion and completed a questionnaire regarding demographic data and surgical experience.

Simulator Training

The virtual reality simulator training was carried out at the Simulation Centre at Rigshospitalet (Copenhagen Academy of Medical Education and Simulation). The cataract (phacoemulsification) interface on the VRmagic EyeSi simulator, version 2.8.10, was used for the study. All participants were given a 10-minute introduction to the simulator. A previously established performance test with evidence of validity was used for the proficiency-based training[10]: All participants trained on the simulator until they achieved a predefined pass/fail score of 600 points (of a maximum of 700 points) in 2 consecutive sessions. This pass/fail level was based on previous study results[10] and evidence indicating that deliberate "overtraining" leads to enhanced skill retention.[6] During training, 1 author (A.S.S.T.) gave instructions to all participants. Table 1 shows details on the proficiency-based training, including settings on the simulator.

Surgical Procedure

The participants performed 3 consecutive phacoemulsification surgeries immediately before and after the training intervention (Fig 1). They were only allowed to operate on uncomplicated cataract cases, defined as follows: (1) being performed under local anesthesia, (2) patient >60 years of age, (3) preoperative best-corrected visual acuity >1/60 (measured using a standard Snellen chart at 6 meters' distance). Age and visual acuity of the patients were noted by the study participants and disclosed to the primary investigator. Furthermore, the timing of the operations was cross-checked to ensure that the surgeons did not select specific operations based on their own preferences to be included in the study. The novices informed the primary investigator about which surgical steps had been performed by their supervisor. All phacoemulsification techniques (including divide-and-conquer and phaco-chop techniques) were accepted. The participants were not allowed to operate on patients while they underwent training on the simulator. Exclusion criteria were as follows: (1) more than 2 weeks between operations and training intervention, and (2) inability to provide the 6 video recordings of performed operations.

Data Anonymization and Masking

The surgeries were video-recorded and thereafter anonymized regarding the identity of both the patient and the surgeon. This was done by cropping the recordings before and after performance of the actual procedure in addition to removing logos, person identifiable data, and sound using Final Cut Pro video editing software version 7 (Apple, Inc, Cupertino, CA). The videos were presented to 3 masked cataract surgeons in a random order through a secured web-based video-rating software.[20] The outcome assessors were also masked to the identity of the surgeons until data were collected and saved in a database.

Outcome Measures

The primary outcome measure was technical performance, measured by the Objective Structured Assessment of Cataract Surgical Skill (OSACSS) rating scale.[21] The rating scale consists of task-specific items and global indices, which are rated from 1 point ("inadequately performed") to 5 points ("well performed"). The first item, concerning draping, was omitted because the surgical assistant usually performs this step of the procedure in Denmark. Global indices were rated but not included in the final assessment score to make comparison between non−independently operating surgeons and independently operating surgeons possible. Thus, the assessment of technical performance included 13 task-specific items, which were rated using the original 5-point rating scale (Fig 2). After recoding the scores from 0 to 4, the final assessment scale ranged from 0 to 52 points, with 52 points representing superior performance. Three raters evaluated all videos independently. Before the initiation of the study, raters were trained to ensure a standardized assessment and to avoid rater errors. Specifically for the novices, steps performed by their supervisor were adjusted to the lowest score ("inadequately performed") post hoc by the primary investigator.

Reliability

Generalizability theory, a statistical method developed by Cronbach et al, was used to analyze the reliability of the OSACSS scores[22]. We used a fully crossed design for every factor in the assessment, meaning that all surgeons performed 3 procedures before and after virtual reality training, and all 3 raters evaluated all procedures. This study design made it possible for us to investigate different sources of bias, including observer bias. Another possible bias would be the influence of the novice surgeons' natural learning curve on the estimated size effect, that is, the performance improvement from case to case leading to an improvement in performance score not caused by the virtual reality training itself. To attribute an observed effect to the

