Virtual reality simulation training of mastoidectomy – studies on novice performance

Steven Arild Wuyts Andersen

This review has been accepted as a thesis together with five previously published papers by University of Copenhagen on 1 June 2016 and defended on 24 June 2016.

Tutor(s): Mads Sølvsten Sørensen, Per Cayé-Thomasen and Lars Konge.

Official opponents: Gregory J. Wiet, Anne Mette Mørcke and Torben Schroeder.

Correspondence: Department of Otorhinolaryngology—Head & Neck Surgery, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark.

E-mail: stevenarild@gmail.com

Dan Med J 2016;63(8):B5277

THE FIVE ORIGINAL PAPERS ARE:

I. Andersen SA, Cayé-Thomasen P, Sørensen MS. Mastoidectomy performance assessment of virtual simulation training using final-product analysis. Laryngoscope 2015;125:431–435.
II. Andersen SA, Foghsgaard S, Konge L, Cayé-Thomasen P, Sørensen MS. The effect of self-directed virtual reality simulation on dissection training performance in mastoidectomy. Laryngoscope 2015; Oct 9 [Epub ahead of print].
III. Andersen SA, Konge L, Cayé-Thomasen P, Sørensen MS. Learning curves of virtual mastoidectomy in distributed and massed practice. JAMA Otolaryngol Head Neck Surg 2015;141:913–918.
IV. Andersen SA, Konge L, Cayé-Thomasen P, Sørensen MS. Retention of mastoidectomy skills after virtual reality simulation training. JAMA Otolaryngol Head Neck Surg. 2016; Apr 28 [Epub ahead of print].
V. Andersen SA, Konge L, Mikkelsen PT, Cayé-Thomasen P, Sørensen MS. Mapping the plateau of novices in virtual reality simulation training of mastoidectomy. Laryngoscope. 2016; Apr 14 [Epub ahead of print].

INTRODUCTION

High quality training of surgical skills and procedures is essential for excellent patient care and safe surgery and surgical education needs to evolve according to the changing landscape of surgery. Advances in computer technology have enabled virtual reality (VR) simulation of complex surgical procedures including mastoidectomy in temporal bone surgery. VR simulation appears to be an attractive training platform for the surgical novice, who needs to acquire basic competencies in a safe environment. Novel technology such as VR simulation offer new and exciting possibilities for learning, but should be used where it makes sense and then applied appropriately.

One of the advantages of VR simulation is the opportunity to practice advanced surgical procedures repeatedly, supporting progressive skills development and consolidation. VR simulation allows hands-on surgical training to occur outside the traditional learning environment of bedside teaching, operating room (OR) experience or specialist training courses, and independent of surgical tutors. Such independent learning calls for a strong educational design and organization, and this should be guided by evidence.

This thesis will investigate novices' performance in directed, self-regulated VR simulation of mastoidectomy to increase evidence on VR simulation training in temporal bone surgery.

BACKGROUND

CHALLENGES IN SURGICAL TRAINING

Surgical skills have traditionally been taught by the means of apprenticeship. In the cognitive apprenticeship model, the surgical teacher starts as a role model and a coach who gradually guides the trainee towards competency.[1] By scaffolding, the surgical trainee is provided with increasing challenges and subsequently achieves the capability for articulation of the procedural steps.[1] Lastly, expertise arises from reflection and the ability to explore and invent new strategies.[1]

Surgical skills are complex and encompass general skills (e.g. situational awareness, communicative skills), specific knowledge (e.g. relevant anatomy, procedural steps, recognition of critical errors) and psychomotor skills that all need to be integrated to become an apt surgeon.[1] This requires a systematic approach to surgical training and a well-planned surgical curriculum.

The apprenticeship model presently faces many challenges: minimally invasive surgery, robotic surgery, and other technological advances have changed the way we do surgery and consequently how we need to train surgeons; restrictions on working hours are limiting the available time for training contra service for both trainees and supervisors; and importantly, there are increased awareness of patient, quality and safety issues.

Many of these challenges apply also to otorhinolaryngology (ORL), resulting in increasing specialization, discontinuity in patient care, and an increasing duration of training as consequences.[2] Work hours, clinical duties, surgical experience, and supervision during ORL training varies greatly depending on country even within Europe,[3] and even though much effort has been made in improving safety and quality of patient care during the last 15 years, marked progress remains faraway.[4]

Clearly, high quality surgical training can no longer rely simply on apprenticeship and the surgical curriculum must be adjusted to these challenges. Simulation can be part of the solution but should be a "*part of a coherent strategy based on clear educational aims and must mirror actual practice*".[5]

SIMULATION-BASED SURGICAL SKILLS TRAINING

The surgical apprenticeship has traditionally been supplemented by a range of other teaching methods such as lectures and videotaped demonstrations, and hands-on training using dissection, animal models, and other simulation-based methods including inanimate objects such as mannequins and box trainers, and more recently, simulated patients, and VR simulation. Overall, the simulation-based teaching methods have a large and positive effect on knowledge, skills, and behaviors in addition to some effect on patient related outcomes.[6] Each training modality has specific benefits and limitations, and the context, objective of training, and learning situation should be considered when choosing the appropriate model. Importantly, it should be an integrated part of the surgical curriculum to achieve the largest effect on learning.[7]

For surgical technical and procedural skills, simulation-based methods can help the trainee gain technical proficiency, provide tailored feedback, situational learning, and address affective aspects of learning.[5] Similar to the optimal surgical apprenticeship, simulation-based training should allow for repeated and deliberate practice for progressive skills development and consolidation, provide access to tutors, map onto real-life clinical experience and provide a positive and motivational learning environment.[5,7] Other important features of simulation-based training leading to effective learning include providing a range of difficulty levels, multiple learning strategies, the ability to capture the clinical variation, a controlled and safe environment, and opportunity for individualized learning.[7] Still, there is a gap in understanding the mechanisms that lead to effective learning in all these areas.[8]

Although simulation-based training has fewer of the time and safety related constraints compared with the surgical apprenticeship, implementation into the surgical curriculum infrequently considers the individual and timely need of the trainee. Instead, simulation-based skills training is often organized as intensive courses, boot camps, and similar, isolated, single-instance training opportunities. From an educational point of view, this can result in simulation-based training being uncoupled from the trainees' everyday work and the transference of the acquired skills to clinical practice can therefore be a challenge.[5]

The development of VR simulators in the late 1990s has transformed simulation-based training of surgical technical skills, promoted by the development of advanced laparoscopic simulators and extensive research on their use. Today, VR surgical simulation training is supported by evidence as an effective training tool in many areas of surgery and VR simulation training is routinely provided in different surgical specialties.[9]

Simulation-based training in ORL includes a full spectrum of simulation models including 1) box models for task training such as suturing, anastomosis, ties and grommet insertion,[10] 2) mannequins for peritonsillar abscess drainage, endoscopy, nasal packaging and cricothyroidotomy,[9,11,12] 3) simulated patients,[13] and 4) VR simulators for bronchoscopy, endoscopic sinus surgery, myringotomy, temporal bone surgery and more.[12,14–17] In the United States, most ORL residency programs include some form of simulation-based training but the integration into the curriculum varies and more best-practice evidence is needed.[18]

TRAINING IN TEMPORAL BONE SURGERY

Temporal bone surgical training is considered important for all ORL trainees even with increasing specialization because it forms the basis for understanding temporal bone and middle ear disease, the surgical management, and the complex anatomy of the region. Temporal bone surgical skills require precise motor skills in handling the drill, suctioning and irrigation, microsurgical skills with the use of the operating microscope, and knowledge that includes a detailed three-dimensional understanding of the anatomy of the temporal bone.

Temporal bone surgical training is mainly based on apprenticeship with direct supervision in the OR by a senior otologist after having completed a temporal bone skills training course or training in a temporal bone laboratory.[19] Temporal bone courses often include lectures, video demonstrations, and temporal bone drilling on cadavers and/or simulated temporal bones. The drilling of human cadaveric temporal bones closely mimics real-life conditions including the variable pneumatization of the temporal bone. Cadaveric temporal bone dissection is considered the gold standard training method even though this is based more on tradition than on scientific evidence for its efficacy. Only recently have the effect of cadaveric temporal bone dissection training been investigated.[20,21]

Temporal bone training on cadaveric temporal bones is typically organized with either access to an open temporal bone lab[22] where the trainees can practice as needed on their own, or as formalized, intensive temporal bone courses. Only few but the largest centers worldwide can provide open access to temporal bone lab facilities due to high maintenance costs and limited availability of donated, human cadaveric temporal bones. Temporal bone courses are also costly and require dedicated instructors in addition to appropriate facilities and temporal bones. This leaves the trainee limited opportunity for repeated practice to support skills acquirement and consolidation, and to gradually hone and develop more advanced skills.

Due to ethical issues, human cadaveric temporal bones are a scarce resource and a low-fidelity 2D model is suboptimal.[23] Animal models are not suitable for temporal bone training due to species differences[24] but for a number of years, temporal bone models of plaster or plastic has offered an alternative.[25] Very recently, reports on temporal bone models made by 3D printing have been published, which could offer a range of anatomical variance equal to cadaveric temporal bones.[26,27] Nonetheless, educational evidence of plaster or plastic models remains sparse and the physical properties of plastic contra bone remains an issue.[28]

VR TEMPORAL BONE SIMULATORS

The earliest report on developing a VR environment for temporal bone dissection can be dated back to 1997.[29] This attempt at creating a virtual temporal bone was based on a polygon rendered model. Shortly thereafter, another group reported on building a VR temporal bone simulator based on histological slices also using a polygon rendered model.[30,31] However, such polygon-rendered surface models have limited use in the simulation of temporal bone surgery and this approach was abandoned for high-fidelity simulation. Instead, volume rendered models based on computed tomography (CT) were introduced.[32,33]

Volumetric models based on CT-derived data, haptic interaction, and 3D stereo graphics have since been used in most VR temporal bone simulator projects: a project from Hamburg, Germany, that was later commercialized as the Voxel-Man simulator;[34] the Ohio State University temporal bone simulator;[35] the IERAPSI project by an European consortium, which ultimately did not prevail;[36-39] a temporal bone simulator from the Stanford BioRobotics lab;[40] and a temporal bone simulator developed at the University of Melbourne.[41]

Others have reported on using commercially available tools and CT-derived data to render a volumetric, virtual model of the temporal bone for neurosurgery.[42] Advances in high-resolution imaging could be used to improve the virtual temporal bone model[43] and a polygon-rendered model based on micro CT-derived data that retains some of the properties of the volumetric models has been reported.[44]

The Visible Ear Simulator, which will be further detailed in the methods section, also uses a volumetric model. In contrast to using CT-derived data, the Visible Ear Simulator is based on high-resolution digital photos of cryo-sections of a fresh frozen human temporal bone.[45,46]

THE ROLE OF VR SIMULATION IN MASTOIDECTOMY TRAINING

The VR simulators that until now have been most thoroughly investigated in relation to training of the mastoidectomy procedure are the Ohio State University, Voxel-man, Stanford, and Melbourne temporal bone simulators. Acknowledging that there are technological differences between the simulators, the following brief overview will consider the evidence of VR temporal bone simulation across the board.

In VR temporal bone simulation, novice surgeons have been found to use more time, remove less bone, more often have the drill-tip obscured, and make more injuries to the sigmoid sinus compared with ORL residents and consultants.[47] Expert raters trended to be able to discriminate between novices and experts based on their VR simulation performance.[48] Experts achieved better scores on a global rating scale (GRS) than novices and a different force pattern was observed.[49] Another study corroborated that the force applied during simulation displays a distinct pattern according to the level of expertise.[50]

A number of other simulator-gathered metrics have been found to correlate with experience and expertise[51,52] and can be used to discriminate between the performance of novices and experts.[53] Sampling of expert performances and a decision tree can be used to analyze metrics and used for automatic scoring of the virtual mastoidectomy performance.[54–56]

Residents improve on task scores in addition to using less time and having more efficient hand movements from their first to their second VR performance.[57] With repeated practice, novices improve on a range of parameters measured by the simulator including time for completion and number injuries.[58]

VR temporal bone simulation can improve trainees' knowledge of the surgical anatomy of the temporal bone.[41] More interestingly, a group trained with supervised VR simulation performed better in cadaveric dissection than a group receiving traditional teaching methods based on small group tutorials, videos, and models.[59] Correspondingly, two hours of self-directed VR simulation training with automated guidance was also superior to traditional training.[60] In contrast, both a pilot study and a subsequent large multicenter study found no difference in performance between a VR simulation-trained group and a group that had practiced on cadaveric temporal bones.[61,62]

Overall, there is extensive evidence for VR simulation having a role in temporal bone surgical skills training: 1) the performance of novices and experts in VR simulation can be discriminated, 2) many simulator-gathered metrics are correlated with experience, 3) repeated practice leads to improvement in performance, establishing a learning effect, and 4) VR simulation training seems to be equal to cadaveric dissection training in improving mastoidectomy performance and superior to traditional methods such as video demonstrations.

ASSESSMENT OF SURGICAL TECHNICAL SKILLS

The dogma "*assessment drives learning*" is one way of viewing the relationship between learning and assessment. In the framework of competency-based training, assessment is crucial in defining the standard level for proficiency and verifying the achievement and maintenance of competency. In medical educational research, assessment is the foundation for documenting the effect of educational interventions. Understandably, assessment is also a key component in evaluating the effect of VR simulation-based training.

Traditionally, assessment of surgical skills has been based on 1) hours of training and a log of procedures, both which yield no information on quality,[1] 2) direct observation or videotaping of surgical procedures,[1] and 3) written or oral tests. In addition, assessment of technical competence in surgery is often primarily informal and unstructured.[63] In modern surgery and high quality training, assessment that relies on self-reporting, log books, hours of training, or written tests to determine procedural competency does not suffice.[64]

In the most recent decades, a new paradigm in surgical education has been the introduction of objective, standardized, assessment of surgical technical skills (OSATS).[1,65] This includes task-specific checklists, global rating scales, and final-product analysis, and can be used to assess the surgical performance in a structured manner both in real-life procedures or in a simulated setting. VR simulation can provide a standardized, reproducible environment for such objective assessment. Any good assessment tool must, however, be feasible, reliable and valid.

A recent systematic review of simulation-based assessment in health professional education found that most validity evidence was concentrated within specific areas such as laparoscopic surgery and certain assessment tools (i.e. the OSATS).[64] In addition, validity evidence was found mostly to concern relation with other variables with limited evidence for content, response process, internal structure, and consequences of testing,[64] all key components of validity in Messick's framework, which constitutes the current standard for validity evidence.[66]

Assessment of the surgical performance comprises a range of domains, which includes but is not limited to surgical technical skills. Nonetheless, technical skills are very important in surgical specialties, and within procedure-based assessment of trainees in ORL, the technical skills domain has been found to most greatly impact on the overall rating and be the best predictor of total score.[67]

ASSESSMENT OF MASTOIDECTOMY PERFORMANCE

Mastoidectomy is one of the key procedures in ORL and in the literature, several tools for assessment of performance and competency in mastoidectomy have been described, attempting to evaluate technical skills, process, and final-product of the mastoidectomy performance.

