UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCE

PhD thesis

Andreas Frithioff, MD

3D-printed models for training temporal bone surgery

Principal supervisor: Professor Mads Sølvsten Sørensen, MD, DMSc

This thesis has been submitted to the Graduate School of Health and Medical Science,
University of Copenhagen, April 4, 2023

3D-printed models for training temporal bone surgery

PhD thesis by Andreas Frithioff, MD

Copenhagen Hearing and Balance Center, Department of Otorhinolaryngology—Head & Neck Surgery and Audiology, Rigshospitalet, Copenhagen, Denmark.

Academic supervisors

Principal supervisor:

Professor Mads Sølvsten Sørensen, MD, DMSc

Copenhagen Hearing and Balance Center, Department of Otorhinolaryngology—Head & Neck Surgery and Audiology, Rigshospitalet, Copenhagen, Denmark.

Primary co-supervisor:

David Bue Pedersen, MSc, PhD

Department of Civil and Mechanical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark.

Co-supervisors:

Steven Arild Wuyts Andersen, MD, PhD

Copenhagen Hearing and Balance Center, Department of Otorhinolaryngology—Head & Neck Surgery and Audiology, Rigshospitalet, Copenhagen, Denmark.

Martin Frendø, MD, PhD

Copenhagen Hearing and Balance Center, Department of Otorhinolaryngology—Head & Neck Surgery and Audiology, Rigshospitalet, Copenhagen, Denmark,

Copenhagen Academy for Medical Education and Simulation (CAMES), Center for HR & Education, Region H, Copenhagen, Denmark.

Department of Plastic Surgery, Herlev & Gentofte Hospital, Copenhagen, Denmark

Assessment Committee

Associate professor Sune Land Bloch, MD, PhD (Chairperson)

Copenhagen Hearing and Balance Center, Department of Otorhinolaryngology—Head & Neck Surgery and Audiology, Rigshospitalet, Copenhagen, Denmark

Department of Clinical Medicine, University of Copenhagen, Denmark

Associate professor Rune Dall Jensen, MSc, PhD

Department of Clinical Medicine, Aarhus University, Denmark

Professor Robert Mlynski, MD, DMSc

Department of Otorhinolaryngology, Head and Neck Surgery "Otto Körner", Rostock University Medical Center, Germany

Table of Contents

Abbreviations

3D: Three-dimensional

AAO-HNSF: American Academy of Otolaryngology-Head and Neck Surgery Foundation

ABS: Acrylonitrile Butadiene Styrene

AM: Additive Manufacturing

BAHS: Bone Anchored Hearing Systems

CAD: Computer-aided Design

CBCT: Cone Beam Computerized Tomography

CI: Cochlear Implantation

CT: Computed Tomography

FPA: Final Product Analysis/Assessment

GRS: Global Rating Scale

MERSQI: Medical Education Research Study Quality Instrument

OR: Operating Room

ORL: Otorhinolaryngology

PLA: Polylactic Acid

SBT: Simulation-based Training

STL: Standard Tessellation Language

TBC: Task-Based Checklists

USD: United States Dollars

UV: Ultra Violet

VES: Visible Ear Simulator

VR: Virtual Reality

WHO: World Health Organization

Acknowledgments

Both professionally and personally, I have genuinely enjoyed every aspect of the research for this thesis and I am very grateful to have had the opportunity to work with so many talented people. I have received excellent guidance during my research and I would like to express my deep and sincere gratitude, especially to my four supervisors:

Professor, MD, DMSc, Mads Sølvsten Sørensen who has not only given me tremendous support during this project but also a unique introduction to the field of ear surgery. Mads' attention to any detail in both research and during surgery is something I will carry with me for the rest of my career.

MD, PhD Steven A.W. Andersen who, already during my bachelor thesis, ignited my interest in temporal bone research. I cannot underestimate the impact that Steven have had on all aspects of my research and I'm truly thankful for what I have learned so far.

PhD, David Bue Pedersen for his deep knowledge and enthusiastic introduction to the world of 3D-printing, together with ambitious visions for the technical and academic work.

MD, PhD Martin Frendø with whom I shared office with during the first year of conducting my PhD project. Martin has supported me in all parts of my research and provided me with fantastic guidance. Martin has a unique ability to motivate people around him and it has been a true joy working with him.

Next, I would like to thank Kenneth Weiss, whose technical knowhow and help have been outstanding. Without Kenneth’s help printing models during COVID-lockdown, it would not have been possible to conduct Study IV during the 2021 national temporal bone course. Also, a special thanks to PhD-fellow, computer scientist, Peter Trier Mikkelsen for help creating the printable file.

I thank the PhD-fellows at CAMES with whom I have shared many good hours. We have had many academic discussions but also a lot of fun.

A sincere thank you to all residents and consultant otosurgeons who participated in our studies. Without their eager willingness to help, this type of research would not be possible.

Finally, I wholeheartedly thank my parents, Niels and Birgitte, and not least my partner, Nanna, for relentless support and encouragement which has been invaluable during my work.

Summary in English

Mastoidectomy is a fundamental but important procedure in ear surgery. Performing the procedure requires good surgical skills as drilling takes place in close relation to the facial nerve, controlling facial expressions, the dura, the membrane surrounding the brain, the vestibular organs, controlling the sense of equilibrium, ossicles, and larger vessels of the brain. Consequently, excellent training in the procedure is required.

Cadavers have been the gold-standard for simulation-based training for many years. This moves initial training away from the operating room and offers a risk-free alternative to operating on real patients. However, a decline in availability of cadavers suitable for temporal bone training has created interest for other training alternatives. Cadaver training is typically organized as formalized dissection courses, rarely available locally at the departments, limiting access to training. Virtual Reality (VR) simulation addresses these problems and is highly beneficial for accelerating trainees' mastoidectomy skills. Moreover, VR simulation enables competency-based training where trainees can progress in training only when demonstrating the ability to perform a certain task. Nevertheless, VR training lacks essential physical aspects of the procedure such as the use of a microscope and otosurgical drill handling. For this training, 3D-printed models are recognized to hold great potential as the interaction closely mimics that of cadaveric bones. However, little is currently known about the educational value of 3D-printed models, how to manufacture them most effectively, and how to best implement this alternative training modality in the temporal bone training curriculum.

This thesis aims 1) to map the existing knowledge on manufacturing 3D-printed temporal bone models and the current educational evidence supporting their use for training, 2) to create a cost-effective 3D-printed model that can be manufactured locally at clinical training departments, 3) to

collect educational validity evidence for using the model for training purposes according to a contemporary validity framework, 4) to evaluate how surgical skills acquired on the model transfer to cadaver dissection and 5) to evaluate the reliability of video-recorded assessment of mastoidectomy compared with physical assessment, opening the possibility of remote assessment and decentralized training.