Table 1. Proficiency-Based Training Program on the EyeSi Simulator

Module No.	Task Name	Task Description	Test Level	Level	Points
1	Intracapsular navigation	Aiming at objects within the capsule with the tip of instrument (abstract task)	3	2	0–100
2	Antitremor training	Following a circular path on the capsule with the tip of instrument (abstract task)	7	4	0–100
3	Intracapsular antitremor training	Following a circular path within the capsule with the tip of instrument (abstract task)	5	2	0–100
4	Forceps training	Collecting objects in the anterior chamber with the forceps (abstract task)	4	4	0–100
5	Bimanual training	Aiming at objects simultaneously with 2 instruments (abstract task)	5	5	0–100
6	Capsulorrhexis	Performing a continuous curvilinear capsulorrhexis (procedural task)	3	1*	0–100
7	Phaco divide and conquer	Performing phacoemulsification on a medium-hard lens (procedural task)	8	5	0–100
		Total score in two consecutive sessions			>600

*Capsulorrhexis: Weak zonula. No initial tear.

actual intervention, we expected performance score improvement between the 3 pretraining procedures and between the 3 posttraining procedures to be negligible, as demonstrated in Figure 1. A generalizability (reliability) coefficient was calculated as a measure of the accuracy of the assessments, and a value greater than 0.8 was considered an acceptable level.[23] Furthermore, we analyzed the dependency of rater and procedure quantity on the generalizability coefficient. G String IV statistical software package version 6.3.8 (Papaworx, Hamilton, Canada) was used for the analysis.

Statistical Analysis

Stata version 14 (StataCorp, College Station, TX) was used for statistical analysis: Mixed-effects regression was used to examine differences in pre-post simulation effects, controlling for items and rater effects.[24,25] This method allows comparison of pre-post effects while taking into account the nested structure of the data, where each participant was assessed on 3 items by 3 raters each.

Results

Of the cohort of 19 enrolled cataract surgeons, 18 were included in the final data analysis; 1 experienced surgeon was excluded because only 2 operations were video-recorded after training. There was a mean of 5 days between the pretraining operations and the intervention and 7 days between the intervention and post-training operations. One pretraining operation for 2 surgeons was not video-recorded owing to technical problems, 1 patient was excluded owing to young age, and 1 complicated case was excluded (white cataract). Instead, the subsequent operation was video-recorded and included in all 4 cases.

The generalizability (reliability) coefficient for the performance assessments was 0.92 and 0.86 for pretraining and posttraining, respectively. Table 2 shows the results of the reliability analysis, specifying the different sources of variance in the study. Our prognostic analysis indicates that including only 2 raters or assessing only 2 procedures per surgeon would result in a reliability coefficient >0.8.

Table 3 shows the descriptive statistics, stratified by group and by pre-post performances, where means and standard deviations were calculated by averaging scores across raters and items. Table 4 shows the pre-post differences in simulation performance, as estimated using regression coefficients from the mixed-effects model.

Both novices and intermediates had significant improvements in performance, by 5.0 and 9.8 points, corresponding to an