An array of assessment tools was introduced by a group from Toronto in 2007, consisting of a global rating scale (GRS), a task-based check-list (TBC), and a tool for final-product analysis (FPA) to comprehensively assess mastoidectomy performance.[68] Similarly, a two-part assessment tool developed by a group from

Johns Hopkins consists of a checklist of procedural steps (TBC) and a 10-item global preparation and process scale (GRS).[69,70] Conflicting evidence on the correlations between GRS, TBC and FPA in these studies suggests that they capture different attributes and domains of the performance. This is further corroborated by a study on developing a cross-institutional grading scale for temporal bone dissection performance: 24 key items in mastoidectomy was identified using a Delphi-like approach and these items were related to both technical skills and process as well as final product.[71]

In general, these complex assessment tools rely on direct or videotaped observation of the entire (or large parts of the) procedure and this has limited widespread use in the clinical context as well as in research. In contrast, final-product-based assessment considers only the outcome of the procedure by a visual inspection of the drilled temporal bone. This has resulted in FPA being the most commonly reported performance outcome in relation to mastoidectomy training.[20,59,61,62,72–74]

None of the current mastoidectomy performance assessment tools have validity evidence concerning all five sources of validity in Messick's framework and consequently, the evidence for each specific tool is scant.

The most widely used FPA tool is the Welling Scale (WS1), which was developed for assessment of temporal bone dissection mastoidectomy performance at the Ohio State University.[75] An analysis of this FPA tool using generalizability theory has demonstrated that the inconsistent performance of trainees contributed to most of the measurement error and that assessment of multiple performances would be needed to ensure reliable assessment.[76]

For feasibility in the context of multiple raters, performances, and modalities, we therefore used FPA based on a modified Welling Scale as the primary outcome measure as detailed in the methods section. An example of a VR simulation and cadaveric dissection final-product is illustrated in *Figure 1*.

Figure 1. An example of the final products in VR simulation (left) and cadaveric dissection (right).

REPEATED PRACTICE, LEARNING CURVES AND RETENTION

VR simulation-based training of surgical skills is one educational strategy but the effect of VR simulation is dependent on the learning context and objectives, and the instructional design. A systematic review and meta-analysis investigated the role of different instructional design features and identified range of difficulty, repeated practice, distributed practice, cognitive interactivity, multiple learning strategies, individualized learning, mastery learning, and feedback as having a major impact on the effect of simulation-based training.[77] In the following, the evidence on some of these factors will be elaborated.

A learning curve illustrates the correlation between learning effort, for example repetitions or time spent practicing, and the resultant learning outcome (*Figure 2*).[78] Learning curves have implications for training and are useful in evaluating the effect of educational strategies and interventions. In addition, learning curves can be used as an assessment metric, inform competency frameworks, and be used to support self-regulated learning (SRL).[78]

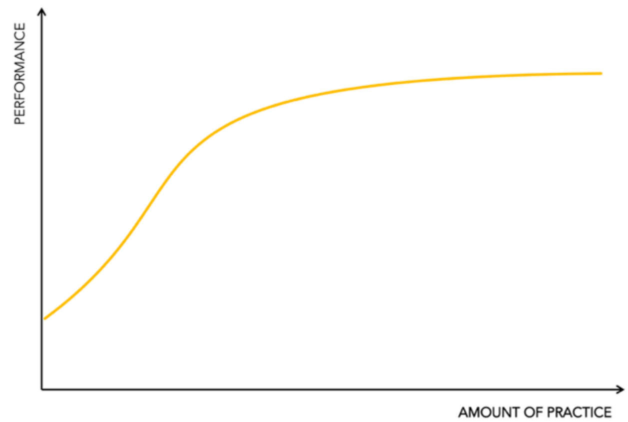

Figure 2. An example of a classical, negatively accelerated learning curve with an upper asymptote (*after Pusic et al.*[78]).

In mastoidectomy, at least 10–15 procedures are needed for technical and basic competency.[79,80] However, for many surgical procedures, there is a substantial slope of the learning curve and procedures need to be performed up to 100 times to reach proficiency, resulting in patient discomfort, longer examination time, and increased risk of complications.[81] Simulation-based training is hypothesized to be able shorten this learning curve.[81] However, learning curves are highly individual and whereas the average learning curve of a surgical procedure is useful for some purposes, different learning curve patterns in surgical training should be acknowledged:[82,83] a few novices demonstrate immediate proficiency and therefore modest further improvement with repeated practice; the largest group improve with repeated practice but at different rates; and a small proportion of learners (~12 %) consistently underperforms and might not improve even with extensive practice and a substantial tutoring effort.[84] There is limited knowledge on the reasons for these differences in learning curves and often the learning curve plateaus after an initial steep phase.[78]

It seems that experience and repeated practice are necessary but not sufficient for true expertise and other mechanisms are needed to overcome the learning curve plateau. Ericsson introduced the concept of deliberate practice in the early 1990s to explain why some professionals become true experts whereas most professionals reach and maintain a stable, average level of performance after having improved with experience initially (*Figure 3*).[85] True experts have not just achieved automaticity of the procedure but use different strategies to continuously improve.[85] The notion of deliberate practice also has implications for simulation-based training[86] and performance plateaus have limitations in determining training endpoints.[87] Although novices seemingly plateau early during simulation-based training, there is some evidence that continued practice increases automaticity of the procedure and reduces the cognitive load of the learner.[88,89] Mastoidectomy performance in VR simulation appears to plateau early,[58] and recently, deliberate practice and its role in skills improvement has gained some interest in ORL and temporal bone surgery.[90,91]

Equally important to the learning curves, retention of what has been learned is needed when evaluating learning strategies and design interventions: the retention of skills is a better indicator of actual learning than performance during practice because the goal of training is consolidated skills and a consistent performance.[92] In general, the retention of surgical skills is determined by the procedure or task, and time since training, but this is modified by factors such as deliberate practice, part-task training, task variability, cognitive demands, and overlearning after reaching proficiency.[93] This leads to heterogeneous reports on the long-term retention of surgical technical skills after simulation-based training.[92,94–97] There is very limited knowledge on the retention of mastoidectomy skills after VR simulation training and the effect of different training strategies.

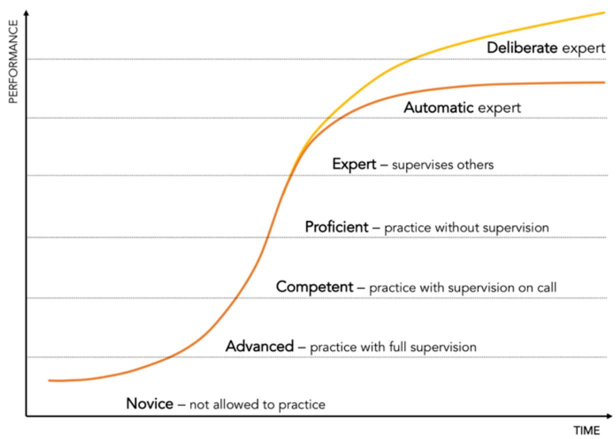

Figure 3. Model of expertise and deliberate practice (*after Pusic et al.*[78]).

PRACTICE ORGANIZATION—DISTRIBUTED AND MASSED PRACTICE

Surgical skills training can principally be organized either with practice distributed in multiple or recurring learning events or with practice massed into a single, finite learning event. Distribution of practice can refer both to distribution of content into several lessons and to practice sessions being spaced by time, often days or weeks, and is also termed (time) spaced or interval training. In this thesis, 'distributed practice' will refer to practice sessions being distributed in time with each session consisting of repeated practice of the exact same procedure.

Clinical training and surgical apprenticeship can be considered a longitudinal learning event with learning and skills acquisition distributed over a longer period of time. An open skills and dissection lab, where the surgical trainee can practice repeatedly during the entire residency, is an example of skills training being organized with distributed practice. In contrast, dissection courses and the increasingly popular surgical boot camps[98,99] are examples of short and intense learning events with massed practice. Distributed practice has consistently been demonstrated to result in superior psychomotor skills acquisition, retention, and skills transfer.[92,100,101] Time-dependent consolidation of memory[100] and other factors[102] add to the positive effect of distributed practice. In myringotomy, distributing practice sessions by a single day was, however, insufficient to improve novice performance[103] and in general, the ideal inter-training interval is still debated.[104]

Another concept concerning practice organization that needs to be introduced is dyad training because participants in *study I* were teamed in pairs. In general, dyad training refers to peers practicing in pairs which could potentially reduce the cost of individual training in complex simulation-based training.[105] There is some effect on motor skills acquisition from observational learning but it is less efficient than actual, physical practice.[105] Current evidence on the effect on motor skills performance of dyad training is unclear, but dyad training could benefit non-technical skills such as communication and cooperation, motivation, and meta-cognition.[106]

FEEDBACK, TUTORING, AND DIRECTED, SELF-REGULATED LEARNING

Constructive and timely feedback by the surgical teacher is a core component of the apprenticeship model.[1] Feedback is also consistently identified as the single most important factor in effective simulation-based training.[7,77] Feedback includes both formative feedback (ongoing/concurrent, typically qualitative) and summative feedback (benchmarking/terminal feedback, often quantitative) feedback, and both are essential for the trainee to develop surgical competency.

A meta-analysis has found feedback in simulation-based procedural skills training to have a considerable beneficial effect on skill outcomes with terminal feedback being more effective than concurrent feedback on long term skill retention.[107] Multiple sources of feedback, including concurrent instructor feedback, improve immediate performance but the long term effects remain mostly uninvestigated, and consequently there is little knowledge on whether the effect on performance is sustained for a prolonged period.[107] In contrast, the guidance hypothesis suggests that constant feedback could lead to over-reliance on feedback and tutoring, resulting in a decline in performance when feedback is withdrawn.[107]

A potential advantage of VR simulation-based surgical skills training is providing the trainee with simulator-based feedback, guidance and tutoring. Ideally, this could support self-directed (as in autonomous) practice and allow the trainee to practice repeatedly without needing human instructors, which as previously discussed can be a limiting factor due to financial, duty, and time constraints. However, the term 'self-directed learning' (SDL) is ambiguously used in the literature and some distinctions should be made:[108] SDL often refers to a design feature of the learning environment that allows for personal learning strategies but a strict definition would include that in SDL, the learning task is always defined by the learner such an example being problem-based learning (PBL). In contrast, self-regulated learning (SRL) allows for the learner to use individual learning strategies but the task can be pre-defined. Both educational strategies fosters personal learning strategies and student engagement in the learning situation as well as goal-directed behavior and self-evaluation of the learning process.[108] SDL can encompass SRL, but not the opposite, and the substantial overlap between the concept SDL and SRL results in the terms often being used synonymously.[108] Some SDL methods use unguided or minimally guided approaches in which the learner must discover or construct essential information for themselves, which leads to suboptimal learning.[109] In contrast to such experiential learning methods, directed, self-regulated learning (DSRL) combines unsupervised training with a strong instructional design supporting SRL by a structured approach.[110] A conceptual framework in relation to SDL could include supervision (by instructors or peers) as one dimension and instructional design features supporting SRL as the second dimension (*Figure 4*).[111] Several studies have found DSRL training to be effective,[60,112,113] and compared with instructor-regulated learning, DSRL led to a better long term outcome.[110]

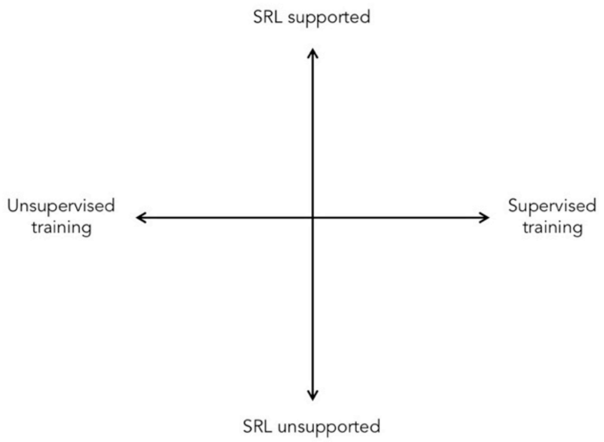

Figure 4. A framework for self-directed learning. Directed, self-regulated learning (DSRL) is an example of unsupervised training with strong support for self-regulated learning (SRL)(*after Brydges et al.*[111]).

In the studies included in this thesis, the VR simulation training was organized with unsupervised training (in *study I* there was some degree of dyad practice), and although the general term SDL applies, DSRL is the more accurate term in our studies: the training was based on a structured approach to the procedure, supporting SRL with sub-goals and suggestions for technique provided by the built-in instructions, feedback by the simulator-integrated tutor-function, while supporting personal learning strategies and engagement in the practice.

When turned on, the simulator-integrated tutor-function provided formative feedback by visually highlighting the volume to be drilled in each step corresponding to the built-in instructions. The participants were not provided with summative feedback at the end of the procedure because the current simulator-integrated metrics do not correlate with final-product performance[114] and because immediate final-product analysis was not feasible.

RESEARCH QUESTIONS
The overall aims of this thesis are to investigate novices' performance in directed, self-regulated VR simulation of mastoidectomy and increase evidence on VR simulation-based training in temporal bone surgery. The specific research questions relating to the papers included in this thesis are:

Study I: Can a modified version of an established FPA tool—the Welling Scale—be used for the assessment of mastoidectomy performance in VR simulation training?

Study II: What is the effect and transferability of two hours of directed, self-regulated VR simulation training of novices on cadaveric dissection performance in mastoidectomy?

Study III: Do different practice conditions (distributed and massed practice) and initial simulator-integrated tutoring in VR simulation-based mastoidectomy training affect the learning curves of novices and can this be used to inform an optimal VR simulation-based training program?

Study IV: Are mastoidectomy skills and secondary task performance after VR simulation-based training of mastoidectomy with distributed or massed practice retained three months after initial training?

Study V: Can the progression of novices in relation to the performance plateau in directed, self-regulated VR simulation training of mastoidectomy be mapped and what are key areas of difficulty, the role of practice condition, simulator-integrated tutoring, and self-assessment?

HYPOTHESES
Study I: FPA can be used to assess virtual mastoidectomy performances.

Study II: Directed, self-regulated VR simulation training can improve mastoidectomy skills and these skills are transferable to cadaveric dissection.

Study III: Practice organization and simulator-integrated tutoring in VR simulation training of mastoidectomy affect the performance curves of novices and can be used to design a training program.

Study IV: Different training strategies affect the 3-month retention of mastoidectomy skills and the cognitive load during the retention test performances in VR simulation.

Study V: Key factors causing the plateau in novice performance of VR simulation mastoidectomy can be identified and novices exhibit limited self-assessment skills in relation to the procedure.

CONTEXT OF THE STUDIES—THE DANISH MEDICAL EDUCATIONAL SYSTEM
To provide some context to the studies and the participants' background, the Danish medical educational and health care system is briefly introduced in the following.