First, we conducted a systematic review of the use of 3D-printed models for the training of temporal bone surgery. We found that it is feasible to 3D-print temporal bone models, albeit a variety of material and print technologies are used and there is currently no consensus on how to manufacture these models most effectively. Altogether, considerations on technical knowhow, time available, and price are important when choosing how to print a training model. We found that rather than assessing if training using 3D-printed models improved surgical performance, most studies only evaluated their models using expert and/or trainee opinion (e.g., whether the users liked the model).

We then designed a new 3D-printed temporal bone model for training purposes. We created a printable file from a digitized human temporal bone and printed the model using an inexpensive consumer-grade printer. We applied some minor modifications to the printer installing a direct drive and ruby nozzle. These modifications increased the number of successful prints and enhanced the durability of the printer. The model was printed using a unique plastic filament with a high load of chalk for mimicking bone properties. After printing the model, the facial nerve, dura, and sigmoid sinus (blood vessel) were represented by manually inserting a yellow 1 mm wire (to represent the facial nerve), and colored latex layers (to represent the dura and sigmoid sinus). Eleven ORL trainees attending a temporal bone course found the model to be a suitable training tool for learning the mastoidectomy procedure.

In the third study, we systematically collected validity evidence for using the 3D-printed model for mastoidectomy training according to a contemporary validity framework (i.e., Messick's framework). We used drillings from trainees and expert otosurgeons to establish a pass/fail score of 21 out of 25 points on a modified Welling Scale (a well-established assessment tool for the mastoidectomy procedure). This score can be used for competency-based training and guide training progression e.g., when a trainee is ready to progress to cadaver training. Using generalizability theory (G theory), we found assessments of the 3D-printed model to be highly reliable. Subsequent decisions (D) studies established that one rater assessing two performances or two raters assessing one performance would be sufficient for a reliability level suited for high-stakes assessment.

In the fourth study, we investigated whether mastoidectomy skills training with 3D-printed models transferred to performance on a human cadaver. ORL residents received three hours of self-directed mastoidectomy training on the 3D-printed model and then performed a mastoidectomy on a human cadaver. The cadaver performances were then compared to the performances of residents who completed similar amounts of VR SBT in a previous study. Cadaver mastoidectomy performances substantially improved after training with 3D-printed models, suggesting that skills do transfer to cadaver surgery.

Lastly, in the fifth study, we compared the reliability of assessment using video-recordings with direct physical assessment. We found that the reliability of assessment using video-recordings was similar to that of physical rating. This supports that reliable assessment in a decentralized setting is feasible as video recordings can easily and instantly be shared for assessment rather than physical specimens (drilled temporal bone models) being transported to the raters.

Altogether, the studies in this thesis provide evidence for using 3D-printed models for temporal bone training and offer key knowledge for training departments implementing such models in their surgical training curricula.

Summary in Danish

Mastoidektomi er en fundamental men vigtig procedure i ørekirurgi. Ved proceduren benyttes et operationsmikroskop og en del af tindingebenet udbores ved brug af otokirurgiske bor. Boringen foregår i et meget komplekst anatomisk område og man borer i tæt relation til ansigtsnerven, som styrer ansigtsmotorik, hjernehinden (dura), øreknoglerne, balanceorganerne, og større kar som forsyner hjernen. For ikke at kompromittere patientsikkerheden, er det derfor nødvendigt med god kirurgisk træning forud for superviseret kirurgi.

Tidligere har brug af kadavere været guldstandarden til at træne proceduren. I modsætning til kirurgi på rigtige patienter, udgør kadavertræning et risikofrit læringsmiljø til gavn for både uddannelsessøgende og patienter. Desværre er tilgængeligheden af kadavertræning faldet og man er derfor begyndt at interessere sig for alternativer. Ydermere er kadavertræning ofte indrettet som dissektionskurser. Sådanne kurser løber over få dage med meget træning på kort tid. Dette har vist sig at føre til dårligere læring sammenholdt med at sprede den samme mængde træning ud over længere tid. Virtual Reality (VR) simulationstræning kan adressere nogle af disse problemer og har vist sig at være en effektiv måde at udvikle uddannelsessøgendes kompetencer. I øvrigt kan VR-simulationstræning understøtte kompetencebaseret læring, hvor uddannelsessøgende først avancerer i deres træning når de har opnået et prædefineret kompetenceniveau. VR-simulationstræning mangler dog mange fysiske aspekter af proceduren, inklusiv brugen af bor og operationsmikroskop. Til dette formål forventes det at 3D-printede tindingebensmodeller kan have et stort potentiale. På trods af de åbenlyse fordele der kunne være ved at bruge sådanne modeller, mangler der fortsat viden om den læringsmæssige værdi, hvordan man bedst printer modellerne og hvordan de bedst implementeres i de eksisterende træningsprogrammer.

Denne afhandling har til formål 1) at kortlægge den eksisterende viden indenfor produktionen af 3D-printede tindingebensmodeller, samt viden om hvilken læringseffekt der er ved at træne på disse modeller. 2) At skabe en omkostningseffektiv 3D-printet tindingebensmodel som kan produceres lokalt på uddannelsesafdelingerne. 3) Systematisk at samle validitetsevidens for brugen af disse modeller til træning ved hjælp af moderne validitetsværktøjer. 4) At undersøge hvordan kirurgiske færdigheder tilegnet ved brugen af 3D-printede modeller overføres til et mere realistiske træningsalternativ, nemlig kadaverdissektion og 5) at evaluere pålideligheden af video-baseret vurderinger af mastoidektomi, dette for at åbne muligheden for fjernvurdering og decentraliseret træning.

Først udførte vi en systematisk gennemgang af den eksisterende litteratur omhandlende brugen af 3D-printede modeller til tindingebenstræning. Her fandt vi, at det er muligt at 3D-printe modeller til træningsbrug ved brug af mange forskellige printteknologier og materialer. På trods af mange rapporter og brugen af flere forskellige teknologier og materialer, fandtes der ingen konsensus om hvordan man bedst 3D-printer en tindingebensmodel. Alt i alt synes både tid, teknisk viden og pris at være vigtige elementer i overvejelserne når man skal beslutte hvordan man vil 3D-printe en tindingebensmodel. Ydermere fandt vi, at de fleste eksisterende studier ikke evaluerede om træning på modellerne forbedrede kirurgiske færdigheder, men udelukkende undersøgte eksperters og uddannelsessøgendes holdninger til simulationsbaseret træning med modellerne.