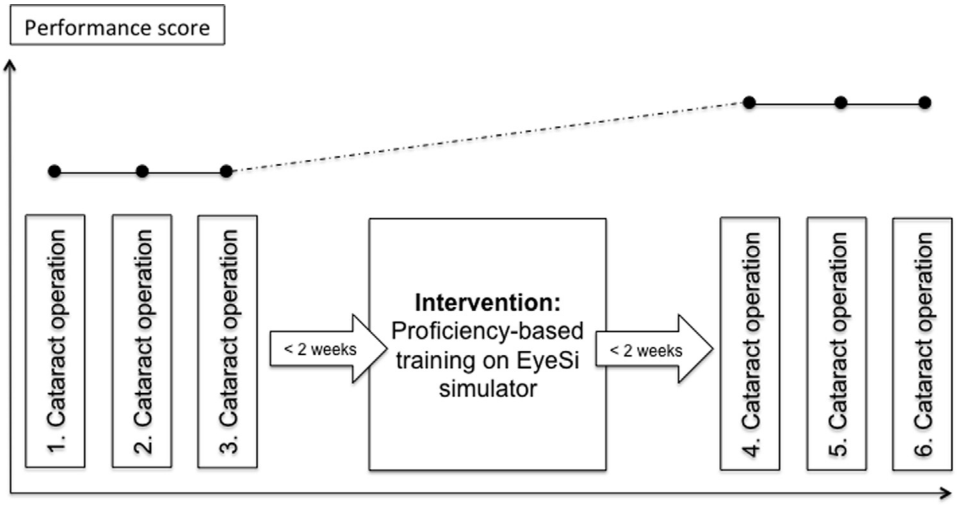

Figure 1. Study design. To investigate possible bias owing to case-to-case learning, the performance scores during the 3 pretraining operations and during the 3 posttraining operations were compared for all participants; a variance less than 15% of the total variance was accepted. The dots represent the expected mean performance score for all participants (study hypothesis).

	Poorly or inadequately performed		Performed with some errors or hesitation		Performed well with no prompting or hesitation	Tick if not performed
Incision and Paracentesis: Formation and Technique	1	2	3	4	5	
Viscoelastic: Appropriate Use and Safe Insertion	1	2	3	4	5	
Capsulorrhexis: Commencement of Flap	1	2	3	4	5	
Capsulorrhexis: Formation and Circular Completion	1	2	3	4	5	
Hydrodissection: Visible Fluid Wave and Free Nuclear Rotation	1	2	3	4	5	
Phacoemulsification Probe and Second Instrument: Insertion Into Eye	1	2	3	4	5	
Phacoemulsification Probe and Second Instrument: Effective Use and Stability Within the Eye	1	2	3	4	5	
Nucleus: Sculpting or Primary Chop	1	2	3	4	5	
Nucleus: Rotation and Manipulation	1	2	3	4	5	
Nucleus: Cracking or Chopping With Safe Phacoemulsification of Segments	1	2	3	4	5	
Irrigation and Aspiration Technique With Adequate Removal of Cortex	1	2	3	4	5	
Lens Insertion, Rotation, and Final Position of Intraocular Lens	1	2	3	4	5	
Wound Closure (Including Suturing, Hydration, and Checking Security as Required)	1	2	3	4	5	

Figure 2. Modified Objective Structured Assessment of Cataract Surgical Skill (OSACSS) rating scale. The OSACSS tool was developed for standardized assessment of technical skills in cataract surgery. The modified scale consists of 13 task-specific items.

improvement of 32% and 38%, respectively. There were no significant improvements for experienced and expert groups.

Figure 3 shows pre-post differences in operating room performance by experience group. The relationship between experience level (only for the independent surgeons) and OSACSS score is depicted in Figure 4.

Discussion

The results of this prospective clinical trial show a significant effect on cataract surgeons' performance on patients after attending a proficiency-based training program on a virtual reality simulator. More specifically, both surgeons-in-training and surgeons who have performed fewer than 75 independent operations achieved a statistically significant effect following the training.

We found an improvement in the OSACSS scores of 32% and 38%, respectively, for the novice group and the intermediates, compared with baseline values. Notably, there were no statistical differences in effect size between the 2 groups (Fig 4). Taking into consideration the standard deviations, the possible size range varied from 3.1 points (novices) to 13.9 points (intermediates), corresponding to an improvement between a 20% and 54% increase compared with baseline values. The mean value of the novices' performance score approached the score of the intermediates; a similar trend was noticed for the intermediates as compared to the experienced surgeons' score, indicating that the training effect is

Table 2. Reliability Analysis: Using G-Theory to Estimate Relevant Sources of Variance's Contribution to the Total Variance in Objective Structured Assessment of Cataract Surgical Skill Score