PRE-GRADUATE MEDICAL TRAINING
In Denmark, completion of a 3-year bachelor in medicine (BSc in medicine) is a prerequisite for the 3-year masters' degree program in medicine (MD). Both are only offered at the four major universities of Denmark: The University of Copenhagen, the University of Aarhus, the University of Southern Denmark (SDU) and the University of Aalborg. Higher education including medical school is free for national students as well as students from the European Union member states and Norway. The government provides a state educational grant, which can be supplemented by cheap state educational loans and/or student jobs. For 90 % of the yearly intake of students (50 % at SDU), admission into medical school is centralized and based on the high school grade point average with the remainder being admitted on the basis of a written, motivated application and/or a multiple choice test. This results in a 3:1 female to male ratio among students (~1:1 at SDU).

The curriculum varies between the different medical schools and has recently undergone major revisions to underpin clinical relevance. Generally, the bachelor in medicine includes traditional basic science subjects such as anatomy, physiology, molecular biology etc. as well as biomedical statistics and research methodology, medical psychology, communication and professionalism. The masters' program includes theoretical medicine and surgery, psychiatry, laboratory medicine etc. in addition to hands-on communication and skills training at the skills laboratory and several clinical placements.

PRE-GRADUATE OTORHINOLARYNGOLOGY

The School of Medicine at the Faculty of Health and Medical Sciences at the University of Copenhagen is the largest of the four medical schools in Denmark with a total of more than 3,800 medical students in the bachelor and masters' programs.[115] At the time this research was conducted, clinical and theoretical otorhinolaryngology was taught during the first semester of the final year of medical school. The course consists of 10 one-hour lectures and 10 three-hour classes in which theory is combined with practical otorhinolaryngology in relation to patients brought into the classroom. Students are invited to follow the clinical work at the department for 1–2 days during the course but this is not mandatory.

Other opportunities for pre-graduate experience with otorhinolaryngology at the University of Copenhagen currently includes a 4-week clinical elective during the final year, international exchange, research for bachelors' or masters' thesis, or involvement with the students' interest group for otorhinolaryngology (SØNHKS). SØNHKS arranges presentations on different subjects relating to the specialty and hands-on courses including animal and cadaveric dissection.

Studies III–V included medical students from all years of study thereby including both bachelor and masters' students. Students were recruited through notices in the weekly magazine for medical students and through SØNHKS. The participants were all considered novices because otorhinolaryngology is introduced during the final year and includes limited clinical exposure to otorhinolaryngology in general and even less to temporal bone surgery. VR simulation training was organized as a voluntary extracurricular activity and participants received no financial compensation for participating. Upon completion of training, participants received a certificate of participation.

POST-GRADUATE MEDICAL TRAINING

In order to obtain Danish authorization and permission to work independently as a doctor, a 1-year internship for basic medical training (KBU) must be completed after medical school. This consists of two internship rotations, most often starting with 6 months at a hospital (in emergency medicine/internal medicine/orthopedic surgery/general surgery) followed by a 6-month rotation in general practice. Currently, otorhinolaryngology is never a part of the internship. Internship is allotted by the national board of health by draw and positions are chosen by each intern in sequence according to the number drawn. After internship, specialist training can be commenced and consists of a 1-year introductory position (6 months for general medicine) offered by any specialty department (PGY1) followed by a 4-year residency (PGY2–5) program, which is coordinated regionally. There is a 2:1 ratio between introductory positions and residency positions resulting in competition for residency. Regional appointment committees select residency candidates on the basis of a written application with curriculum vitae and an interview.

POST-GRADUATE TRAINING IN OTORHINOLARYNGOLOGY

After the one-year introductory position in otorhinolaryngology, the 4-year residency consists of structured rotations with placements at 1) a central hospital for at least one year, 2) a university hospital for at least one year, 3) a department of audiology for 6 months, and 4) independent specialist practice for 6 months. It should be noted that with the increasing centralization over the past decades even the smallest central hospitals have a patient population of at least 250,000 persons. In the eastern part of Denmark, three ORL departments service approximately 2.6 million citizens (45 % of the Danish population). Denmark has a national health care system financed by income taxes and health care remains largely free. Private hospitals constitute only a small proportion of the health care system in Denmark with much financing originating from government to reduce public hospital waiting lists. About half of Danish otorhinolaryngologists are in independent specialist practice, servicing community needs and the majority of patients. Referral is not needed to consult an otorhinolaryngologist in specialist practice and consultation and treatment is also paid by the national health care system. Temporal bone surgery is a hospital-based subspecialty and practiced mainly by sub-specialized otorhinolaryngologists.

The number of residency positions is regulated by the national board of health based on projections for the need of specialists. Currently, a total of 16 residency positions are offered yearly in ORL in Denmark. Residency training is regulated by a national training program with a description of competencies and goals for residency. There is no specialist examination upon completion of residency and specialist license is based on formative evaluation with work-place based assessment throughout residency. During residency, 8 mandatory training courses must be completed including a course on middle ear and temporal bone surgery. These mandatory training courses are held on a national level and participants can only participate in each course once. In *study I and II*, participants were recruited from the temporal bone courses held in 2012–2015. Residents were PGY2–5 and were considered novices because course participation is a prerequisite for supervised surgery (aside from temporal bone surgery rarely being performed during residency due to centralization and sub-specialization). A couple of participants had participated in an international temporal bone course and a few had tried the VR simulator (<1 hour).

THE SIMULATION CENTRE AND LOCAL SIMULATION TRAINING

The Simulation Centre at Rigshospitalet is a part of the Copenhagen Academy for Medical Education and Simulation (CAMES) at the Centre for HR in the Capital Region of Denmark. The Simulation Centre at Rigshospitalet provides surgical skills training to doctors at all levels (from internship throughout residency as well as specialists) in most surgical specialties and most of the activity at the Simulation Centre is aimed at trainees from the Eastern part of Denmark.

Training at the Simulation Centre is research-based and uses a four-step approach to training with theoretical preparation, introduction by a clinical specialist, directed, self-regulated practice, and certification assessed by a specialist.[116] In 2015, the number of certificates of competency in a surgical procedure using simulation-based training was more than 1000.[116] Research is a primary mission of the Simulation Centre in order to provide evidence-based training and the Simulation Centre published more than 50 peer-reviewed publications in 2015. Specialists are in charge of each specific procedure as well as the group of instructors for that procedure. A team of dedicated and experienced medical students and two full time nurses assist with technical simulator expertise during directed, self-regulated practice.

A VR simulation set up with the Visible Ear Simulator is found in many ORL departments in Denmark, but our participant questionnaire revealed that the residents have very limited experience (<1 hour) with the simulator, if any at all. At least until now, participants in the national temporal bone courses have not received systematic training locally. However, this might change with more departments investing in the hardware necessary for running the Visible Ear Simulator and past course participants pioneering local training and being able to offer basic technical assistance with the simulator. This will make repeated VR simulation practice prior to course participation feasible in the future.

METHODS

VIRTUAL REALITY SIMULATION PLATFORM

The Visible Ear Simulator (VES) is an advanced temporal bone simulator provided as academic freeware for download of the Internet.[117] The simulator runs on personal computers (*Figure 5*) with Windows (Microsoft, Redmond, WA, USA) as the operating system and a high-end GeForce GTX graphics card (Nvidia, Santa Clara, CA, USA).[45,46] 3D stereo graphics can be accomplished using standard, anaglyphic glasses and other outputs. The simulator supports drilling with force-feedback and haptic interaction using the Geomagic Touch (previously Phantom Omni)(3D Systems, Rock Hill, SC, USA) and simulation of sharp and diamond drills of different sizes (0.5–7 mm). The simulator does not simulate bone dust or bleeding and because suction/irrigation is not needed, the simulator only uses one haptic device (for drilling). The haptic device can be moved to accommodate right and left handed learners.

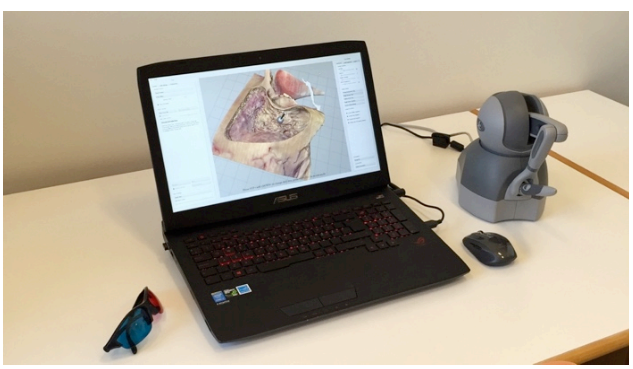

Figure 5. The Visible Ear Simulator running on a laptop with the Geomagic Touch haptic device (right) and anaglyphic glasses (left).

VES currently has a single temporal bone model, which is based on high-resolution digital photos of cryo-sections of the temporal bone of a human cadaver.[118] This provides authentic colors of the virtual specimen. Each of the 200µ slices was manually segmented to build a precise model with the ability for visualization and color coding of each structure and allowing for metrics such as collisions (and in the version 2.0 also deformability).

In *study I*, VES version 1.2 was used. An experimental edition of version 1.3 of the VES was developed specifically for studies II–V. The difference between version 1.2 and 1.3 was minor technical improvements. In version 1.2, the instructional guide and the tutor-function needed to be manually loaded and the final-products needed to be saved manually at the end of session as well. In the experimental version 1.3, this was automatized with individual user login with preset conditions for auto-loading the instructional guide and the tutor-function, and auto-save of the final-product at pre-defined intervals.

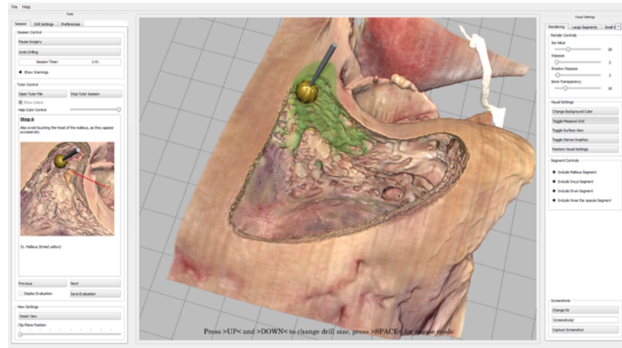

Figure 6. Screenshot from the Visible Ear Simulator with the built-in step-by-step guide (left) and simulator-integrated tutor-function greenlighting the corresponding volume to be drilled (center).

The Visible Ear Simulator has a step-by-step guide with text and illustrations *(Figure 6*). This provides a structured guide to a complete, anatomical mastoidectomy and posterior tympanotomy with anatomical landmarks illustrated, sub-goals for each part of the procedure, and suggestions for technique and selection of an appropriate drill type and size.

The optional simulator-integrated tutor function highlights the volume corresponding to each step of the tutorial with a green color, to visually and intuitively guide the procedure. Both the step-by-step guide and the simulator-integrated tutor function can be turned on and off.

The Visible Ear Simulator can record some basic metrics in addition to time used and steps completed in the tutorial: the volume drilled inside and outside of the reference volume (corresponding to the volume highlighted by the tutor function); and collisions with the dura (collisions with the sigmoid sinus included), the facial nerve, chorda tympani, malleus, incus and stapes, and the inner ear space. Study participants were not provided with this information.

Finally, for studies IV and V, a small computer program was developed by computer graphics engineer Peter Trier Mikkelsen, the Alexandra Institute, Aarhus, Denmark, for calculating the volume difference between two saved files. This tool can be used to map the progression between saved files (final-products) from different time points or for calculating the volume drilled inside and outside the reference volume.

TEMPORAL BONE COURSE ORGANIZATION AND DISSECTION SET-UP

For a number of years, the national temporal bone training course for ORL residents has consisted of: 1) two days with lectures on the diseases of the temporal bone and the middle ear, the related anatomy and physiology, and a theoretical introduction to surgical anatomy, technique and treatment, and VR simulation (*Figure 7*), immediately followed by 2) two days of cadaveric dissection (*Figure 8*) on donated material at the dissection laboratory at the Department of Cellular and Molecular Medicine, the Faculty of Health and Medical Sciences, University of Copenhagen, Denmark.

During dissection training, each participant successively completes a mastoidectomy with entry into the antrum and posterior tympanotomy, tympanoplasty type I–III, cochleostomy, radical mastoidectomy and labyrinthectomy on one side of a human head. Participants are provided personal safety equipment, an operating microscope, an otosurgical drill, a standard array of different drill bits, and suction/irrigation.

Participants have access to a number of standard dissection manuals in the course material, which they can bring with them during dissection. In the original course format, participants received individual and plenary feedback by senior faculty (and peer feedback) as in *study I*. However, for *study II*, this was changed so that participants did not receive feedback during the mastoidectomy procedure and only after the assessment of the mastoidectomy and for the following procedures, participants received feedback by the instructors.

Figure 7. The VR simulation setup: ten dedicated personal computers from the Department of Otorhinolaryngology at Rigshospitalet running the Visible Ear Simulator at the Simulation Centre at Rigshospitalet during the temporal bone course.

Figure 8. Dissection setup during the temporal bone course.

FINAL-PRODUCT ASSESSMENT

Performances were assessed using a modified version of the original Welling Scale[75] for analysis of mastoidectomy final-product performance. The assessment tool was modified to reflect the steps in the VR simulator's built-in instructions to the procedure and local tradition in teaching the procedure.[114] The modified Welling Scale consists of 25 or 26 items, depending on whether the posterior tympanotomy is also evaluated. Similar to the original assessment tool, each item is rated dichotomously with 0 points for an inadequate/incomplete performance and 1 point for an adequate/complete performance. The modified Welling Scale is provided in *Appendix 1*.

In all the studies, performances were rated by the same two expert raters (Prof. Mads Sølvsten Sørensen and Dr. Per Cayé-Thomasen) with the addition of a third rater in *study II* (Dr. Søren Foghsgaard). The raters could not be blinded in relation to training modality (cadaveric dissection vs. VR simulation), but were blinded to participant, participant level (PGY, and in the case of VR simulation also medical student vs. resident), repetition number (in repeated practice), practice condition (distributed vs. massed practice), and tutoring condition (simulator tutored vs. non-tutored). A handful of the performances were rated on paper forms whereas the majority were rated using a digital version of the assessment form running on an iPad.[119]

SUMMARY OF THE STUDIES

STUDY I: MASTOIDECTOMY PERFORMANCE ASSESSMENT OF VIRTUAL SIMULATION TRAINING USING FINAL-PRODUCT ANALYSIS

Background

Objective, structured assessment of technical competency in surgical procedures such as mastoidectomy is needed to evaluate the outcome of surgical training, educational interventions and training tools including VR surgical simulators. Various assessment tools can capture different domains of the surgical performance such as technical skills, process, or final product. Objective, structured assessment can also guide feedback and self-monitoring in directed, self-regulated learning. Finally, defining a proficiency level for competency-based surgical training and future automated simulator-based assessment is dependent on validation against established assessment tools.