I andet studie designede vi vores egen 3D-printede model til tindingebenstræning. Vi brugte et digitaliseret humant tindingeben til at skabe en printbar fil. Denne fil printede vi ved brug af en bruger- og prisvenlig 3D-printer. For at hæve succesraten af print, samt sænke behovet for vedligehold, installerede vi et direct-drive og en rubin-dyse på printeren. For at efterligne knogle, brugte vi et unikt plastik filament med et højt indhold af kalk til at printe modellen. Efter modellen blev printet, var det nødvendigt at imitere ansigtsnerven, dura og det store kar sinus sigmoideus. Dette gjorde vi ved hjælp af en gul ledning samt et farvet silikonelag. Modellen blev evalueret til et tindingebenskursus, og her rapporterede de uddannelsessøgende at modellen, efter deres opfattelse fremstod som et godt træningsredskab forud for kadaverdissektion.

I det tredje studie, samlede vi systematisk validitetsevidens efter et moderne validitetsframework (Messick's framework) for brugen af den 3D-printede model til tindingebenstræning. Ved brug af boringer fra uddannelsessøgende og erfarne kirurger, etablerede vi en beståelsesgrænse på 21 ud af 25 point på en modificeret udgave af Welling Scale; et evalueringsværktøj som bliver brugt til at vurdere mastoidektomi-præstationen. Denne score kan bruges af uddannelsessøgende som et led i kompetence-baseret træning af tindingebenskirurgi f.eks. hvornår en uddannelsessøgende er klar til at gå videre til kadaver træning. I studiet brugte vi G teori til at vurdere reliabiliteten af vurderingerne og fandt generelt at der var et højt niveau af reliabilitet. Ved hjælp af D studier, fandt vi at én bedømmer skal vurdere to boringer, eller at to bedømmere skal vurdere én boring, for at opnå det reliabilitets-niveau der er krævet for bedømmelser til brug af f.eks. certificering.

I det fjerde studie undersøgte vi om mastoidektomi-evner tilegnet ved brug af 3D-printede modeller kunne overføres til kadaverdissektion. En gruppe uddannelsessøgende modtog 3 timers selvstændig træning på den 3D-printede model før boring af en mastoidektomi på et kadaver. Kadaver-præstationerne blev herefter sammenlignet med uddannelsessøgende kirurger fra et tidligere studie. Denne gruppe modtog VR-træning forud for kadaverdissektionen i stedet for 3D-printede modeller. Vi fandt at de uddannelsessøgende som havde trænet på den 3D-printede model, klarede sig bedre end den gruppe som havde modtaget VR-træning. Dette viser at mastoidektomi-træning på 3D-printede modeller har en positiv effekt på den kirurgiske præstation på kadavere.

Til sidst, i det femte studie, undersøgte vi hvordan reliabiliteten bliver påvirket, når man bruger videooptagelser i stedet for fysiske vurderinger af boringsresultatet. Her fandt vi, at fjernvurdering ved brugen af videooptagelser har samme niveau af reliabilitet som ved fysiske vurderinger. Dette betyder at den uddannelsessøgende kan træne decentralt uden tilstedeværelsen af en supervisor, men stadig modtage troværdig feedback på sin kirurgiske præstation. Dette understøtter den kompetence-baserede træning.

Afhandlingens fund understøtter brugen af 3D-printede modeller i tindingebenstræning. Ydermere, kan dette arbejde hjælpe uddannelsesafdelinger i flere trin af implementeringen af disse modeller i deres træningsprogrammer.

List of publications

I. Frithioff A, Frendø M, Pedersen DB, Sørensen MS, Andersen SAW. 3D-printed models for temporal bone training: A systematic review. *Otolaryngol Head Neck Surg. 2021;165(5):617-625.*

II. Frithioff A, Weiss K, Frendø M, Seen P, Mikkelsen PT, Sieber D, Sørensen MS, Pedersen DB, Andersen SAW. 3D-printing a cost-effective model for mastoidectomy training – *Accepted for publication in 3D-printing in Medicine*

III. Frithioff A, Frendø M, Weiss K, Foghsgaard S, Pedersen DB, Frederiksen TW, Sørensen MS, Andersen SAW. 3D-printed models for temporal bone training: A validity study – *Submitted to Otology and Neurotology*

IV. Frithioff A, Frendø M, Weiss K, Foghsgaard S, Pedersen DB, Sørensen MS, Andersen SAW. Effect of 3D-printed models on Cadaveric Dissection in Temporal Bone Training *OTO Open. 2021;5(4):2473974X211065012.*

V. Frithioff A, Frendø M, Foghsgaard S, Sørensen MS, Andersen SAW. Are video-recordings reliable for Assessing Surgical Performance? A Prospective Reliability Study Using Generalizability theory *[published online ahead of print, 2022 Oct 18]. Simul Healthc. 2022;10.1097/SIH.0000000000000672.*

Introduction

Hearing loss is a rising problem globally and the WHO predicts that by 2050, nearly 2.5 billion people will have some degree of hearing loss. In this group, at least 700 million will require rehabilitation services.[1] Untreated hearing loss is associated with social isolation, depression and dementia.[2–4] Unattended hearing loss is currently estimated to cost nearly 1 trillion international dollars a year.[1] Otosurgical procedures can improve or recreate hearing in patients with congenital or acquired hearing loss, where conventional hearing aids alone are insufficient. This includes middle ear surgical procedures such as removal of cholesteatomas, reconstruction of the ossicular chain (tympanoplasty type II–IV), Bone Anchored Hearing Systems (BAHS) for conductive hearing loss and Cochlear Implantation (CI) for profound sensorineural hearing loss.

Several of these procedures involve drilling the temporal bone in close relation to important anatomical structures. Therefore, excellent surgical skills are required, necessitating repeated practice. When novices start to learn the procedure, they are naturally prone to making mistakes due to their lack of experience.[5] For that reason, simulation-based training (SBT) has prospered during the last decade in all surgical specialties including in otorhinolaryngology.[6–8] SBT offers trainees the possibility to develop skills in a risk-free environment by moving the initial learning outside the OR, for the benefit of both patients and trainees. Overall, simulation-based training has a positive effect on knowledge, skills and patient-related outcomes, meaning that the old paradigm of “see one, do one, teach one” is increasingly obsolete.[9]

In temporal bone surgery, cadavers have traditionally been the gold-standard for simulation-based training. However, a shortage in availability remains a major challenge and limits the possibility of a sufficient amount of training.[10] Animal models are not a suitable alternative due to species differences, and this has led to an increased interest in developing other alternative training modalities.[11–13] VR simulators are, for now, the most frequently used alternative and has repeatedly been found to be an effective training tool.[8,14] VR simulation offers the trainee haptic drilling and a realistic visual representation, but lacks important physical aspects of temporal bone surgery such as drill and microscope handling. Physical models, either cast- or 3D-printed, are other alternatives that support the training of instrument-handling but educational evidence on their effectiveness is lacking.[15] As 3D-printing technologies advance, and printing costs decrease, 3D-printed models are perceived to hold great potentials in the future of temporal bone training and several studies have described the creation of such models.[16–19] However, for 3D-printed models to be relevant, the educational effectiveness, and how best to implement them in the surgical curriculum, must be established. This thesis aims to investigate 3D-printed models as a tool in training of temporal bone surgery.