Source of Variance	*Description*	Proportion of Total Variance *(Before training)*	Proportion of Total Variance *(After training)*	*Interpretation*
Surgeon	Systematic variance among surgeons	76%	64%	Most of the variance is attributable to different competence levels between the surgeons
Procedure	Systematic variance among procedures	0%	0%	The procedures had the same difficulty level
Rater	Systematic variance among raters	9%	14%	Medium level of consistency between the raters' degree of stringency
Interaction between surgeon and procedure	Systematic trend for a surgeon to perform differently from procedure to procedure	5%	12%	The individual surgeons' performance was consistent between procedures
Interaction between surgeon and rater	Systematic trend for a rater to assess a particular surgeon differently	4%	4%	The anonymization of data and masking of raters were effective (no assessment bias)
Interaction between surgeon, procedure, and rater	Remaining error variance	6%	6%	Very small proportion of unexplained error
Total variance		100%	100%	

clinically relevant. Furthermore, other research findings point toward virtual reality training being *more* efficient than traditional apprenticeship training at the beginning of the learning curve: McCannel et al[15] reported a significant decrease in the rate of errant capsulorrhexes after implementation of a proficiency-based capsulorrhexis training curriculum on the EyeSi simulator. This may be explained by more efficient hands-on training time in a simulator environment as compared to clinical training, where training opportunities also depend on the available patients.[26] However, because of the retrospective study design and lack of prospective studies—presumably owing to ethical concerns—ascertaining the specific cause-effect relationships is rather speculative. Nevertheless, virtual reality training should not stand alone as a training modality in the cataract surgical curriculum but rather should be supplemented by other effective training tools.[8]

Our study results demonstrate that surgeons who already started operating on patients (<75 operations) received a significant beneficial effect from virtual reality training. The 75-operation threshold was chosen as a result of our particular grouping of surgeons based on previous study findings, which demonstrated a significant reduction in complication rates after the first 75 cases.[19] Looking at the relationship between experience level and training effect size in our cohort (as depicted in Fig 4), we see a decrease in the size of the effect (illustrated by the distance between lines) at around 100 operations. A significant training effect may also be found for even more experienced surgeons if a more advanced evidence-based training program was established.

On the basis of these study findings, we propose a cataract surgical training curriculum consisting of 3 parts: a *pre-patient training program* (steps 1−3), *supervised practice on patients* (step 4), and a *follow-up training program* for surgeons operating independently (steps 5 and 6) (Fig 5). Evident from previous publications on the field, the adoption of proficiency-based training—including objective assessment of surgical skills—is challenging.[7] In Europe, most countries still use only the apprenticeship model for cataract surgical training.[27] The findings of this study underline the benefits of using a proficiency-based approach based on automated, objective assessments of surgical skills using virtual reality technology. We believe that integrating our 6-step approach to cataract surgical training will benefit future surgeons and patients, as well as educators in ophthalmology.

Previous retrospective studies and 1 case-series study have evaluated the effect of implementation of structured training programs—including virtual reality training—on patient-related outcomes, but the reported findings are rather divergent[8]: 2 studies did not find a reduction in complication rates.[12,13] Possible reasons for inconclusive and conflicting results are the absence of valid outcome measures, type II errors owing to small sample sizes, and failure to adapt efficient instructional approaches.[9] There has been a move toward proficiency-based training,[7] but other instructional approaches are also becoming increasingly accessible as additional evidence-based tools and more efficient training of surgeons becomes available.