The Welling Scale (WS1) is a tool for final-product analysis of mastoidectomy performance and consists of 35 items rated dichotomously as adequate/complete or inadequate/incomplete.[75] In contrast to other assessment tools, which require direct or video-taped observation of the performance, final-product analysis enables assessment at a later point in time. This advantage makes larger studies with repeated performances and multiple raters feasible.

However, final-product analysis has previously only been used to assess cadaveric dissection mastoidectomy.[59,62,75] We made modifications to the original Welling Scale to reflect the procedural steps incorporated into the Visible Ear Simulator and investigated the use of FPA for mastoidectomy performance in VR simulation.

Methods

The study included ORL residents participating in the national temporal bone courses in 2012 (17 participants) and 2013 (17 participants). Participants performed a single, complete mastoidectomy with entry into the antrum (but not posterior tympanotomy) in the Visible Ear Simulator version 1.2. On the following day, participants performed a similar procedure during dissection training at the Faculty of Health and Medical Sciences, University of Copenhagen.

The participants were teamed in pairs during both VR simulation and dissection training. During the VR simulation session, participants were provided with the on-screen step-by-step guide and simulator-integrated tutoring but received no further instructions or feedback by faculty. In contrast, participants got individual and plenary instruction and feedback by faculty during dissection training and had access to a traditional temporal bone dissection manual.

The VR simulation and cadaveric dissection mastoidectomy performances were assessed using the 25-item modified Welling Scale for final-product analysis. Performance rating was done by two senior otologists blinded to participant.

Results

The VR simulation and cadaveric dissection mastoidectomy final-product performance scores were significantly correlated (Pearson's r=0.31, p=0.01). For VR simulation performances, the modified Welling Scale was found to have substantial inter-rater reliability (=0.77, 95 % CI [0.72–0.81]) whereas it for cadaveric dissection performances demonstrated moderate inter-rater reliability (=0.59, 95 % CI [0.54–0.64])(*Figure 9*).

VR simulation mastoidectomy performance correlated significantly with the number of steps completed in the simulator's on-screen step-by-step guide (Pearson's r=0.62, p<0.001) and there was a trend towards association with the amount of reference volume removed (Pearson's r=0.33, p=0.06). Other basic metrics such as time and collisions did not correlate with final-product performance.

Concerning a potential effect of peer feedback and dyad practice, there was no significant difference on final-product performance of performing the procedure first or last, having observed the first procedure, in neither VR simulation (mean difference 0.03; ANOVA, p=0.98) nor cadaveric dissection (mean difference -1.22; ANOVA, p=0.31).

Conclusion

Final-product analysis can be used to assess VR simulation mastoidectomy performance with a substantial inter-rater agreement. VR simulation and cadaveric dissection performance correlates to some extent.

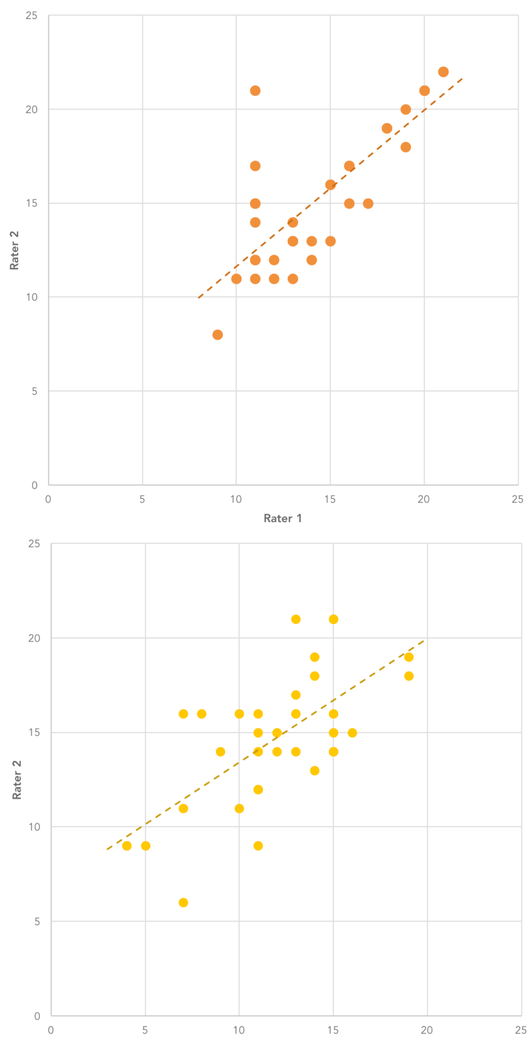

Figure 9. Inter-rater correlation for VR simulation performances (top) and cadaveric dissection performances (bottom).

STUDY II: THE EFFECT OF SELF-DIRECTED VIRTUAL REALITY SIMULATION ON DISSECTION TRAINING PERFORMANCE IN MASTOIDECTOMY

Background

In temporal bone surgery, gold standard training has traditionally been based on dissection of human cadaveric temporal bones.[19] In the recent decade, several sophisticated temporal bone VR simulators have emerged, enabling advanced surgical procedures to be trained outside of the operating room with mounting evidence supporting these simulators as effective training tools for novices.[47,50,58–60,62,120,121]

Previous studies on the effect of VR simulation training on cadaveric dissection performance of mastoidectomy have used multilayered training and instructional interventions.[59,60,62] In this study, we wanted to establish the isolated effect of two hours of self-directed VR simulation training of novices on cadaveric dissection mastoidectomy performance using the Visible Ear Simulator.

Methods

Two cohorts of ORL residents participating in the national temporal bone courses in 2014 (20 participants) or 2015 (20 participants) were included in this study. In these courses, training consisted of two hours of self-directed VR simulation training (3 repeated procedures) in the Visible Ear Simulator version 1.3 at the Simulation Centre at Rigshospitalet and traditional cadaveric dissection training (1 procedure) at the Faculty of Health and Medical Sciences, University of Copenhagen. In 2014, participants completed VR simulation training before cadaveric dissection, and in 2015, participants performed the cadaveric dissection mastoidectomy before receiving VR simulation training (*Figure 10*).
In both training modalities, participants performed a complete mastoidectomy with entry into the antrum to the point of posterior tympanotomy without peer or faculty feedback. In VR simulation, the simulator-integrated tutor-function was turned on for the first and second repeated procedure but not for the third. During VR simulation training, participants had access to the on-screen step-by-step guide, and during cadaveric dissection participants had access to a traditional temporal bone dissection manual.

The last of the virtual mastoidectomies and the dissection mastoidectomy were assessed using the 25-item modified Welling Scale for final-product analysis.[114] Performances were rated by three senior otologists blinded to participant.

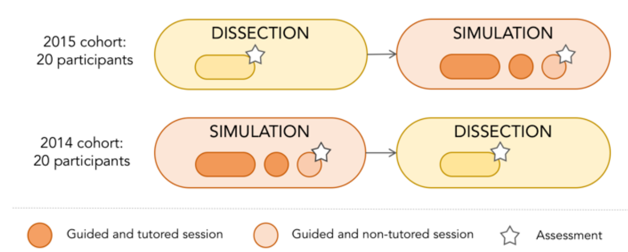

Figure 10. Flow-chart for study II. One cohort received dissection training before VR simulation training and vice versa.

Results

The two cohorts had similar background and characteristics except for self-reported, weekly, average computer usage, which, however, did not correlate with a better VR simulation performance.

Two hours of self-directed VR simulation training significantly increased cadaveric dissection performance by 52 % from 9.8 points (95 % CI [8.4–11.1]) to 14.9 points (95 % CI [12.6–16.9]) (independent samples t-test, p<0.0001)(*Figure 11*) corresponding to the performance in the last VR simulation procedure of 15.5 points (95 % CI [14.2–16.8]). In contrast, cadaveric dissection training first did not improve VR simulation performance (mean difference -1.1 points; independent samples t-test, p=0.22). VR simulation training improved cadaveric dissection performance especially in relation to thinning the bone at the boundaries of the mastoidectomy even though participants scored low on these items in the VR simulation mastoidectomies.

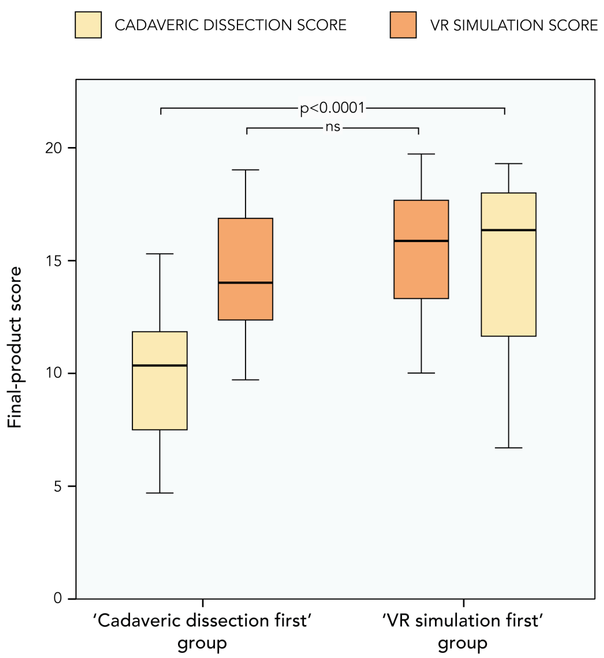

Figure 11. Boxplot of the final-product scores for the group performing cadaveric dissection first (left) and the group receiving VR simulation training first (right).

Conclusions
Two hours of self-directed VR simulation training significantly improved cadaveric dissection mastoidectomy performance in novice ORL residents. Mastoidectomy final-product skills translate from the VR simulation environment to cadaveric dissection and VR simulation training improves performance especially in relation to respecting the boundaries of the mastoidectomy.

STUDY III: LEARNING CURVES OF VIRTUAL MASTOIDECTOMY IN DISTRIBUTED AND MASSED PRACTICE
Background
Surgical skills acquisition requires repeated and deliberate practice[5,85] and organization of practice matters: distributed practice is superior to massed practice in learning complex psychomotor skills because consolidation of skills occur over time.[100,101] VR simulation can support repeated and distributed practice, provide simulator-integrated tutoring, support directed, self-regulated learning, altogether accommodating the trainees' individual learning needs.

The learning curve of any surgical procedure is pivotal because of the implications for training. For mastoidectomy, technical competency in the operating rooms requires performing approximately 10–15 procedures.[79,80] Nonetheless, mastoidectomy performance seemingly plateaus after just 4 repeated procedures in another VR temporal bone simulator.[58]

In this study, we wanted to explore the learning curves of VR simulation of mastoidectomy in novices, using different practice conditions in combination with and without initial simulator-integrated tutoring. Based on this knowledge, we aimed at proposing an improved VR simulation training program in mastoidectomy.

Methods
Forty-three medical students from the Faculty of Health and Medical Sciences, University of Copenhagen participated in this study of repeated practice of mastoidectomy in directed, self-regulated VR simulation training at the Simulation Centre at Rigshospitalet. All participants performed 12 repeated complete mastoidectomies with entry into the antrum and posterior tympanotomy in the Visible Ear Simulator: 21 of 24 participants completed distributed practice with blocks consisting of two repeated procedures spaced at least three days apart; 19 of 19 completed massed practice with all repeated procedures performed consecutively in one day. Participants in both practice conditions were randomized to receive initial simulator-integrated tutoring during the first five procedures or not at all (flowchart, *Figure 12*). All participants had access to the on-screen step-by-step guide at all times but did not receive any further instructions or feedback by faculty.

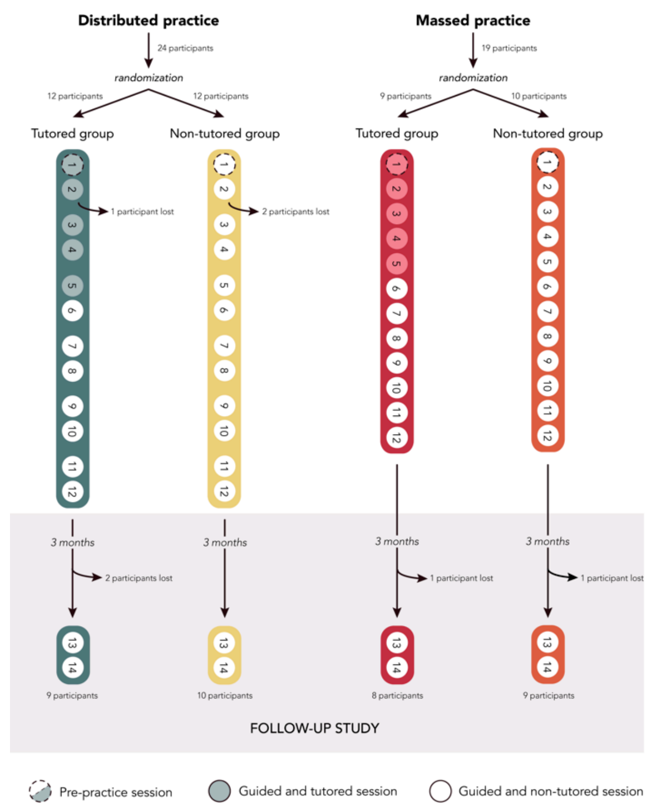

Figure 12. Flowchart for studies III-V.

For the first procedure, which served as an introduction, participants were allowed 60 minutes. For the subsequent procedures, participants were allowed 30 minutes after which the simulator auto-saved the final-product and closed. Moreover, the VR simulator auto-saved final-products at 10-minute intervals.

The mastoidectomy performances at 30 minutes were assessed using the 26-item modified Welling Scale for final-product analysis. Performances were rated by two senior otologists blinded to participant, practice condition, session number and simulator-integrated tutoring.

Results

Distributed practice was superior to massed practice in increasing mastoidectomy final-product performance in VR simulation (*Figure 13*). After four sessions, performance of the massed practice group started to gradually deteriorate whereas performance in the distributed practice group followed a negatively accelerated learning curve towards a performance plateau at 16.0 points (non-linear regression using a sigmoid function, 95 % CI [15.3–16.7]), which was statistically reached by the 9th session.

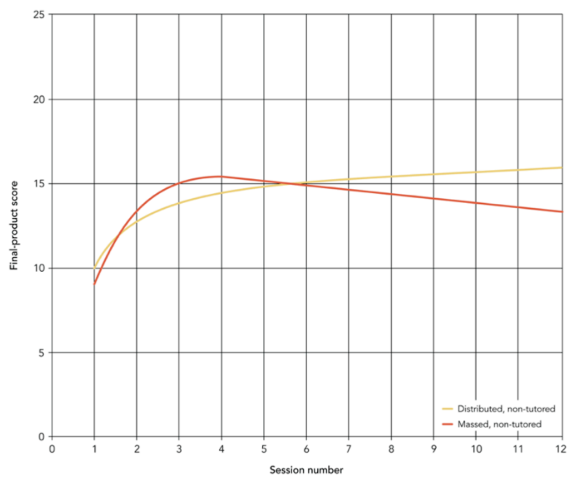

Figure 13. The learning curves of non-tutored participants in distributed and massed practice.