Background

Mastoidectomy

Mastoidectomy is considered a fundamental and very important procedure in otologic surgery. The procedure is used for several purposes such as treatment of acute mastoiditis, removal of pathology including cholesteatomas but also for accessing the middle or inner ear for removal of vestibular schwannomas or restoring hearing through cochlear implantation.[20–22] A mastoidectomy comprises drilling away the air-filled cells in the mastoid bone using an otosurgical drill and suction/irrigation, while the surgical field is magnified by a microscope.[20] The temporal bone houses key anatomical structures such as the facial nerve, chorda, dura, sigmoid sinus, ossicles and semicircular canals. These structures serve as surgical landmarks and boundaries. A precise comprehension of the anatomy as well as the acquisition of fine motor skills are therefore crucial for avoiding injury and serious adverse events.[20,23] Such incidents include facial nerve injury, brain injury, vertigo, bleeding, or iatrogenic hearing loss.[20]

Acquisition of surgical skills

Becoming a good surgeon is challenging and puts high demands on both knowledge and fine motor skills. The overall goal for all future surgeons is to rapidly escalate expertise to perform accurate and safe surgery. Several theoretical frameworks describe the acquisition of motor skills and one of the most widely used theories is the Fitts-Posner division of motor skills learning.[5] In this theory, learning is divided into three stages: the cognitive, associative, and autonomous stages. In the cognitive stage, the trainee understands the task but must keep attention on every movement, which the trainee performs slowly and inconsistently. With repeated practice, the trainee reaches the associative stage and become more fluent, and movements are occupying less cognitive capacity. At the final, autonomous stage, the trainee performs the task with accurate and efficient movements with minimal mental effort.[5,24–26]

This transition does not depend strictly on the duration of training or the number of repetitions of a procedure but evolves from continued and deliberate practice[27,28] adjusted to accommodate individual differences.[27] According to Ericsson, experience and repeated practice are necessary but not sufficient for true expertise.[28,29] Instead, true expertise depends on continued and cognitively engaged efforts to improve.[28] Deliberate practice comprises elements such as feedback, error correction/monitoring, motivation, clearly defined goals, and progression.[28–30] One way to support such training is through the instructional concept known as mastery learning.[31] In mastery learning, trainees must achieve proficiency in one domain of training before proceeding to the next level. Training must be supported by elements that relate to the concept of deliberate practice. Mastery learning is relevant to competency-based training and dictates that educational progress is based on demonstrated competency rather than time spent on training.[32]

Despite the Fitts-Posners division, trainees have traditionally obtained surgical skills primarily by supervised surgery (i.e., the apprenticeship model). This means that even the early stages of learning happen in the OR. At this stage of learning, trainees are prone to mistakes,[5] of which are undesirable when working with patients. Instead, simulation-based training is a way to move the initial learning outside the OR, providing a supplement to clinical practice in a risk-free environment.[25,33] Also, simulation-based training can support elements of both deliberate practice and mastery learning which can be difficult to integrate fully into the apprenticeship model.[30,33]

Educational outcome measures

Evaluating the impact of educational interventions requires suitable outcome measures. Not all educational outcomes are equally important and evaluations are therefore guided by taxonomy levels. One widely used framework is Kirkpatrick's four-level hierarchy of educational outcomes.[34] Kirkpatrick's model ranges from outcomes related to the learner's reaction to the training intervention (Level 1) to outcomes demonstrating actual training benefit on patient health/well-being (Level 4; Table 1).[35] In addition to being used independently in the literature, Kirkpatrick's hierarchy is also an integrated part of the Medical Education Research Study Quality Instrument (MERSQI); a recognized tool for appraising methodological quality in medical educational research.[36–38] Kirkpatrick's model has been criticized for only being focused on simple outcomes, failing to evaluate the complex processes that can impact training.[39,40]

Table 1 *– Kirkpatrick's level of hierarchy*

Hierarchy	Description of outcome
Level 1: Learners' reaction	Trainee satisfaction with the training i.e. did the trainees enjoy it?
Level 2a: Modifications of attitudes/perception	Reciprocal attitude or perception between groups towards the training.
Level 2b: Acquisition of knowledge and/or skills	Trainees demonstrate actual knowledge and/or skill development after training.
Level 3: Changes in behavior	Transfer from the learning environment to the clinic.
Level 4a: Changes in organizational practices	Changes in the care delivery due to the training.
Level 4b: Benefits to patients	Improvement in patient health/well-being.

**Modified from Hammick*[35]

Transfer of learning

Skills acquired in a simulated setting are only relevant if skills can be *transferred*.[41] *Transfer of skill* is defined as the *"application of knowledge and skills learned in one context to another"*.[42] In simulation-based training, *"transfer"* most often refers to skills learned in the simulated setting impacting real-life performance, i.e., patient care.[41,43] *"Transfer of skill"* can also be between different simulation modalities or two different procedures which share traits. Specific for mastoidectomy training, skills obtained during VR training transfer to cadaveric dissection. This means that VR simulation can be a relevant supplement to cadaver dissection which is a scarce resource.[10,44,45] In Study IV, we investigate the transfer of mastoidectomy skills from 3D-printed models to cadaveric dissection.

Mastoidectomy training

Basic mastoidectomy skills have typically been acquired by cadaver dissection and supervised surgery.[8,45] Few institutions offer their trainees access to an open temporal bone lab where they can perform cadaver dissection freely as needed.[46] Cadaver training is typically organized during national or international dissection courses comprising a few days of intensive training.[10,12]

Such courses are costly and require dedicated instructors and dissection facilities. Further, most trainees only drill one cadaver during a dissection course, limiting the possibility for repeated and deliberate practice. VR simulators and 3D-printed models offer an alternative to cadaver dissection and can be used locally at the department at the convenience of the trainees.[47] While these alternatives cannot substitute cadaver dissection entirely, they can serve as a supplement and offer trainees more diverse training opportunities.

VR simulators are, for now, the more frequently used of these two and different simulators are available both as freeware and commercially.[10,13,48] Currently the most well-known simulators are:

The Ohio State University simulator (USA),[49] University of Melbourne simulator (Australia)[50], CardinalSim (USA),[51] VOXEL-MAN (Germany)[52] and the Visible Ear Simulator (VES; Denmark)[53].