The generalizability (reliability) coefficient of the performance assessments in our study was very high, reflecting that between 86% and 92% of the OSACSS score could be ascribed the true performance score of the surgeons.[28] This generalizability coefficient is specific for this study design: a fully crossed assessment design with 3 independent, masked raters and 3 different procedures (pretraining and posttraining) using a modified OSACSS rating scale. However, our prognostic analysis indicates that including only 2 raters or 2 procedures would also result in an acceptable generalizability coefficient of >0.8, which is the recommended minimum level of reliability for high-stakes assessments.[28] Translating this to other clinical settings, to acquire a reliable assessment of surgical

Table 3. Descriptive Statistics: Performance Before and After Intervention

Group (No. of Operations Performed)	No. of Surgeons	Before		After	
		Mean	*SD*	*Mean*	*SD*
Novice (0)	4	15.33	7.03	20.31	12.10
Intermediate (1–75)	4	25.81	9.66	35.58	9.22
Experienced (76–999)	7	42.97	3.08	42.60	3.01
Expert (≥1000)	3	48.26	2.42	47.78	4.64
Overall	18	33.90	14.00	36.95	12.08

Descriptive statistics were derived by taking the means across 3 items and across 3 raters.

trainees—using the modified OSACSS scale—either 2 observers rating 3 procedures or 3 observers rating 2 procedures are needed. Assessment bias (or observer bias) is a major threat to the validity of study results when using human-based scoring because of a possible influence of the raters' predispositions, including the rater–trainee relationship.[29] The OSACSS rating tool was developed for evaluation of video-recorded surgeries, enabling masking of raters, thus minimizing the risk of assessment bias compared with direct observation. The results of our generalizability analysis show that masking was successful, with only 4% of the total variance caused by the interaction between surgeon and rater.

One major limitation to our study is the lack of a control group: owing to a limited sample size it was not possible to conduct a randomized trial. Also, some would argue that it would be unethical to expose 1 group of surgeons to virtual reality training and the other to no training. It would have been optimal to randomize 1 group of surgeons to virtual reality training and another group to wet-lab training, or 1 group to proficiency-based training and another to a different type of educational intervention.[30] However, such a study design would not directly answer the intended research question: Does proficiency-based training of cataract surgeons have an effect on performance in the operating room, and if so, is a possible effect limited to only novice surgeons? Importantly, the lack of a control group leads to a risk of confounding and bias to the results. To address this risk, we have applied generalizability theory to the *before-and-after* study design.

Table 4. Pre-Post Differences: Mixed-Effects Regression Estimates

Group (No. of Operations Performed)	Difference	*P* Value
Novice (0)	4.97 (1.87)	0.008
Intermediate (1–75)	9.78 (4.12)	0.018
Experienced (76–999)	−0.37 (1.28)	0.775
Expert (≥1000)	−0.48 (1.18)	0.682
Overall	3.06 (1.15)	0.008

Estimates are mixed-effects regression coefficients. Estimates represent pre-post differences, controlling for participants, items, and raters.

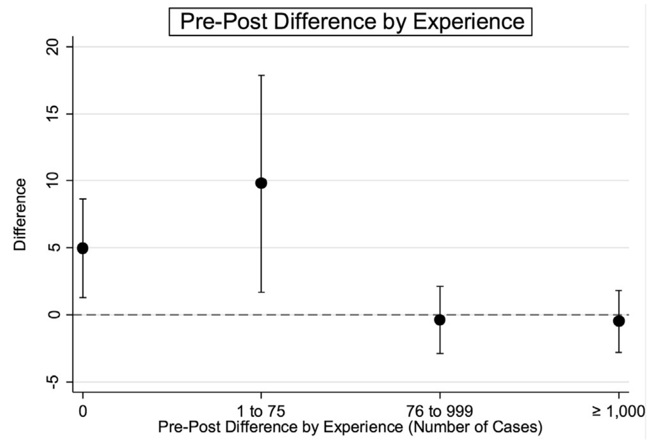

Figure 3. Pre-post differences by experience: point estimates and 95% confidence intervals. Estimates are coefficients based on mixed-effects regression, with random intercepts on participant, item, and raters.