Simulator-integrated tutoring significantly improved final-product performances (ANOVA, p<0.01) and accelerated the initial slope of the learning curve (*Figure 14*). Nonetheless, when simulator-integrated tutoring was discontinued, final-product performance dropped markedly especially if combined with the massed practice condition. With continued non-tutored practice, tutored participants' final-product performance caught up with the performance of the corresponding non-tutored participants.

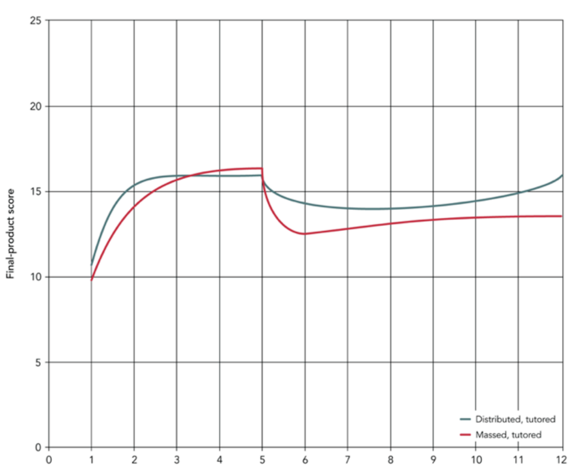

Figure 14. The learning curves of simulator-tutored participants in distributed and massed practice.

Conclusions

Directed, self-regulated VR simulation mastoidectomy training of novices should be organized with distributed practice. Simulator-integrated tutoring can accelerate the learning curve but should be used in conjunction with distributed practice to prevent over-reliance on tutoring. Based on the different learning curves, a more optimal training program could consist of distributed practice with practice blocks of one simulator-tutored procedure immediately followed by two non-tutored procedures. The total number of practice blocks and spacing between blocks need further investigation and VR simulation-based training should also consider the individual learning curves.

STUDY IV: RETENTION OF MASTOIDECTOMY SKILLS AFTER VIRTUAL REALITY SIMULATION TRAINING

Background

The goals of surgical training are consolidated skills and a consistent and high performance but surgical skills are heterogeneously retained and depend on a variety of factors including the procedure, cognitive demands and practice organization.[93,96] Nonetheless, most studies on the effect of VR simulation training in mastoidectomy report only performance during or immediately following practice even though retention of acquired skills is a better indicator of actual learning.[92] In addition, highly complex motor skills such as those required for mastoidectomy can cause a substantial cognitive load, reducing the capacity for learning.[122]

All together this has implications for the organization of training and there is a gap in knowledge of the retention of mastoidectomy skills after VR simulation training. In this study, we wanted to determine the retention of mastoidectomy skills in VR simulation and the cognitive load after distributed and massed practice of the procedure.

Methods

All participants in the previous study on the learning curves of repeated practice of mastoidectomy in a VR temporal bone simulator were invited back for a 3-month retention test at the Simulation Centre at Rigshospitalet (*Figure 12*). The retention test consisted of two mastoidectomies identical to the procedure previously trained with 30 minutes allowed for each procedure and full access to the on-screen step-by-step guide but without simulator-integrated tutoring.

Participants were medical students at the Faculty of Health and Medical Sciences, University of Copenhagen. 19 of 21 participants who had completed distributed practice and 17 of 19 participants who had completed massed practice responded to the invitation and returned for retention testing. None of the participants had practiced the procedure in the intervening time.

Final-product performances were assessed by two senior otologists blinded to participant, session number, practice condition and initial simulator-integrated tutoring. In addition, participants' cognitive load was estimated by performance on a secondary task measuring reaction time in response to a visual cue. Supplementary analyses of the volume removed during VR simulation sessions were used to explore differences in time to completion of the procedure.

Results

Final-product performance did not deteriorate significantly during the 3-month non-practice period regardless of initial practice condition (*Figure 15*). However, the final-product performance of the massed practice group increased significantly from the first to the second retention procedure (paired samples t-test, p=0.001), which rendered the performance of the massed practice group similar to the performance of the distributed practice group (ANOVA, p=0.89). Also, based on the volume analysis, the massed practice group used more of the allowed time to achieve a retention performance similar to the end-of-training performance, suggesting that skills were less consolidated.

For both groups, the reaction time increased to the pre-training level, suggesting that the decrease in cognitive demands by repeated practice is not retained as consistently as the technical skills.

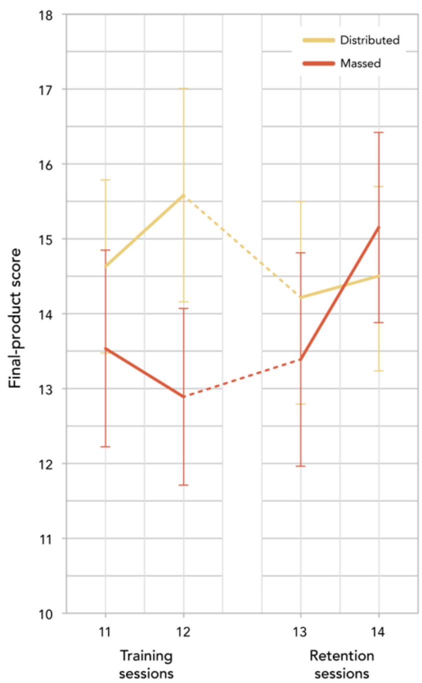

Figure 15. Final-product scores in end-of-training session (left) and retention sessions (right) for the distributed and massed practice conditions.

Conclusions

VR simulation mastoidectomy skills are largely retained at three months after directed, self-regulated training in a distributed practice program, whereas more drilling time was needed to compensate for the less consolidated skills in the massed practice group. Although massed practice leads to a suboptimal outcome of initial training, the learning curve can be resumed. Finally, cognitive skills seem to deteriorate faster than motor skills. Continuing practice of surgical technical skills should therefore be scheduled regularly with a frequency that maintains both motor and cognitive skills.

STUDY V: MAPPING THE PLATEAU OF NOVICES IN VIRTUAL REALITY SIMULATION TRAINING OF MASTOIDECTOMY

Background

Evidence points to self-regulated learning in simulation training being effective[111] but unsupervised or self-guided practice needs special considerations.[113,123] The early and low plateau of novices in directed, self-regulated VR simulation training of mastoidectomy[124] could relate to the complexity of the procedure.[79] However, cognitive skills such as self-assessment could also contribute to the learning curve ceiling[112] and accurate self-assessment and self-regulatory skills are crucial components of successful directed, self-regulated training. Nonetheless, self-assessment can be difficult for the novice.[125] Potential modifiers of self-assessment skills include consolidation of psychomotor skills occurring with distributed but not massed practice[124] and over-reliance on tutoring.[107]

In this study, we wanted to map the progression of novices in relation to the performance plateau in directed, self-regulated VR simulation training of mastoidectomy and to explore key areas of difficulty and the role of practice organization, simulator-integrated tutoring, and self-assessment.

Methods

Data from the performances of 40 medical students from the Faculty of Health and Medical Sciences, University of Copenhagen, who had completed repeated directed, self-regulated VR simulation training at the Simulation Centre at Rigshospitalet in a previous study[124] were included for study.

Participants had completed a training program consisting of 12 repeated mastoidectomy procedures in the Visible Ear Simulator with different training conditions: 21 participants had completed distributed practice with practice blocks of two procedures spaced by at least three days; and 19 participants had completed massed practice with all 12 procedures performed in one day. The VR simulator had auto-saved the final-products at 10, 20 and 30 minutes.

The 30-minute final-products had previously been rated by two expert raters using a 26-item modified Welling Scale for mastoidectomy performance assessment. For this study, the mean rating for each item of the assessment tool was calculated according to different factors such as simulator-integrated tutoring and practice condition.

To map progression, the total amount of virtual bone drilled between 20 and 30 minutes was extracted as well as the amount of bone drilled inside and outside the reference volume for the mastoidectomy. These data were used to identify procedures in which the participant had stopped early.

Results

Several key items of difficulty for novices in the mastoidectomy procedure were identified. These included avoiding to drill holes in the anatomical boundaries of the mastoidectomy and avoiding violations of vital structures such as the lateral semicircular canal and the facial nerve. Simulator-integrated tutoring significantly improved performance on these items but over-reliance on tutoring led to the effect of tutoring not exceeding the effect of repeated practice alone. Distributed practice had a positive effect on performance on many key items of difficulty compared with massed practice. Interestingly, massed practiced with immediate repetition induced self-regulation in areas where the tutor-function enticed risky behavior, indicating issues with the simulator-integrated tutor-function that needs future improvement.

Participants stopped drilling early in 19.3 % of the later sessions (session 6–12). These performances had a mean final-product score of 14.1 points and were not significantly better or worse than the average performance (mean 14.3 points). In addition, the 10 % of participants who consistently did not use all the allowed time performed similar to the remaining participants. Altogether this indicated poor self-assessment skills in novices and a lack of knowledge on when to stop or how to excel.

Conclusions
Several areas need to be improved to overcome the initial performance plateau of novices in directed, self-regulated VR simulation training of surgical technical skills such as the mastoidectomy procedure: key items of difficulty should be specifically addressed in the instructions; over-reliance on simulator-integrated tutoring emphasize the need for tutoring to be embedded in a strong instructional design with process-oriented goals; and lastly, novices exhibit poor self-assessment skills and have difficulty knowing when to stop or how to excel. This needs additional attention when designing a directed, self-regulated training program to stimulate cognitive effort and avoid arrested development.

DISCUSSION
In this thesis on VR simulation training of mastoidectomy, we have investigated the final-product performance of novices under different practice conditions and directed, self-regulated learning to increase evidence on VR simulation training in temporal bone surgery. In the following, the research questions and key issues will be discussed after a general discussion of the limitations of the studies.

LIMITATIONS
None of the studies in this thesis were randomized. *Study I* was non-interventional and the original course design and curriculum was not changed. This could potentially cause differences in the degrees of peer and faculty feedback, in addition to participants dividing the time unequally between them. For *study II*, randomization to an intervention and a control group was planned but due to the dissection equipment arriving one day late in 2014 this had to be abandoned. The raters could therefore not be blinded to whether participants had received VR simulation training first. In *study III*, randomization to distributed or massed practice was not feasible because the time commitment needed of participants was considerably different. This resulted in minor differences in participants' background and characteristics in the two practice groups with participants in the distributed practice group being slightly older, more often male and having a higher frequency of gaming. However, these factors were not found to be associated with final-product performance. *Study IV* was a follow-up study on *study III* and *study V* was based on the data from study III and the lack of randomization for practice group therefore applies to these studies as well.

Another general limitation is the small number of participants in each study. In *study I* and *study II* participants were recruited from the national temporal bone course, which has a yearly number of participants of 17–20 participants. Based on this number of participants in each group, a final-product score difference between the groups of at least 2.5 points would be needed for statistical significance (using the standard deviation and means from *study* I). A meaningful minimal relevant difference on the modified Welling Scale is 2–3 points and therefore corresponds well to the sample size calculation. Sample size calculations for learning curves are not well defined and in *study III* we recruited participants to achieve approximately 10 participants in each group. This small number of participants in combination with the loss to follow-up potentially makes *study IV* underpowered, introducing a considerable risk of a type II error (failure to detect an actual difference between groups). In *study V* we considered the performances from multiple sessions in each group, giving more power to the calculations and adjusted significance using the Bonferroni correction for mass testing due to post-hoc analysis of multiple items.

A limitation specific to *study V* was the definition of the cut-off for performances stopped early. This definition included a small amount of volume to include participants who stopped drilling during the last 5–7 minutes. The cut-off was based on an analysis of the volume progression at 10, 20 and 30 minutes to determine an appropriate cut-off value.

FINAL-PRODUCT ANALYSIS IN THE ASSESSMENT OF MASTOIDECTOMY PERFORMANCE
Final-product-based assessment has previously only been used to assess physical (i.e. cadaveric and plaster) temporal bone drilling. In *study I*, we therefore investigated final-product assessment in relation to the assessment of VR simulation performances, which would be the basis for the performance assessment and comparisons in the subsequent studies.

We made modifications to the original Welling Scale to reflect the procedural steps built in to the Visible Ear Simulator instructions but kept the fundamental design principles with dichotomous rating of the items, many of which were unchanged. These modifications limit direct comparison with results reported for the original assessment tool. Strictly, any modification warrants validation anew, keeping in mind that validity evidence for the original assessment tool is only partial in relation to Messick's framework of validity.66 Nonetheless, we found an inter-rater reliability kappa for the modified Welling Scale (=0.59, 95 % CI [0.54–0.64]) for the dissection performances similar to the original scale (ranging between 0.49–0.64).

For the VR simulation performances, we found a higher inter-rater reliability kappa (=0.77, 95 % CI [0.72–0.81]). Several factors could contribute to this: the single, well-pneumatized temporal bone model featured in the Visible Ear Simulator; the raters' familiarity with this particular temporal bone; and the additional use of simulator metrics to support decisions about suspected collisions.

The standardized and controlled VR simulation environment and the higher inter-rater reliability could be an advantage in high stakes assessment, where validity and reliability are key concerns. Generalizability theory can be used to statistically determine the reliability in different conditions and in high stakes assessment the generalizability coefficient (GC) should preferably be at least 0.8. An analysis of the original Welling Scale using generalizability theory found that much of the measurement error introduced was related to the inconsistent performance of the trainees.[76] Therefore, the rating of several performances of each trainee is vital: with two raters, the assessment of three dissection mastoidectomy performances would result in a GC of 0.7.[76] In other words, high stakes assessment based on cadaveric dissection would be both costly and time consuming. A specific analysis of the VR simulation performances using generalizability theory could establish the effect on the GC of the standardized VR environment but this analysis has not been carried out so far.

Whereas the single temporal bone model in the Visible Ear Simulator might be beneficial for assessment purposes, task variability remains an important component of effective simulation-based training.[7,77] From a training perspective, the lack of alternative training scenarios is therefore currently a limitation of the Visible Ear Simulator.

Final-product analysis for the assessment of mastoidectomy performance considers only the end result and not the technical skills or process and the correlation with these domains are poor.[68] This general limitation should be acknowledged in relation to all the studies in this thesis.

Other issues of final-product analysis need consideration as well. First, to the best of our knowledge, expert performance has not been assessed using the Welling Scale or other final-product analysis tools. Although it could be expected that an expert would achieve the maximum score, this remains uninvestigated and might prove difficult especially in VR simulation. Next, dichotomous rating could easily introduce a ceiling effect—a point where the assessment instrument lacks discriminative abilities or sensitivity[126] (this effect is also found for continuous grading scales). We found such a ceiling effect for the very first performance of novices, where none of the participants achieved a maximum score and few got a score above 20 out of 25 even when provided with sufficient time—but more interestingly, we also observed a peak effect.[127] This peak effect was the result of final-product performance deteriorating after initial progress because a large proportion of the participants kept drilling rather than stopping, thereby making mistakes such as introducing collisions when provided with too much time. Time and performance are dependent factors and this should be considered when using final-product analysis for assessment. Automated final-product analysis could be used for real-time feedback[54] and could potentially reduce the peak effect by making the trainee aware when the performance, e.g. the thinning of a bony surface overlying the dura or facial nerve etc., is satisfactory.