All these simulators use volumetric models built on either CT-derived data or cryo sections of fresh frozen temporal bones and offer the trainees a haptic drilling experience.[13,54] Further, they include a range of learning supports such as integrated tutoring or automated feedback which is only possible in a VR environment.[55–57]

Physical models, either casted or 3D-printed, offer the trainee a training experience closer to the setup of real surgery. Instead of viewing a PC screen, the trainee operates under the surgical microscope and the haptic device used in VR simulation is replaced with an actual otosurgical drill (Figure 1). The first report on a physical model for temporal bone training was in 1998[58] with many more being presented during recent years.[8,13] Nevertheless, integration of these models in the surgical curricula has so far been limited for several reasons. First, the air cells of the mastoid bone can be difficult to reproduce in a plastic model. Second, the haptic resemblance is often low as the use of high-speed drills creates so much heat that the plastic melts. Third, the manufacturing cost has been a barrier which, however, is decreasing due to technological development.[8,13,15,59,60]

Figure 1 *– VR training using VES (left) and training on a physical 3D-printed model (right)*

Assessment of surgical skills

Assessment of surgical skills is essential in both clinical training and medical educational research. Historically, assessment of surgical skills was largely informal and unstructured, based on hours of training, number of procedures, or by using direct observations without well-defined assessment criteria.[61,62] This is problematic as such assessment comes with a substantial risk of bias. The structured assessment addresses this through tools, enabling a way to standardize and structure observations, thereby helping the assessor to provide objective assessment of surgical performance.[61] This helps measure learning progress and is not only used for *testing* but can be a powerful tool in *supporting* learning.[63] Evaluating surgical technical skills using objective assessment is a key element in competency-based training to ensure that trainees have reached a certain competence level before progressing in their training.[64] Further, assessment is a prerequisite in medical educational research for measuring the effects of training interventions. This highlights the need for reliable assessments supported by validity evidence in many contexts of surgical training including simulation-based training.

Validity

Validity evidence is scientific evidence supporting that an assessment measures what it is intended to measure.[65] This is, according to Downing and Yudkowski, *"the single most important characteristic of assessment data"*.[65] If assessments are not supported by validity evidence, the interpretations based on these assessments are meaningless and no conclusions can be drawn from them. This has implications for many aspects of surgical training including competency-based training: If assessments do not reflect surgical proficiency, it is impossible to make any conclusions on the level of the trainee. In summary, for competency-based training to be relevant, assessments of performance must be supported by validity evidence. Validity evidence is not a dichotomous outcome and no assessment can be deemed "*valid*" or "*not valid*". Rather, validity research is hypothesis-driven, either supporting or opposing the validity *argument*.[64,66] This unitary view on validity has replaced the classic view consisting of four different *types* of validity evidence (i.e., construct, face, criterion and concurrent).[64] Unfortunately, the use of this classic framework is still very common in the surgical training literature.[67,68]

A widely accepted approach for gathering validity evidence is Messick's framework. It consists of five *sources* of validity evidence: Content, response process, internal structure, relationship with other variables and consequences (Table 2).[64,66,69–71] All five sources add to the body of evidence supporting the validity of an assessment; exploring the sources should be seen as an ongoing process.[64,70,72] Messick's framework is widely accepted in the educational research community and has been adopted by American Educational Research Association, the American Psychological

Association, and the National Council on Measurement in Education for being the standard for educational and psychological testing.[71] In Study III, we used Messick's framework to structure the collection of validity evidence.

Table 2 *– Overview of the different sources of validity according to Messick's framework*

Source	**Description**
Content	Relevance of the test/training tool content when comparing this with the domain of interest.
Response process	Relates to the stringency in data management/collection for eliminating errors from test administrations etc.
Internal structure	The reliability, reproducibility of a test result.
Relationship with other variables	Correlations to other external, relevant and independent measures.
Consequences	Consequences of the test use and impact of assessment

Modified from Assessment in Health Professions Education[65]

Reliability

Another important part of the validity evidence and Messick's third source of validity (i.e., internal process) is reliability. Reliability concerns the consistency of test results meaning that assessments should be consistent if repeated multiple times under the same conditions.[65] This is key in surgical training as assessments should not only measure performance but also do so consistently. For example, if a trainee must reach a certain score to pass a test, the assessment score should reflect the "true" performance of the trainee. The observed test score reflects the "true" performance of the trainee plus an error of measurement. When reliability levels are high, the observed test score approaches the "true" performance and the risk of a non-competent individual passing a test is low.

Contrarily, low-reliability levels would lead to the opposite situation. This emphasizes that reliability is highly relevant for both researchers and educators.

Similarly, when looking at test scores across a group of trainees, variances in scores contain both "true" scores (i.e., differences in skills) and error variances:[73]

$$\textit{Observed variances} = \textit{True score variance} + \textit{Error variance.}$$

In classic test theory, a reliability coefficient is calculated as the proportion of variance which attributes to the trainees' true performance:

$$\textit{Reliability coefficient (ICC)} = \frac{\textit{True score variance}}{\textit{True score variance} + \textit{Error variance}}$$

Several factors contribute to the error variance including 1) how different raters assess performances (inter-rater reliability), 2) the rater's consistency over time (intra-rater reliability), 3) if the items of the test itself are consistent with each other (i.e., internal consistency). However, more factors can contribute to error variance which is not merely isolated to rater differences. Therefore, classic test theory has been criticized for giving an overly simplified view of the different contributors to the error variance.[65,74]

Generalizability (G) theory addresses this problem by classifying the error variance in multiple different factors (named facets) which all affect the score.[74] G analysis can estimate the contribution of each facet, thereby helping to identify both large and small contributors to the total amount of error and estimating an overall reliability coefficient called a G coefficient.[74,75] Based on G analyses, so-called Decisions (D) studies can be conducted. D studies are statistical simulations of multiple assessment scenarios, helping to estimate the effect on reliability when employing changes to the original G study (e.g., how does it affect reliability to change the number of raters?). Consequently, G theory can help researches and educators to optimize study designs or make recommendations on conducting test/training programs. [65]

Standard setting

Standard setting is the process of identifying when a trainee is competent or not.[65] Standards can be divided into relative (norm-based) and absolute (criterion-based) standards where the latter is the most used for testing technical competence.[65] An absolute standard uses a predetermined level reflecting the minimum level of competence for a trainee. Standard setting, therefore, helps answer the question: *When is a trainee competent enough?* This is a central element in competency-based training.[65,76] Nevertheless, establishing a credible standard can be challenging and several methods apply for different situations, with no single gold-standard method.[65,76–79] Overall, methods for standard setting can either be item-based (focusing on the content of the test) or examinee-based (focusing on the performance of the examinees). In the item-based methods, assessors review the test items and set the level for what they see as a "just adequate" performance. In the examinee-based methods, assessors utilize the actual performance of a group of examinees for setting a standard.[65,80] Table 3 provides an overview of four common methods for standard setting, but many more are found in the literature.[81]