When designing this study, we were concerned about the impact of the surgeons' natural learning curve confounding a possible training effect. However, as is evident from our reliability analysis, an individual surgeon's performance was consistent between procedures when comparing the 3 pretraining operations or the 3 posttraining operations (as depicted in Fig 1), with a variation of <15%. If the effect of a natural learning curve had been pronounced, we would expect a significant source of variance between each surgeon and corresponding procedures, either before or after the intervention. A variation in the difficulty level of the included cases is another possible source of bias—although we defined 3 criteria (performed with the patient under local anesthesia, patient >60 years of age,

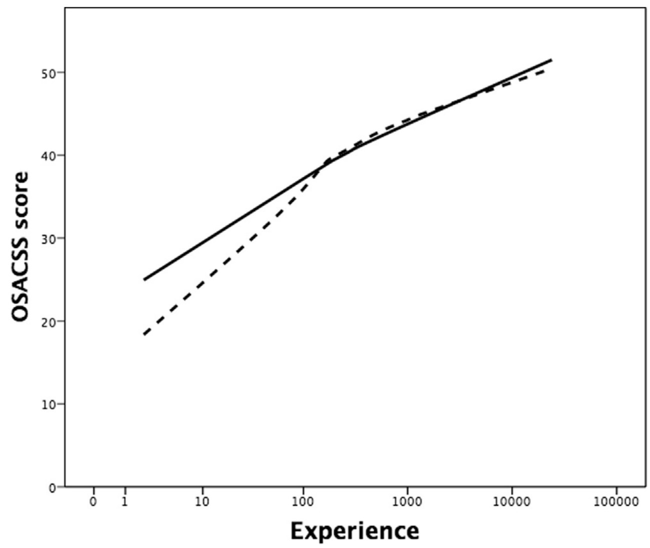

Figure 4. Relationship between experience level and Objective Structured Assessment of Cataract Surgical Skill (OSACSS) score. Illustration shows relationship between experience levels of the independent surgeons (number of operations performed) and operating room performance before (*dashed line*) and after (*solid line*) virtual reality training.

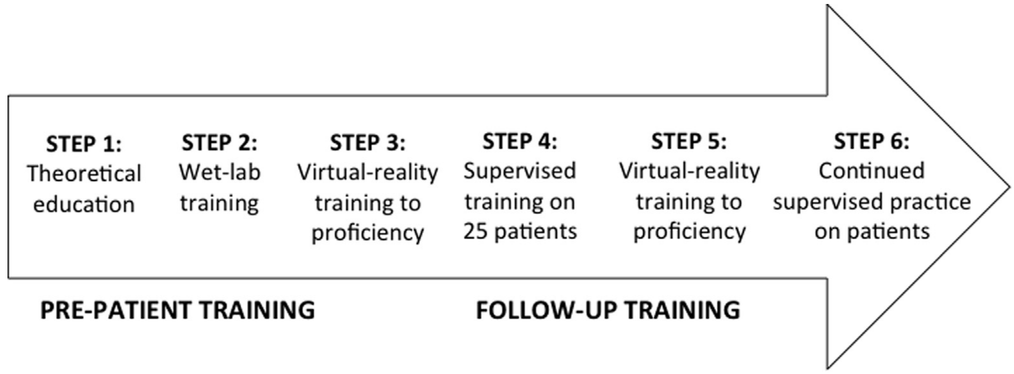

Figure 5. Mastery learning in cataract surgery: Suggestion for an evidence-based training program with follow-up training.

preoperative best-corrected visual acuity >1/60) to standardize the included cases. An observed effect may be caused by participants performing surgery on less complicated cases after attending the virtual reality training, and not the training itself. Yet, the results from the reliability analysis suggest that no systematic variance was introduced owing to the different procedures (either pretraining or posttraining). Furthermore, a commonly used approach in surgical training programs would be to allocate more complicated cases to the surgical trainee as the training progresses, and not the opposite. However, we cannot rule out that there may have been a difference in the difficulty levels of cases from pretraining to posttraining, as this was not specifically investigated.