THE EFFECT OF VR SIMULATION TRAINING ON CADAVERIC DISSECTION PERFORMANCE

VR simulation training of mastoidectomy has consistently been demonstrated to improve the cadaveric dissection performance of novices compared with traditional teaching methods[59,60] and VR simulation training and practice on cadaveric temporal bones result in a similar improvement in dissection performance.[61,62] None of these studies have compared VR simulation training to novices' first unsupervised cadaveric dissection performance, establishing the isolated effect of directed, self-regulated VR simulation and the transferability of skills to the dissection setting. In *study II*, we assigned one cohort to perform a cadaveric dissection mastoidectomy without feedback from peers or faculty and without receiving VR simulation training first and another cohort to receive two hours of directed, self-regulated VR simulation training before cadaveric dissection. Unsurprisingly, VR simulation training improved cadaveric dissection performance significantly compared with not receiving VR simulation training first. However, this demonstrated that VR simulation training without peer or instructor feedback can improve cadaveric dissection performance substantially.

We also found that VR simulation training improved the dissection performance on most of the items assessed but interestingly, a poor performance in the VR simulation environment on some items led to a correspondingly better cadaveric dissection performance: drilling holes in the sigmoid sinus, tegmen, and the external auditory canal wall during VR simulation training seemed to caution the participants to proceed more carefully in these areas during dissection, suggesting a learning effect from making mistakes as well.

A number of differences between VR simulation and cadaveric dissection could potentially limit transferability of skills: the anatomical variation and degree of pneumatization of the cadaveric temporal bones; the challenge of handling multiple instruments such as the operating microscope and suction device in cadaveric dissection; the unaccustomed translation of hand movements with the haptic device and so on. Nonetheless, our study corroborates that mastoidectomy skills acquired in VR simulation training translates to the dissection condition because the VR simulation training resulted in a 52 % increase in dissection performance compared with the group performing a dissection mastoidectomy without prior VR simulation training.

Our VR simulation training program consisted of two repeated procedures with the simulator-integrated tutor-function turned on and a final procedure without simulator-integrated tutoring, which was assessed. The effect of VR simulation training alone and the effect of simulator-integrated tutoring can consequently not be separated. It is therefore likely that both the effect of VR simulation training, practicing the procedure with the structured approach provided by the on-screen guide, and the simulator-integrated tutor-function, contributed to the superior performance of the group receiving the VR simulation training intervention first. One other study has found that self-directed VR simulation practice of mastoidectomy was effective in improving cadaveric dissection performance.[60] Together, this supports directed, self-regulated training in improving the dissection mastoidectomy outcome. In general, there is strong evidence supporting the use of VR simulation training in early temporal bone training, which could reserve donated temporal bones for subsequent training once basic competencies have been acquired in the VR simulation environment.

THE EFFECT OF PRACTICE ORGANIZATION ON THE LEARNING CURVES

In *study II*, performing a single mastoidectomy on a cadaver did not improve the VR simulation performance. Repeated practice is important regardless of training on cadaveric temporal bone or VR simulation[20,62] and in *study III* we therefore wanted to investigate the learning curves of repeated practice of VR simulation mastoidectomy using different practice conditions. Training was organized either with practice blocks of two repeated procedures spaced by at least three days or as massed practice with all repeated procedures completing during one day.

In line with established knowledge that massed practice causes suboptimal learning outcomes, we found that massed practice beyond 2.5 hours led to the performance gradually declining. The mechanism for the deterioration of performance can most likely be attributed to mental fatiguing, which is less likely to occur with the shorter practice blocks we used in the distributed practice program. Nonetheless, at many institutions including our own, many surgical skills are taught in massed learning events such as intensive skills courses. Despite solid evidence on the negative effect on learning, the recent increase in the number of publications on surgical boot camps[98,99,128–130] demonstrates that massed learning events are not a thing of the past. Acknowledging that there are many good reasons for the popularity of such intensive training events, there is often limited opportunity for the important repeated practice, allowing for the consolidation of skills. In addition, such courses are most often not planned to match subsequent opportunities for using the acquired skills in the clinical context.

The average performance of non-tutored participants in distributed practice followed a classical negatively accelerated learning curve. The learning curve plateaued after nine repetitions and there was only modest improvement in performance after the sixth procedure. Furthermore, the level of the plateau was low (16 out of 26 points). A learning curve plateau is a common finding in studies on VR surgical simulation training of technical skills[58,81,131,132] and the early and low plateau suggests that for novices either significantly more practice is needed[87] or that other factors contribute to the plateau (see later). Participants in *study III* were medical students but their final-product performance was similar to the level achieved by ORL residents after three repetitions in *study II*. We are currently investigating whether ORL residents can surpass the level achieved by medical students with repeated practice, thereby explaining part of the observed ceiling effect with training level.

The main aim of *study III* was to use the learning curves to design an optimal training program. By analyzing the average learning curves of the different practice conditions, an optimal practice protocol could be suggested to consist of distributed training with practice blocks of three procedures. We are currently investigating such a training program and the use of simulator-integrated tutoring in the first procedure of each practice block for acceleration of the learning curve and regular reinforcement. Hopefully, this strategy can prevent over-reliance on simulator-integrated tutoring. The number of practice blocks needed to achieve a degree of automaticity remains unexplored.

THE EFFECT OF PRACTICE ORGANIZATION ON RETENTION

The retention of mastoidectomy skills after VR simulation training is little investigated and in *study IV* we wanted to determine the retention of mastoidectomy skills three months after initial training. The study was designed as a follow-up study on *study III* and the effect of distributed and massed practice on retention was therefore investigated.

It is well-established that massed practice often causes suboptimal long term retention of skills.[92,100,101] In line with this, we found that the massed practice group needed more time to complete the task and achieve a performance similar to the distributed practice group in order to compensate for loss of skills. In contrast, the retention mastoidectomy performance of the distributed practice group was similar to the end-of-training performance. Mastoidectomy skills were in other words largely retained, but since retention was only tested at three months, this limits the conclusions on longer term retention. Ideally, retention could have been tested at different or multiple points in time, but this would have come at the cost of a smaller sample size, a longer study period, and loss to follow-up.

The massed practice group improved significantly from their first to their second retention performance, suggesting that the learning curve can be resumed, which also has been observed in simulation-based training of endoscopic sinus surgery.[132] The literature also suggests, that some degree of ‘overlearning’ (continued practice after proficiency) is beneficial for skills retention.[95] Overlearning could be favorable for novices because automaticity occurs with continued practice after proficiency has been attained.[88] In the deliberate practice model by Ericsson, automaticity is the level of performance that requires a minimal amount of mental effort—the level achieved by most professionals.[85] Continuing learning towards true expertise and avoiding this ‘arrested development’ requires substantial cognitive effort and deliberate practice.[85] With repeated and distributed practice of mastoidectomy in VR simulation, the cognitive load decreased whereas with massed practice it did not.[89] This could reflect beginning automaticity, but because the secondary task performance did not plateau, 12 repetitions were not enough for novices to achieve automaticity even though their primary performance on the final product had plateaued. This was further corroborated in *study IV* because we found that after three months, the cognitive load had returned to pre-training levels, substantiating that cognitive skills deteriorate faster than motor skills.[93]

DIRECTED, SELF-REGULATED LEARNING AND THE ROLE OF SIMULATOR-INTEGRATED TUTORING AND FEEDBACK

Many of the benefits of VR simulation-based training relates to the opportunity for self-directed training because of the general limitations in time and resources as previously discussed. In contrast to more experiential learning, evidence supports DSRL as effective in surgical skills training.[111] We have previously found the simulator-integrated tutor-function to improve performance especially in the later and most difficult part of the procedure[127] and in *study III*, we also investigated the role of simulator-integrated tutoring in relation to repeated practice.

We found that simulator-integrated tutoring improved the initial final-product performance in both distributed and massed practice and increased the slope of the learning curve. In agreement with the guidance hypothesis,[107] the increase in performance persisted only while the tutor-function was on and a considerable decline in performance occurred when the simulator-integrated tutoring was discontinued. Another study has demonstrated that guidance is beneficial for novice learners but whereas such feedback induced faster skills acquisition it was also followed by a more rapid deterioration of skills.[110] In our study, the more consolidated skills resulting from distributed practice had some protective effect against this performance drop, suggesting that simulator-integrated tutoring is most optimally used in conjunction with distributed practice.

The low plateau of the learning curves warranted further analysis and in *study V* we wanted to identify key factors contributing to the performance plateau in directed, self-regulated practice. In this analysis we found that the participants had difficulty not drilling holes in the outer anatomical boundaries of the developing cavity and leaving adjacent structures intact. The simulator-integrated tutor-function effectively increased performance in relation to this but at the same time also enticed some risky behavior. This knowledge can be used to further improve the simulator technically as well as the built-in guide and the tutor function.

The identified peak and ceiling effects[127] could be the manifestation of the novice participants lacking the knowledge on when to stop and how to excel, respectively. In contrast to other studies that have found novices to be able to adequately self-assess surgical technical skills,[113,133] our analysis revealed that novices had poor self-assessment skills. A possible explanation could be that the mastoidectomy procedure is more complex than the technical skills studied in other reports. Nonetheless, the learning curve plateau could also be a result of a lack of cognitive effort once a self-assessed proficiency level has been reached.[112] In line with Ericsson’s model of deliberate practice, continuous cognitive effort is essential in avoiding arrested development.[85]

Self-assessment is a key component in directed, self-regulated learning[123] and a directed, self-regulated training program should have specific and explicit process goals and supportive and directive instructions.[134] Our findings suggest that we need a stronger instructional design to improve the trainees’ self-assessment skills, and to better support deliberate practice through scaffolding to improve performance and break the performance plateau.

Summative feedback is critically important for technical skills development[107] and remains a key issue in VR simulation training of mastoidectomy. In our studies, participants were not provided any feedback. This currently makes the directed, self-regulated learning strongly dependent on self-assessment of performance. Development of automated assessment for summative feedback would not only resolve this but could also provide the individual trainee and educators with feedback for continuous monitoring of skills development and progression and would encourage deliberate practice. Automated assessment needs to be based on simulator-gathered metrics to provide immediate post-performance feedback and simulator-based assessment should be validated in accordance with Messick's framework. Automated final-product analysis[54] is one possibility for feedback but could be supplemented by a range of other metrics.[52]

Rather than a repetition-based curriculum based on the average learning curve plateau, which is currently the only feasible option in VR simulation-based training of mastoidectomy, setting a standard for proficiency by defining the level of experts using automated assessment would allow competency-based training.[135] Training to proficiency would also address the issue of the high variability of the individual learning curves. In general, such mastery learning provides not only better primary learning outcomes but also has downstream effects such as improved practice and better patient outcomes.[136] Naturally, mastery learning is the next step in further improving training in temporal bone surgery.

CONCLUSION AND PERSPECTIVES

This thesis has investigated directed, self-regulated practice of mastoidectomy in a VR temporal bone surgical simulator to increase evidence of VR simulation-based training in temporal bone surgery. In this thesis, we used final-product analysis for assessment of mastoidectomy performance in VR simulation and cadaveric dissection and even though this has some limitations as discussed, it is widely used for assessment of technical skills in temporal bone surgery and made the assessment of the many performances in the studies feasible.

We found that two hours of directed, self-regulated VR simulation training of mastoidectomy significantly improved performance in cadaveric dissection and that skills transferred from VR simulation to dissection. This suggests that VR simulation training is an useful adjunct to the current gold standard training on cadaveric temporal bones because this might save donated temporal bones for more advanced training after basic competencies have been acquired in VR simulation.

At many training institutions including our own, trainees participate once in a temporal bone course. However, massed practice results in poor learning and we demonstrated that repeated practice should be distributed over multiple practice sessions to allow for optimal learning and consolidation of skills. This was also reflected in the retention performance after three months where the massed practice group in contrast to the distributed practice group needed to use more time to compensate for skills deterioration during the non-practice period.

Simulator-integrated tutoring was useful in directed, self-regulated training, but over-reliance on this feedback could be an issue and simulator-integrated tutoring should be embedded in the training program for regular reinforcement rather than be used continuously during initial training. We are currently investigating a new, distributed training program based on the results presented in this thesis.

The learning curves of novices were highly individual but the mean learning curve plateaued at an inadequate level. We found the poor self-assessment skills of novices to be a possible explanation for this, and this needs to be addressed specifically in the instructional design. In general, directed, self-regulated learning requires a strong instructional design to encourage deliberate practice and to stimulate cognitive effort to prevent arrested development.

The next step in VR simulation training and a possible solution to these issues is automated assessment for immediate post-procedure feedback. In addition, automated feedback can be used in establishing a proficiency level for competency-based training with mastery learning. Automated assessment should be valid and based on evidence, and several groups working on the development of VR temporal bone simulators have reported on their progress in relation to automated assessment.[54–56,74] In combination with methods such as cumulative sum (CUSUM), the learning curve can be used to monitor progress in competency-based training.[137–139]

In conclusion, there is plenty of evidence favoring VR simulation-based training of mastoidectomy as a supplement to traditional training methods, and this thesis has added evidence on factors leading to effective VR simulation-based training. This knowledge should be used in the VR simulation environment to optimize directed, self-regulated learning as well as implemented into the surgical curriculum, to systematically provide ORL trainees with repeated practice, supporting individual needs and embedded in clinical practice. Ultimately, this will not only provide high quality, evidence-based training for surgical trainees but will in turn also translate to improved safety and patient outcomes.

ABBREVIATIONS

CAMES	Copenhagen Academy for Medical Education and Simulation
CT	Computed tomography
DSRL	Directed, self-regulated learning
FPA	Final-product analysis/assessment
GC	Generalizability coefficient
GRS	Global rating scale
OR	Operating room
ORL	Otorhinolaryngology
OSATS	Objective Structured Assessment of Technical Skills
PBL	Problem-based learning
PGY	Post-graduate year
SDL	Self-directed learning
SRL	Self-regulated learning
SØNHKS	Students association for otorhinolaryngology
TBC	Task-based checklist
VES	The Visible Ear Simulator
VR	Virtual reality

ACKNOWLEDGEMENTS

This PhD thesis is based on studies conducted at the Department of Otorhinolaryngology—Head & Neck Surgery at Rigshospitalet and Gentofte Hospital and the Simulation Centre at Rigshospitalet in the period 2012–2015 as part of the project "Development and validation of evidence based software for integrated self-directed learning and assessment in temporal bone surgery. Improving the

Visible Ear Simulator by scientific iteration." for which the Oticon Foundation, Denmark, generously provided a full grant.