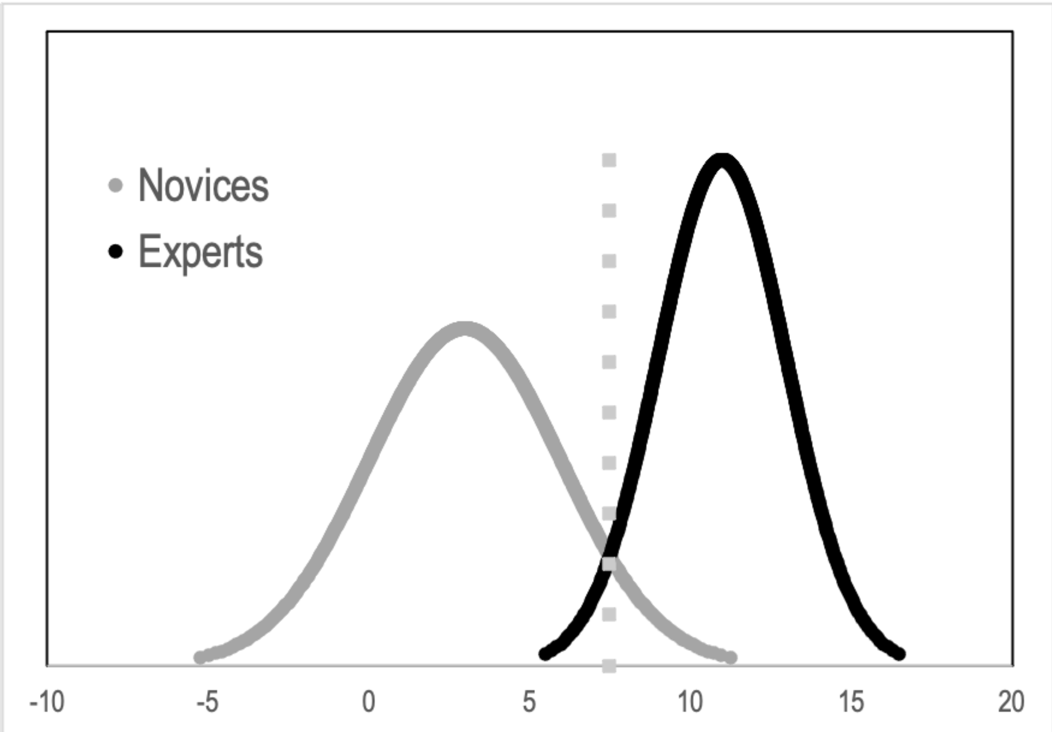
In Study III, we use the contrasting groups method for setting a standard in mastoidectomy training using 3D-printed models. The contrasting groups method is an examinee-based method that uses the normal distribution of test scores between a group of experts and a group of novices setting the intersection as the pass/fail level (Figure 2).[65,82]

***Table 3** – Common methods for standard setting*

Method	Category	Description
Angoff	Item-based	Assessors estimate for each test-item the proportion of passings in a borderline group. The average score of each item is then summed up for setting a standard.
Hofstee (the relative-absolute compromise method)	Item-based	Utilize agreements on the minimum and maximum acceptable passing score and failure rate by the assessors for setting a standard.
Contrasting-groups	Examinee-based	Uses the distribution of test scores for a proficient and non-proficient group. The intersection is set as the standard.
Borderline-group	Examinee-based	Assessors identify examinees with borderline performances. The standard is set as the mean/median score of the borderline group.

Setting a credible standard is an important part of Messick's 5th source of validity, ´"Consequences of testing". Cook et al. argue that *"Both clinical and educational assessments can be seen as an intervention. The act of administering or taking a test, the test interpretation of scores, and ensuing decisions and actions influence those being assessed"*.[83] "Consequences of testing" addresses this, and standard setting is at the center. For example, in certification exams, setting a wrong standard could result in a non-competent individual passing or a competent individual failing. Both situations would represent a problem as the non-competent who passed would constitute a threat to the patient, and the competent who failed could have been a resource in the clinic. Evaluating *consequences of testing* and standard setting is therefore an essential part of the validity process. However, the domain is infrequently reported in the medical educational literature.[67,83]

Figure 2 *– Contrasting groups method*



Example of contrasting groups method analysis. The x-axis represents an imaginary assessment score between 0 and 20. The two bell curves represent the normal distribution of scores in an expert and novice group. The intersection between the two curves (dotted line) represents the pass/fail-level.

Assessment of mastoidectomy performance

There are several tools designed for assessing mastoidectomy performance.[49,84–92] The assessment tools comprise Task-based Checklists (TBC), Global Rating Scales (GRS), and Final-product assessment (FPA).[84,88,92] TBC includes statements and/or questions that can be scored based on the direct actions of the trainee. GRS considers the performance as a whole and assessors are often asked to rate the overall performance or global impressions based on parts of the procedure.[93] Contrary to GRS and TBC assessments, FPA strictly relies on the final result and does not include any evaluation of the process. An advantage of FPA is that assessment can be performed at the end of each performance and does not require the simultaneous presence of an assessor. Inconclusive correlations between GRS, TBC and FPA for mastoidectomy assessment, highlight that the different assessment tools reflect different aspects of the performance.[88]

A well-known final product assessment tool, supported by extensive validity evidence, is the Welling Scale (WS).[84,86,90,91] This assessment tool was developed at the Ohio State University and has been used for grading mastoidectomy performances on cadavers, in VR simulation, and on 3D-printed models, demonstrating good inter- and intra-rater reliability.[84,86,91,94]

Originally, the WS consist of 35 items which are graded binarily (0 = Incomplete/Inadequate or 1 = Complete);[86,91] However, for reflecting the procedure in our context, the assessment tool has been modified to comprise only 26 items for cadaver surgery and 25 items for 3D-printed model performances (Appendix).[94] The 3D-printed model used in our studies does not include a representation of the chorda and consequently item 25 (Tympanic chorda exposed) cannot be assessed.

Additive manufacturing (3D-printing)

Additive manufacturing (AM) is the industrial application of *3D-printing*, which is a wide palette of technologies used for creating physical objects from three-dimensional digital models. Research on additive manufacturing emerged in the 1980's, and gained a foothold during the 00's with many new technologies being introduced.[95] AM has since been widely adopted in the manufacturing industry and is also considered to hold great potential in the field of medicine, as the technologies enable the recreation of anatomically correct, patient-specific structures.[96–98] Also in the field of surgical education, AM creates the possibility of manufacturing high-fidelity models for simulation-based training.[97,99]

AM is a general term for considerably different manufacturing processes, which are based on the same general idea: building a physical object from 3D-model data by adding successive layers of material.[98,100] Materials can either be polymers, ceramics, metals or composites in different feedstocks (i.e. filaments, powder, liquid or sheet stocks) which are processed differently, for example by using laser, hot extrusion, electron beam processing, or UV-light exposure.[95] The nomenclature around AM is not stringently in the literature and often clouded by the interchangeable use of acronyms, tradenames and abbreviations. This is despite efforts by the ASTM/ISO 52900:21 to unify the nomenclature.[95] An overview of the seven most widely adopted AM processes including acronyms and common tradenames is presented in Table 4.