The most important limitation of our study was the relatively small sample size. If more participants were included in the study, it would have been possible for us to conduct a multiple regression analysis including experience level and patient data as variables. However, our study was adequately powered to detect a possible training effect among cataract surgeons of different experience levels.

We used performance assessment as a parameter for technical skills to investigate the effect of virtual reality training. Future research may focus on studying the relationship between technical skills and patient-related outcomes to investigate the direct impact of various training approaches on patient outcomes.

The results of this study suggest that proficiency-based virtual reality training can improve the surgical performance in the theatre of not only novice surgeons but also surgeons at a more intermediate level. The conclusion that may be drawn is that technical skills improvement is transferable from a simulated setting to the operating room.

References

1. Lundström M, Barry P, Henry Y, et al. Evidence-based guidelines for cataract surgery: Guidelines based on data in the European Registry of Quality Outcomes for Cataract and Refractive Surgery database. *J Cataract Refract Surg*. 2012;38(6):1086-1093.
2. Johnston RL, Taylor H, Smith R, Sparrow JM. The Cataract National Dataset electronic multi-centre audit of 55,567 operations: variation in posterior capsule rupture rates between surgeons. *Eye (Lond)*. 2010;24(5):888-893.
3. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369(15):1434-1442.
4. Oetting TA. Surgical competency in residents. *Curr Opin Ophthalmol*. 2009;20(1):56-60.
5. McCannel CA. Simulation surgical teaching in ophthalmology. *Ophthalmology*. 2015;122(12):2371-2372.
6. Zevin B, Levy JS, Satava RM, Grantcharov TP. A consensus-based framework for design, validation, and implementation of simulation-based training curricula in surgery. *J Am Coll Surg*. 2012;215(4):580-586.
7. Lee AG, Oetting T, Beaver HA, Carter K. The ACGME Outcome Project in ophthalmology: practical recommendations for overcoming the barriers to local implementation of the national mandate. *Surv Ophthalmol*. 2009;54(4):507-517.
8. Thomsen ASS, Subhi Y, Kiilgaard JF, et al. Update on simulation-based surgical training and assessment in ophthalmology. *Ophthalmology*. 2015;122(6):1111-1130.
9. Nicholls D, Sweet L, Muller A, Hyett J. Teaching psychomotor skills in the twenty-first century: revisiting and reviewing instructional approaches through the lens of contemporary literature. *Med Teach*. 2016;38(10):1056-1063.
10. Thomsen ASS, Kiilgaard JF, Kjaerbo H, et al. Simulation-based certification for cataract surgery. *Acta Ophthalmol*. 2015;93(5):416-421.
11. Rogers GM, Oetting TA, Lee AG, et al. Impact of a structured surgical curriculum on ophthalmic resident cataract surgery complication rates. *J Cataract Refract Surg*. 2009;35(11):1956-1960.
12. Belyea DA, Brown SE, Rajjoub LZ. Influence of surgery simulator training on ophthalmology resident phacoemulsification performance. *J Cataract Refract Surg*. 2011;37(10):1756-1761.
13. Pokroy R, Du E, Alzaga A, et al. Impact of simulator training on resident cataract surgery. *Graefes Arch Clin Exp Ophthalmol*. 2013;251(3):777-781.
14. Baxter JM, Lee R, Sharp JA, Foss AJ, Intensive Cataract Training Study Group. Intensive cataract training: a novel approach. *Eye (Lond)*. 2013;27(6):742-746.
15. McCannel CA, Reed DC, Goldman DR. Ophthalmic surgery simulator training improves resident performance of capsulorhexis in the operating room. *Ophthalmology*. 2013;120(12):2456-2461.