The idea for the project was conceived during the process of writing my Master's thesis under the supervision of professor Mads Sølvsten Sørensen, who developed the Visible Ear Simulator with the intent of improving education in temporal bone surgery everywhere, including parts of the world where educational and instructional resources are few. Together with principal computer graphics engineer Peter Trier Mikkelsen from the Alexandra Institute in Aarhus, who programmed the simulator from scratch, and with funding from the Oticon Foundation, they made the simulator publicly available as academic freeware. This altruism and vision has served as my greatest inspiration throughout the process.

I am truly grateful and indebted to both Peter, who has taken the time to program my shrewd ideas and odd requests, and Mads, who has not only used much of his spare time on developing and continuously improving the simulator but also in supervising this PhD-project. I feel honored to have been entrusted a project so dear to his heart and I hope to have lived up to expectations and the high standards of the project.

A special mentioning is needed for my co-supervisors Drs. Per Cayé-Thomasen and Lars Konge. I have been lucky to draw on Per's extensive expertise in temporal bone surgery and many years of teaching temporal bone surgery at the highest international level and Lars' great knowledge and experience with virtual reality surgical simulation and his ability to hone any research question. Lars also kindly provided me not only with space for the simulators and trials but also with assistance and resources in every possible way at the Simulation Centre at Rigshospitalet.

Many other people deserve to be mentioned at this point for their assistance and for inspiring me during the project including Kristianna Mey, my fellow PhD student with whom I shared office; Søren Foghsgaard, Susan Jacobsen and all other otologists at our department; our head of department Mads Klokker and professor Christian von Buchwald; all the PhD students and senior researchers at both the Department of Otorhinolaryngology—Head & Neck Surgery and at the Copenhagen Academy for Medical Education and Simulation (CAMES); Sebastian Roed Rasmussen and Niels West who I supervised for their bachelor's and master's thesis, respectively; Jacob Melchiors, Kristine Aabenhus, Daniel Sieber, Leizl Nayahangan, Berit Sachmann, Aase Svenningsen and many more. I would also like to thank professor Teodor Grantcharov and his amazing team at the Center for Surgical Education and Innovation at St. Michael's Hospital, University of Toronto, Canada, for an inspirational study stay.

I also need to thank all the people who volunteered for participation in my studies: medical students from the Faculty of Health and Medical Sciences at the University of Copenhagen, and my fellow colleagues and residents in otorhinolaryngology. I am thankful for the time they took to participate and for their commitment and flexibility even when things did not go according to plan. Finally, I would like to thank my family and friends for their support.

I feel greatly honored for the opportunity to work on the Visible Ear Simulator project where new ideas and evidence are instantly implemented, and where new knowledge and best practice iteratively gained, keep improving the simulator and thereby temporal bone surgical education. We are at the forefront and in the future, virtual reality simulation will not only provide high quality training for surgical novices but also be a useful tool for established surgeons and experts.

SUMMARY

Virtual reality (VR) simulation-based training is increasingly used in surgical technical skills training including in temporal bone surgery. The potential of VR simulation in enabling high quality surgical training is great and VR simulation allows high-stakes and complex procedures such as mastoidectomy to be trained repeatedly, independent of patients and surgical tutors, outside traditional learning environments such as the OR or the temporal bone lab, and with fewer of the constraints of traditional training.

This thesis aims to increase the evidence-base of VR simulation training of mastoidectomy and, by studying the final-product performances of novices, investigates the transfer of skills to the current gold-standard training modality of cadaveric dissection, the effect of different practice conditions and simulator-integrated tutoring on performance and retention of skills, and the role of directed, self-regulated learning.

Technical skills in mastoidectomy were transferable from the VR simulation environment to cadaveric dissection with significant improvement in performance after directed, self-regulated training in the VR temporal bone simulator. Distributed practice led to a better learning outcome and more consolidated skills than massed practice and also resulted in a more consistent performance after three months of non-practice. Simulator-integrated tutoring accelerated the initial learning curve but also caused over-reliance on tutoring, which resulted in a drop in performance when the simulator-integrated tutor-function was discontinued. The learning curves were highly individual but often plateaued early and at an inadequate level, which related to issues concerning both the procedure and the VR simulator, over-reliance on the tutor function and poor self-assessment skills. Future simulator-integrated automated assessment could potentially resolve some of these issues and provide trainees with both feedback during the procedure and immediate assessment following each procedure. Standard setting by establishing a proficiency level that can be used for mastery learning with deliberate practice could also further sophisticate directed, self-regulated learning in VR simulation-based training. VR simulation-based training should be embedded in a systematic and competency-based training curriculum for high quality surgical skills training, ultimately leading to improved safety and patient care.

REFERENCES

1 Reznick RK. Teaching and testing technical skills. Am J Surg. 1993;165(3):358–361.
2 Canter RJ. Training in surgery: time for a new approach. Clin Otolaryngol. 2009;34(1):90–92.
3 Simo R, Hartley C, Saeed SR, Zarod AP, Taylor PH. Otorhinolaryngological training in Europe—a comparative study. Clin Otolaryngol Allied Sci. 1997;22(4):332–342.
4 Shah RK, Brereton J, Roberson DW. Quality in Otolaryngology: Where We Have Been and Where We Are Going. Otolaryngol Head Neck Surg. 201;152(5):776–777.
5 Kneebone RL, Nestel D, Vincent C, Darzi A. Complexity, risk and simulation in learning procedural skills. Med Educ. 2007;41(8):808–814.
6 Cook DA, Hatala R, Brydges R, et al. Technology-Enhanced Simulation for Health Professions Education: A Systematic Review and Meta-analysis. JAMA. 2011;306(9):978–988.
7 Barry Issenberg S, Mcgaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review*. Med Teach. 2005;27(1):10–28.
8 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.
9 Gardner AK, Scott DJ, Pedowitz RA, et al. Best practices across surgical specialties relating to simulation-based training. Surgery. 2015;158(5):1395–1402.
10 Carr ER, Benjamin E. "Surgical skills box": a new training aid for surgical trainees. J Laryngol Otol. 2006;120(2):133–134.

11 Clifton N, Klingmann C, Khalil H. Teaching Otolaryngology skills through simulation. Eur Arch Otorhinolaryngol. 2011;268(7):949–953.
12 Javia L, Deutsch ES. A Systematic Review of Simulators in Otolaryngology. Otolaryngol Head Neck Surg. 2012;147(6):999–1011.
13 Deutsch ES. Simulation in Otolaryngology: Smart Dummies and More. Otolaryngol Head Neck Surg. 2011;145(6):899–903.
14 Fried MP, Uribe JI, Sadoughi B. The role of virtual reality in surgical training in otorhinolaryngology. Curr Opin Otolaryngol Head Neck Surg. 2007;15(3):163–169.
15 Wiet GJ, Stredney D, Wan D. Training and Simulation in Otolaryngology. Otolaryngol Clin North Am. 2011;44(6):1333–1350.
16 Arora A, Lau LY, Awad Z, Darzi A, Singh A, Tolley N. Virtual reality simulation training in Otolaryngology. Int J Surg. 2014;12(2):87–94.
17 Piromchai P. Virtual Reality Surgical Training in Ear, Nose and Throat Surgery. International Journal of Clinical Medicine. 2014;05(10):558–566.
18 Deutsch ES, Wiet GJ, Seidman M, Hussey HM, Malekzadeh S, Fried MP. Simulation Activity in Otolaryngology Residencies. Otolaryngol Head Neck Surg. 2015;153(2):193–201.
19 George AP, De R. Review of temporal bone dissection teaching: how it was, is and will be. J Laryngol Otol. 2009;124(2):119–125.
20 Mowry SE, Hansen MR. Resident Participation in Cadaveric Temporal Bone Dissection Correlates With Improved Performance on a Standardized Skill Assessment Instrument. Otol Neurotol. 2014;35(1):77–83.
21 Awad Z, Tornari C, Ahmed S, Tolley NS. Construct validity of cadaveric temporal bones for training and assessment in mastoidectomy. Laryngoscope. 2015;125(10):2376–2381.
22 Fennessy BG, O'Sullivan P. Establishing a temporal bone laboratory: considerations for ENT specialist training. Ir J Med Sci. 2009;178(4):393–395.
23 Torgerson C, Brydges R, Chen J, Dubrowski A. Low fidelity simulation of temporal bone drilling leads to improved but suboptimal outcomes. Stud Health Technol Inform. 2007;125:470–472.
24 Mills R, Lee P. Surgical skills training in middle-ear surgery. J Laryngol Otol. 2003;117(3):159–163.
25 Vorwerk U, Begall K. Präparierübungen amvkünstlichen Felsenbein. HNO. 1998;46:246–251.
26 Hochman JB, Kraut J, Kazmerik K, Unger BJ. Generation of a 3D Printed Temporal Bone Model with Internal Fidelity and Validation of the Mechanical Construct. Otolaryngol Head Neck Surg. 2014;150(3):448–454.
27 Rose AS, Kimbell JS, Webster CE, Harrysson OL, Formeister EJ, Buchman CA. Multi-material 3D Models for Temporal Bone Surgical Simulation. Ann Otol Rhinol Laryngol. 2015;124(7):528–536.
28 Bhutta MF. A review of simulation platforms in surgery of the temporal bone. Clin Otolaryngol. 10 October 2015. [Epub ahead of print].
29 Kuppersmith RB, Johnston R, Moreau D, Loftin RB, Jenkins H. Building a virtual reality temporal bone dissection simulator. Stud Health Technol Inform. 1997;39:180–186.
30 Mason TP, Applebaum EL, Rasmussen M, Millman A, Evenhouse R, Panko W. The virtual temporal bone. Stud Health Technol Inform. 1998;50:346–352.
31 Mason TP, Applebaum EL, Rasmussen M, Millman A, Evenhouse R, Panko W. Virtual temporal bone: creation and application of a new computer-based teaching tool. Otolaryngol Head Neck Surg. 2000;122(2):168–173.
32 Wiet GJ, Bryan J, Dodson E, et al. Virtual temporal bone dissection simulation. Stud Health Technol Inform. 2000;70:378–384.
33 Pflesser B, Leuwer R, Tiede U, Höhne KH. Planning and rehearsal of surgical interventions in the volume model. Stud Health Technol Inform. 2000;70:259–264.
34 Pflesser B, Petersik A, Tiede U, Höhne KH, Leuwer R. Volume cutting for virtual petrous bone surgery. Comp Aided Surgery. 2002;7(2):74–83.
35 Wiet G, Stredney D, Sessanna D, et al. Virtual temporal bone dissection: An interactive surgical simulator. Otolaryngol Head Neck Surg. 2002;127(1):79–83.
36 John NW, Thacker N, Pokric M, et al. An integrated simulator for surgery of the petrous bone. Stud Health Technol Inform. 2001;81:218–224.
37 Agus M, Giachetti A, Gobbetti E, et al. Mastoidectomy simulation with combined visual and haptic feedback. Stud Health Technol Inform. 2002;85:17–23.
38 Agus M, Giachetti A, Gobbetti E, et al. A haptic model of a bone-cutting burr. Stud Health Technol Inform. 2003;94:4–10.
39 Neri E, Sellari Franceschini S, Berrettini S, Caramella D, Bartolozzi C. IERAPSI project: Simulation of a canal wall-up mastoidectomy. Comp Aided Surgery. 2006;11(2):99–102.
40 Morris D, Sewell C, Barbagli F, et al. Visuohaptic simulation of bone surgery for training and evaluation. IEEE Comput Grap Appl. 26(6):48–57.
41 O'Leary SJ, Hutchins MA, Stevenson DR, et al. Validation of a Networked Virtual Reality Simulation of Temporal Bone Surgery. Laryngoscope. 2008;118(6):1040–1046.
42 Kockro RA, Hwang PY. Virtual Temporal Bone. Neurosurgery. 2009;64(5 Suppl 2):216–229.
43 Wiet G, Schmalbrock P, Powell K, Stredney D. Use of Ultra-High-Resolution Data for Temporal Bone Dissection Simulation. Otolaryngol Head Neck Surg. 2005;133(6):911–915.
44 Wong D, Unger B, Kraut J, et al. Comparison of cadaveric and isomorphic virtual haptic simulation in temporal bone training. J Otolaryngol Head Neck Surg. 2014;43(1):31.
45 Trier P, Noe KØ, Sørensen MS, Mosegaard J. The visible ear surgery simulator. Stud Health Technol Inform. 2008;132:523–525.
46 Sørensen MS, Mosegaard J, Trier P. The visible ear simulator: a public PC application for GPU-accelerated haptic 3D simulation of ear surgery based on the visible ear data. Otol Neurotol. 2009;30(4):484–487.
47 Khemani S, Arora A, Singh A, Tolley N, Darzi A. Objective Skills Assessment and Construct Validation of a Virtual Reality Temporal Bone Simulator. Otol Neurotol. 2012;33(7):1225–1231.
48 Zirkle M, Roberson DW, Leuwer R, Dubrowski A. Using a Virtual Reality Temporal Bone Simulator to Assess Otolaryngology Trainees. Laryngoscope. 2007;117(2):258–263.
49 Sewell C, Morris D, Blevins N, Barbagli F, Salisbury K. Achieving proper exposure in surgical simulation. Stud Health Technol Inform. 2006;119:497–502.
50 Zhao YC, Kennedy G, Hall R, O'Leary S. Differentiating levels of surgical experience on a virtual reality temporal bone simulator. Otolaryngol Head Neck Surg. 2010;143(5):S30–S35.
51 Sewell C, Morris D, Blevins NH, et al. Validating metrics for a mastoidectomy simulator. Stud Health Technol Inform. 2007;125:421–426.
52 Sewell C, Morris D, Blevins N, et al. Providing metrics and performance feedback in a surgical simulator. Comp Aided Surgery. 2008;13(2):63–81.
53 Arora A, Hall A, Kotecha J, et al. Virtual reality simulation training in temporal bone surgery. Clin Otolaryngol. 2015;40(2):153–159.
54 Kerwin T, Wiet G, Stredney D, Shen H-W. Automatic scoring of virtual mastoidectomies using expert examples. Int J CARS. 2011;7(1):1–11.
55 Kerwin T, Stredney D, Wiet G, Shen H-W. Virtual mastoidectomy performance evaluation through multi-volume analysis. Int J CARS. 2012;8(1):51–61.
56 Wiet G, Hittle B, Kerwin T, Stredney D. Translating surgical metrics into automated assessments. Stud Health Technol Inform. 2012;173:543–548.
57 Francis HW, Malik MU, Diaz Voss Varela DA, et al. Technical skills improve after practice on virtual-reality temporal bone simulator. Laryngoscope. 2012;122(6):1385–1391.
58 Nash R, Sykes R, Majithia A, Arora A, Singh A, Khemani S. Objective assessment of learning curves for the Voxel-Man TempoSurg temporal bone surgery computer simulator. J Laryngol Otol. 2012;126(07):663–669.
59 Zhao YC, Kennedy G, Yukawa K, Pyman B, Stephen O'Leary. Can Virtual reality simulator be used as a training aid to improve cadaver temporal bone dissection? Results of a randomized blinded control trial. Laryngoscope. 2011;121(4):831–837.
60 Zhao YC, Kennedy G, Yukawa K, Pyman B, O'Leary S. Improving temporal bone dissection using self-directed virtual reality simulation: results of a randomized blinded control trial. Otolaryngol Head Neck Surg. 2011;144(3):357–364.
61 Wiet GJ, Rastatter JC, Bapna S, et al. Training Otologic Surgical Skills Through Simulation—Moving Toward Validation: A Pilot Study and Lessons Learned. J Grad Med Educ. 2009;1(1):61–66.
62 Wiet GJ, Stredney D, Kerwin T, et al. Virtual temporal bone dissection system: OSU virtual temporal bone system. Laryngoscope. 2012;122(S1):S1–S12.
63 Szalay D, MacRae H, Regehr G, Reznick R. Using operative outcome to assess technical skill. Am J Surg. 2000;180(3):234–237.
64 Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. Acad Med. 2013;88(6):872–883.
65 Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273–278.
66 Cook DA, Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. Am J Med. 2006;119(2):166.e7–16.
67 Awad Z, Hayden L, Robson AK, Muthuswamy K, Tolley NS. Reliability and validity of procedure-based assessments in otolaryngology training. Laryngoscope. 2015;125(6):1328-1335
68 Zirkle M, Taplin MA, Anthony R, Dubrowski A. Objective Assessment of Temporal Bone Drilling Skills. Ann Otol Rhinol Laryngol. 2007;116(11):793–798.
69 Laeeq K, Bhatti NI, Carey JP, et al. Pilot testing of an assessment tool for competency in mastoidectomy. Laryngoscope. 2009;119(12):2402–2410.
70 Francis HW, Masood H, Chaudhry KN, et al. Objective Assessment of Mastoidectomy Skills in the Operating Room. Otol Neurotol. 2010;31(5):759–765.
71 Wan D, Wiet GJ, Welling DB, Kerwin T, Stredney D. Creating a cross-institutional grading scale for temporal bone dissection. Laryngoscope. 2010;120(7):1422–1427.
72 Torgerson CS, Brydges R, Chen JM, Dubrowski A. Drilling Simulated Temporal Bones with Lef-Handed Tools: A Left-Hander's Right? Ann Otol Rhinol Laryngol. 2007;116(11):819–826.
73 Linke R, Leichtle A, Sheikh F, et al. Assessment of skills using a virtual reality temporal bone surgery simulator. Acta Otorhinolaryngol Ital. 2013;33(4):273–281.
74 Wijewickrema S, Piromchai P, Zhou Y, et al. Developing Effective Automated Feedback in Temporal Bone Surgery Simulation. Otolaryngol Head Neck Surg. 2015;152(6):1082–1088.