***Table 4** – Overview of 3D-printing processes*

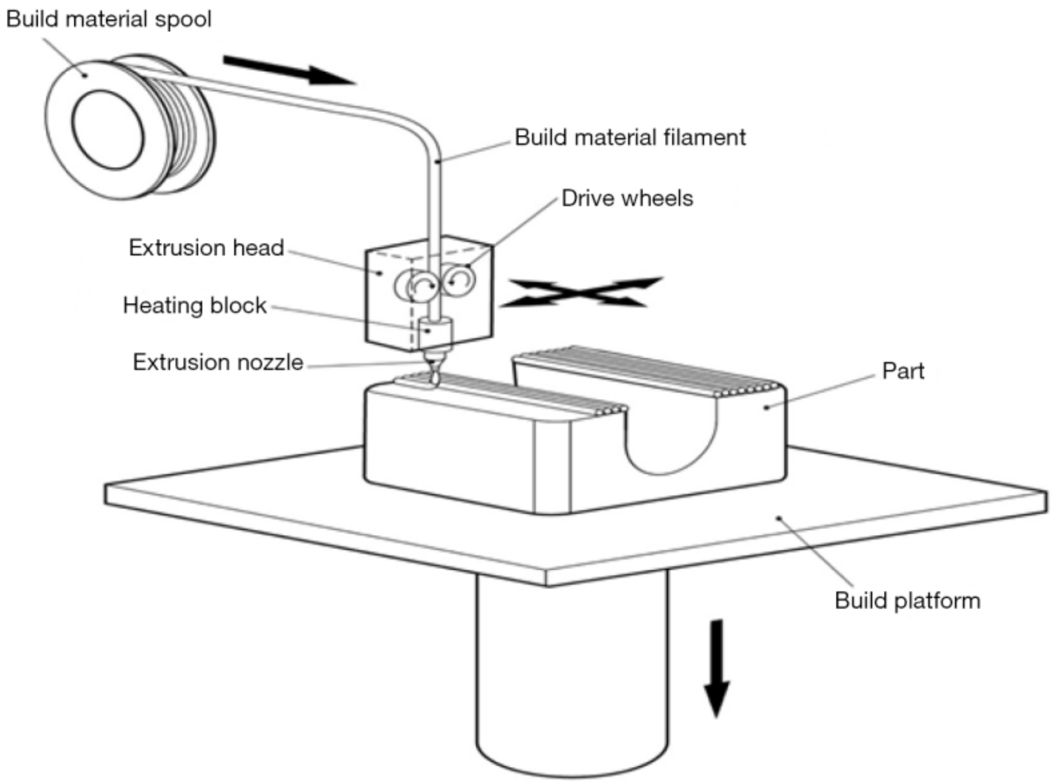
Process name	***Acronyms and tradenames***	***Materials***	***Material feedstock***	***Short process description***
Material Extrusion	Fused deposition modeling (FDM) Fused filament fabrication (FFF)	Polymers	Filament	A filament wire is extruded through a heated nozzle which causes the filament to melt before being deposited on a building plate.
Material Jetting	Multijet Modeling Polyjet	Polymers	Melted	Thermal plastic is jetted on a build plate where it becomes solid.
Powder-bed Fusion	Selective Laser Sintering (SLS) Selective Laser Melting (SLM) Direct Metal Laser Sintering (DMLS)	Polymers and metals	Powder	By either laser or electron beam processing a powdered material is hardened and new powder is successively added. This process happens in a confined chamber (i.e. a powderbed).
Binder Jetting	Powder bed 3D-printing Inkjet 3D-printing	Polymers	Powder + Liquid binder	Comparable to powder bed fusion. However, instead of the powder being hardened by laser/electron beaming a liquid binder is added successively.
Vat photopolymerization	Stereolithography (SLA) Dark Light Processing (DLP)	Polymers	Liquid	A photopolymer (resin) in a vat, is cured from the top layer and down by using UV lighting.
Sheet lamination	Laminated Object Manufacturing (LOM)	Polymers and metals	Sheet stock	Sheets are cut by either a high-power laser or a punch. Subsequently the cut layers are stacked and welded using ultrasonic welding.
Directed energy deposition	Laser Engineered Net Shape (LENS) Cladding Direct Metal Printing (DMP)	Metals	Filament or powder	The metal filament or powder is melted by a laser and deposited through a nozzle.

Regardless of the process, AM entails close relation between digital and physical manufacturing. The process can roughly be divided into three steps: 1) pre-processing, 2) processing, and 3) post-processing. 1) Pre-processing concerns using Computer-Aided Design (CAD) to generate/create a digital 3D model. This model is translated to a printable file (typically STL format), which is then refined and divided into layers (i.e., slicing) before being transferred to the AM system. Further, the pre-processing step includes choosing the appropriate AM technology, printer, and material and setting up the AM system and processing parameters. 2) The AM process is mostly autonomously executed; however, control and human monitoring and intervention of the process is normal. 3) The post-processing also depends on the print-technology, but most frequently concerns removing the object from the machine, as well as removal of support structures or residual material and further refinement of the printed element.

Material Extrusion

Material-extrusion printers are the most common type of printer; they are typically less costly and easier to use than other print technologies.[95] In extrusion-based printers, the material is a filament of thermoplastic (e.g., PLA and ABS), which is coiled onto a spool. The material is pushed by a drive gear through an extrusion-head where the filament is heated above its melting point. Through a nozzle, the semi-liquid material is deposited on a building platform where it cures. Both the extrusion-head and building platform (printing bed) moves in the *x-y* plan. Each time a successive layer is deposited on the building platform, the extrusion-head moves up (or platform down) before starting to deposit the next layer (Figure 3). Printing the object requires that the base and overhanging structures are supported. These support structures are generated in software programs (e.g., Ultimaker-Cura, Slic3er or PrusaSlicer) before printing. This software is also used for adjusting several other parameters during the printing process such as the temperature of the extrusion-head and building platform, the print speed, infill, layer heights etc. before slicing.[101–103] The low cost of printers and materials and ease of operation make extrusion-based printers suitable for technical non-professionals to undertake 3D-printing.