16. Cox T, Seymour N, Stefanidis D. Moving the needle: simulation's impact on patient outcomes. *Surg Clin North Am*. 2015;95(4):827-838.
17. Gillan SN, Saleh GM. Ophthalmic surgical simulation. A new era. *JAMA Ophthalmol*. 2013;131(12):1623.
18. Cheng A, Kessler D, Mackinnon R, et al. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Adv Simul*. 2016;1(12):238-248.
19. Taravella MJ, Davidson R, Erlanger M, et al. Characterizing the learning curve in phacoemulsification. *J Cataract Refract Surg*. 2011;37(6):1069-1075.
20. Subhi Y, Todsen T, Konge L. An integrable, web-based solution for easy assessment of video-recorded performances. *Adv Med Educ Pract*. 2014;5:103-105.
21. Saleh GM, Gauba V, Mitra A, et al. Objective structured assessment of cataract surgical skill. *Arch Ophthalmol*. 2007;125(3):363-366.
22. Bilgic E, Watanabe Y, McKendy KM, et al. Reliable assessment of performance in surgery: a practical approach to generalizability theory. *J Surg Educ*. 2015;72(5):774-775.
23. Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ*. 2004;38(9):1006-1012.
24. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman; 2004.
25. Wiley HE, Thompson DJS, Bailey C, et al. A crossover design for comparative efficacy: a 36-week randomized trial of bevacizumab and ranibizumab for diabetic macular edema. *Ophthalmology*. 2016;123(4):841-849.
26. Konge L, Clementsen PF, Ringsted C, et al. Simulator training for endobronchial ultrasound: a randomised controlled trial. *Eur Respir J*. 2015;46(4):1140-1149.
27. Muttuvelu DV, Andersen CU. Cataract surgery education in member countries of the European Board of Ophthalmology. *Can J Ophthalmol*. 2016;51(3):207-211.
28. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide. *Med Teach*. 2012;34(11):960-992.
29. Hróbjartsson A, Thomsen ASS, Emanuelsson F, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*. 2012;344:e1119.
30. Tolsgaard MG, Ringsted C. Using equivalence designs to improve methodological rigor in medical education trials. *Med Educ*. 2014;48(2):220-221.

Footnotes and Financial Disclosures

Originally received: September 6, 2016.
Final revision: November 3, 2016.
Accepted: November 11, 2016.
Available online: ■■■. Manuscript no. 2016-265.

[1] Department of Ophthalmology, Rigshospitalet, Glostrup, Denmark.

[2] Copenhagen Academy for Medical Education and Simulation, Centre for HR, Copenhagen, Capital Region of Denmark, Denmark.

[3] National Institute for Health Research, Biomedical Research Centre at Moorfields Eye Hospital and the UCL Institute of Ophthalmology, Moorfields Eye Hospital, London, United Kingdom.

[4] Department of Medical Education, College of Medicine, University of Illinois, Chicago, Illinois.

Financial Disclosure(s):

The author(s) have made the following disclosure(s):

Funded by Fight for Sight Denmark and the Synoptik Foundation.

G.S.: Support − the Department of Health through an award made by the National Institute for Health Research Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust, the UCL Institute of Ophthalmology, and the Special Trustees of Moorfields.

The funding organizations had no role in the design or conduct of this research.

Author Contributions:

Conception and design: Thomsen, Højgaard-Olsen, Saleh, la Cour, Konge

Analysis and interpretation: Thomsen, Park, la Cour, Konge

Data collection: Thomsen, Bach-Holm, Kjærbo, Højgaard-Olsen, Subhi

Obtained funding: Not applicable

Overall responsibility: Thomsen, Bach-Holm, Kjærbo, Højgaard-Olsen, Subhi, Saleh, Park, la Cour, Konge

Abbreviations and Acronyms:
G-theory = Generalizability theory; **OR** = Operating Room; **OSACSS** = Objective Structured Assessment of Cataract Surgical Skill.

Correspondence:
Ann Sofia Skou Thomsen, MD, Department of Ophthalmology, Rigshospitalet, Ndr. Ringvej 57, DK-2600 Glostrup, Denmark. E-mail: skouthomsen@dadlnet.dk.