75 Butler NN, Wiet GJ. Reliability of the Welling Scale (WS1) for Rating Temporal Bone Dissection Performance. Laryngoscope. 2007;117(10):1803–1808.
76 Fernandez SA, Wiet GJ, Butler NN, Welling B, Jarjoura D. Reliability of Surgical Skills Scores in Otolaryngology Residents: Analysis Using Generalizability Theory. Eval Health Prof. 2008;31(4):419–436.
77 Cook DA, Hamstra SJ, Brydges R, et al. Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. Med Teach. 2013;35(1):e867–e898.
78 Pusic MV, Boutis K, Hatala R, Cook DA. Learning Curves in Health Professions Education. Acad Med. 2015;90(8):1034–1042.
79 Francis HW, Masood H, Laeeq K, Bhatti NI. Defining milestones toward competency in mastoidectomy using a skills assessment paradigm. Laryngoscope. 2010;120(7):1417–1421.
80 Carr MM. Program directors' opinions about surgical competency in otolaryngology residents. Laryngoscope. 2005;115(7):1208–1211.
81 Eversbusch A, Grantcharov TP. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. Surg Endosc. 2004;18(10):1514–1518.
82 Grantcharov TP, Funch-Jensen P. Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. Am J Surg. 2009;197(4):447–449.
83 Schijven MP, Jakimowicz J. The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance. Surg Endosc. 2004;18(1):121–127.
84 Louridas M, Grantcharov TP. Unpublished data. 2015.
85 Ericsson KA. Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. Acad Med. 2004;79(Supplement):S70–S81.
86 Causer J, Barach P, Williams AM. Expertise in medicine: using the expert performance approach to improve simulation training. Med Educ. 2014;48(2):115–123.
87 Brunner WC, Korndorffer JR Jr., Sierra R, et al. Laparoscopic virtual reality training: Are 30 repetitions enough? J Surg Res. 2004;122(2):150–156.
88 Stefanidis D, Scerbo MW, Sechrist C, Mostafavi A, Heniford BT. Do novices display automaticity during simulator training? Am J Surg. 2008;195(2):210–213.
89 Andersen SA, Mikkelsen PT, Konge L, Caye-Thomasen P, Sørensen MS. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. Laryngoscope. 2016;126(2):E74–79.
90 Malik MU, Diaz Voss Varela DA, Park E, et al. Determinants of resident competence in mastoidectomy: Role of interest and deliberate practice. Laryngoscope. 2013;123(12):3162–3167.
91 Bhatti NI, Ahmed A. Improving skills development in residency using a deliberate-practice and learner-centered model. Laryngoscope. 2015;125 Suppl 8:S1–14.
92 Moulton CA, Dubrowski A, MacRae H, et al. Teaching Surgical Skills: What Kind of Practice Makes Perfect? Ann Surg. 2006;244(3):400–409.
93 Arthur W Jr, Bennett W Jr, Stanush PL, McNelly TL. Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. Human Performance. 1998;11(1):57–101.
94 Maagaard M, Sorensen JL, Oestergaard J, et al. Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer. Surg Endosc. 2010;25(3):722–727.
95 Bjerrum F, Maagaard M, Led Sorensen J, et al. Effect of Instructor Feedback on Skills Retention After Laparoscopic Simulator Training: Follow-Up of a Randomized Trial. J Surg Educ. 2015;72(1):53–60.
96 Sinha P, Hogle NJ, Fowler DL. Do the laparoscopic skills of trainees deteriorate over time? Surg Endosc. 2008;22(9):2018–2025.
97 Stefanidis D, Korndorffer JR Jr., Sierra R, Touchard C, Dunne JB, Scott DJ. Skill retention following proficiency-based laparoscopic simulator training. Surgery. 2005;138(2):165–170.
98 Malekzadeh S, Malloy KM, Chu EE, Tompkins J, Battista A, Deutsch ES. ORL emergencies boot camp: Using simulation to onboard residents. Laryngoscope. 2011;121(10):2114–2121.
99 Chin CJ, Chin CA, Roth K, Rotenberg BW, Fung K. Simulation-based otolaryngology - head and neck surgery boot camp: "how I do it." J Laryngol Otol. 2016;130(3):284–290.
100 Shea CH, Lai Q, Black C, Park J. Spacing practice sessions across days benefits the learning of motor skills. Human Movement Science. 2000;19(5):737–760.
101 Mackay S, Morgan P, Datta V, Chang A, Darzi A. Practice distribution in procedural skills training. Surg Endosc. 2002;16(6):957–961.
102 Benjamin AS, Tullis J. What makes distributed practice effective? Cogn Psychol. 2010;61(3):228–247.
103 Kesser BW, Hallman M, Murphy L, et al. Interval vs Massed Training: How Best Do We Teach Surgery? Otolaryngol Head Neck Surg. 2013;150(1):61–67.
104 Stefanidis D, Walters KC, Mostafavi A, Heniford BT. What is the ideal interval between training sessions during proficiency-based laparoscopic simulator training? Am J Surg. 2009;197(1):126–129.
105 Shea CH, Wulf G, Whitacre C. Enhancing Training Efficiency and Effectiveness Through the Use of Dyad Training. J Mot Behav. 1999;31(2):119–125.
106 Räder SB, Henriksen A, Butrymovich V, et al. A Study of the Effect of Dyad Practice Versus That of Individual Practice on Simulation-Based Complex Skills Learning and of Students' Perceptions of How and Why Dyad Practice Contributes to Learning. Acad Med. 2014;89(9):1287–1294.
107 Hatala R, Cook DA, Zendejas B, Hamstra SJ, Brydges R. Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. Adv in Health Sci Educ. 2013;19(2):251–272.
108 Loyens SM, Magda J, Rikers RM. Self-Directed Learning in Problem-Based Learning and its Relationships with Self-Regulated Learning. Educ Psychol Rev. 2008;20(4):411–427.
109 Kirschner PA, Sweller J, Clark RE. Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. Educational Psychologist. 2006;41(2):75–86.
110 Brydges R, Nair P, Ma I, Shanks D, Hatala R. Directed self-regulated learning versus instructor-regulated learning in simulation training. Med Educ. 2012;46(7):648–656.
111 Brydges R, Manzone J, Shanks D, et al. Self-regulated learning in simulation-based training: a systematic review and meta-analysis. Med Educ. 2015;49(4):368–378.
112 Jowett N, LeBlanc V, Xeroulis G, MacRae H, Dubrowski A. Surgical skill acquisition with self-directed practice using computer-based video training. Am J Surg. 2007;193(2):237–242.
113 Wright AS, McKenzie J, Tsigonis A, et al. A structured self-directed basic skills curriculum results in improved technical performance in the absence of expert faculty teaching. Surgery. 2012;151(6):808–814.
114 Andersen SA, Caye-Thomasen P, Sorensen MS. Mastoidectomy performance assessment of virtual simulation training using final-product analysis. Laryngoscope. 2014;125(2):431–435.
115 University of Copenhagen. Studenterbestand Ved Københavns Universitet. http://us.ku.dk/studiestatistik/studiestatistikker/bestand/KU_BESTAND_2015.pdf. Accessed 3 March 2016.
116 Konge L, Bjerrum F, Nayahangan LJ, Schroeder TV. Developing and running a surgical simulation centre: experiences from Copenhagen, Denmark. J Surg Simul. 2015;2:47–52.
117 The Visible Ear Simulator. http://ves.alexandra.dk. Accessed 3 March 2016.
118 Sørensen MS, Dobrzeniecki AB, Larsen P. The visible ear: a digital image library of the temporal bone. ORL J Otorhinolaryngol Relat Spec. 2002;64:378–381.
119 Andersen SA. Cross-platform digital assessment forms for performance rating of surgical skills. J Educ Eval Health Prof. 2015;12:13–14.
120 Sewell C, Morris D, Blevins NH, Barbagli F, Salisbury K. Evaluating drilling and suctioning technique in a mastoidectomy simulator. Stud Health Technol Inform. 2007;125:427–432.
121 Andersen SA, Cayé-Thomasen P, Sørensen MS. Novices perform better in virtual reality simulation than in traditional cadaveric dissection training of mastoidectomy. J Surg Simul. 2015;2:68–75.
122 Sweller J. Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science. 2010;12(2):257–285.
123 Brydges R, Dubrowski A, Regehr G. A New Concept of Unsupervised Learning: Directed Self-Guided Learning in the Health Professions. Acad Med. 2010;85:S49–S55.
124 Andersen SA, Konge L, Caye-Thomasen P, Sørensen MS. Learning Curves of Virtual Mastoidectomy in Distributed and Massed Practice. JAMA Otolaryngol Head Neck Surg. 2015;141(10):913–918.
125 Butler DL, Brydges R. Learning in the health professions: what does self-regulation have to do with it? Med Educ. 2013;47(11):1057–1059.
126 Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA. Ceiling effect in technical skills of surgical residents. Am J Surg. 2004;188(3):294–300.
127 West N, Konge L, Cayé-Thomasen P, Sørensen MS, Andersen SA. Peak and ceiling effects in final-product analysis of mastoidectomy performance. J Laryngol Otol. 2015;129(11):1091–1096.
128 Deutsch ES, Orioles A, Kreicher K, Malloy KM, Rodgers DL. A qualitative analysis of faculty motivation to participate in otolaryngology simulation boot camps. Laryngoscope. 2013;123(4):890–897.
129 Malekzadeh S, Deutsch ES, Malloy KM. Simulation-based otorhinolaryngology emergencies boot camp: Part 2: Special skills using task trainers. Laryngoscope. 2014;124(7):1566–1569.
130 Malloy KM, Malekzadeh S, Deutsch ES. Simulation-based otorhinolaryngology emergencies boot camp: Part 1: Curriculum design and airway skills. Laryngoscope. 2014;124(7):1562–1565.
131 Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. Am J Surg. 2003;185(2):146–149.
132 Uribe JI, Ralph WM, Glaser AY, Fried MP. Learning curves, acquisition, and retention of skills trained with the endoscopic sinus surgery simulator. Am J Rhinol. 2004;18(2):87–92.
133 MacDonald J, Williams RG, Rogers DA. Self-assessment in simulation-based surgical skills training. Am J Surg. 2003;185(4):319–322.
134 Brydges R, Carnahan H, Safir O, Dubrowski A. How effective is self-guided learning of clinical technical skills? It's all about process. Med Educ. 2009;43(6):507–515.
135 Bhatti NI, Cummings CW. Viewpoint: Competency in Surgical Residency Training: Defining and Raising the Bar. Acad Med. 2007;82(6):569–573.

136 Mcgaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. Med Educ. 2014;48(4):375–385.
137 Bolsin S, Colson M. The use of the Cusum technique in the assessment of trainee competence in new procedures. Int J Qual Health Care. 2000;12(5):433–438.
138 Biau DJ, Williams SM, Schlup MM, Nizard RS, Porcher R. Quantitative and individualized assessment of the learning curve using LC-CUSUM. Br J Surg. 2008;95(7):925–929.
139 Maguire T, Mayne CJ, Terry T, Tincello DG. Analysis of the surgical learning curve using the cumulative sum (CUSUM) method. Neurourol Urodyn. 2013;32(7):964–967.

APPENDIX I: THE MODIFIED WELLING SCALE

Grade each item:
0 = incomplete/inadequate dissection
1 = complete/adequate dissection.

Mastoidectomy margins defined at		
1. Temporal line	0	1
2. Posterior canal wall	0	1
3. Sigmoid sinus	0	1
Antrum mastoideum		
4. Antrum entered	0	1
5. Lateral semicircular canal exposed	0	1
6. Lateral semicircular canal intact	0	1
Sigmoid sinus		
7. Exposed, no overhang	0	1
8. No cells remain	0	1
9. No holes	0	1
Sinodural angle		
10. Sharp	0	1
11. No cells remain	0	1
Tegmen mastoideum/tympani		
12. Attic/tegmen tympany exposed	0	1
13. Ossicles intact (untouched)	0	1
14. Tegmen mastoideum exposed	0	1
15. No cells remain	0	1
16. No holes	0	1
Mastoid tip		
17. Digastric ridge exposed	0	1
18. Digastric ridge followed towards stylomastoid foramen	0	1
19. No cells remain	0	1
External auditory canal		
20. Thinning of the posterior canal wall	0	1
21. No cells remain	0	1
22. No holes	0	1
Facial nerve		
23. Facial nerve identified (vertical part)	0	1
24. No exposed nerve sheath	0	1
25. Tympanic chorda exposed	0	1
Posterior tympanotomy		
26. Facial recess completely exposed	0	1