Figure 3 *– Material Extrusion additive manufacturing.*



Reproduced from reference[104], published as open access under Creative Commons License

3D-printing in temporal bone training

In general, 3D-printed models have gained favor as a suitable tool for surgical training, but implementation is still lacking.[97] Temporal bone surgery is no exception[13,59,105,106] and many 3D-printed models already exist.[16,18,107–116] This is probably because the 3D-printed models can provide physical and reproducible simulation-based training for future ORL surgeons.[15] Currently, temporal bone surgery/mastoidectomy is the most common usage of 3D-printed models in otolaryngology training.[106] The existing models are printed using a variety of different printing techniques including Powder Bed Fusion, Binder Jetting, Vat Photopolymerization and Extrusion-based printers.[117] This suggests that more than one technology can be suitable for accurate replication of the temporal bone, but also highlights that no gold-standard exists.

In 2018, the American Academy of Otolaryngology-Head and Neck Surgery Foundation (AAO-HNSF) created a work group of temporal bone experts to coordinate future work in the development and validation of 3D-printed models for temporal bone surgery. This work was initiated based on the opinion that 3D-printing should be part of the future training curricula in ORL.[118] The group concluded that adequate models can be created using a variety of print types and materials[118] but did not systematically compare the educational evidence supporting the existing models; nor did they provide specific recommendations/guidance for training institutions to create their 3D-printed models. Without such recommendations, the large number of available printer and print materials could represent a major barrier to implementation. Establishing 3D-printing facilities in the clinical setting is expensive and time-consuming.[119] Specific recommendations for a print-routine using affordable and easy-to-use print technologies could help lower this barrier and aid training institutions in starting their manufacturing. In Study II, we provide recommendations on how to print a cost-effective temporal bone model using an extrusion-based printer.

The Open Ear library

The Open Ear Library is a publicly available dataset consisting of eight different digitized temporal bone models.[120] Each of the eight datasets is built on a combination of Cone Beam Computerized Tomography (CBCT) and micro-slicing imaging.[120] In contrast to other available datasets, the Open Ear datasets offer a reconstructed three-dimensional model of the human temporal bones including colors.[120] The datasets have been used for offering case-variation in the Visible Ear Simulator[121] but can also be translated to STL-files for 3D-printing. In Study II, we utilized a dataset from the Open Ear library for creating a 3D-printed temporal bone model.

Research aims

This thesis aims to explore the current educational evidence supporting 3D-printed models for temporal bone training, and develop and evaluate a cost-effective 3D-printed model that can be manufactured by clinicians in most ORL training institutions. Further, the studies aim to collect validity evidence for using the model in resident training of temporal bone surgery. The specific aims of the five studies are:

Study I: To explore the methods used for manufacturing 3D-printed temporal bone models and how existing educational evidence supports the training of temporal bone surgery.

Study II: To create a cost-effective 3D-printed temporal bone model suited for surgical training using entry-level print technologies.

Study III: To gather validity evidence for the use of a 3D-printed model for mastoidectomy training and explore a credible pass/fail score using this model.

Study IV: To explore if mastoidectomy skills acquired by training on 3D-printed models transfer to cadaveric dissection performance, and how this compares to VR simulation training.

Study V: To investigate the reliability of video-based assessments *vs* physical assessments of mastoidectomy performances on 3D-printed models and cadavers.

Research hypotheses

Study I: 3D-printing temporal bone models is feasible but failure to use evidence-based methods in research on their educational impact has led to inadequate knowledge in the field.

Study II: A 3D-printed model designed for mastoidectomy training can be printed using affordable entry-level print technologies.

Study III: Validity evidence supports using 3D-printed models for mastoidectomy training.

Study IV: Training using 3D-printed models can improve mastoidectomy skills and these skills are transferable to cadaveric dissection.

Study V: Video-based assessment of cadaver dissections and 3D-printed models is as reliable as hands-on physical assessment.

Summary of studies

Study I: 3D-printed models for temporal bone training: A systematic review

Background

Temporal bone surgery takes place in an anatomically complex area and requires good surgical skills. Obtaining these skills should happen in a risk-free environment not compromising patient safety, and for that purpose, cadaver dissection has been the gold standard. Nevertheless, a shortage of available cadavers suited for temporal bone training has led to a rising interest in alternative training methods. 3D-printed models are a promising alternative due to their physical similarities with human temporal bones. Even though several studies have described 3D-printing models for temporal bone training, educational evidence on their effectiveness seems scarce in comparison to VR simulation training.

Methods

Following the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guideline[122], we searched PubMed, Embase, the Cochrane library and Web of Science to identify relevant studies on manufacturing, training, and/or validation of 3D-printed models suited for temporal bone training. Two authors independently screened records identified for eligibility, and data from relevant studies were extracted using a data extraction form piloted on five studies. Quality assessment was performed using Medical Education Research Study Quality Instrument (MERSQI) and further, educational outcomes were evaluated according to Kirkpatrick's level of hierarchy.

Results

Searches yielded a total of 595 unique articles whereof 36 studies were included in the final analysis. Image acquisition was based on both CT and micro-CT varying in resolution and slice thickness (0.23–0.38mm for clinical scans and as low as 12µm for cadaveric specimens). Many studies reported using semi-automatic segmentation of the bone volume based on imaging thresholds. Delineation of key structures, such as the facial nerve, still warranted manual segmentation. There were a variety of print technologies and materials used for printing including ABS, PLA, resin and various cast-powders. Several studies attempted to find the most suitable material and resin seems promising. Nevertheless, resin necessitated the removal of residual material from hollow structures after printing, and no final conclusion on best practices can be made from the current literature. The printer and printing material have a big impact on manufacturing price which varied from 0.9 to 400 USD per model. Only a few studies reported the manufacturing costs and an accurate cost-analysis was often absent.

Educational evidence on the use of 3D-printed models for temporal bone training was also largely absent and the was quality low. Most evaluations were solely based on trainee/expert opinion, corresponding to Level 1 in Kirkpatrick's hierarchy and no studies included outcomes above Level 2a (i.e., Modifications of attitudes/perception). Overall, there was a positive attitude toward the use of 3D-printed models for educational purposes. In general, MERSQI scores were low with a median score of eight. This was often due to a lack of objective outcomes and exploration of validity evidence.

Conclusion

3D-printing temporal bone models suitable for surgical training is feasible using many different print technologies and materials. For now, it is not possible to give specific recommendations on the most suitable print technology and materials. Further details are needed on the optimal manufacturing workflow before conclusions can be made.

Most importantly, evidence supporting the educational effectiveness (i.e., whether training improves surgical performance) is currently lacking.

Study II: 3D-printing a cost-effective model for mastoidectomy training

Background

Commercially available extrusion-based 3D-printers are low-cost printers that are easy to operate. These printers could enable surgical training institutions without prior 3D-printing experience to manufacture their own cost-effective temporal bone models. Such models could be an alternative to the expensive commercially available models and help departments offer their trainees ample training opportunities. In this study, we aimed to create a cost-effective model for temporal bone training using inexpensive and commercially available print technologies.

Methods

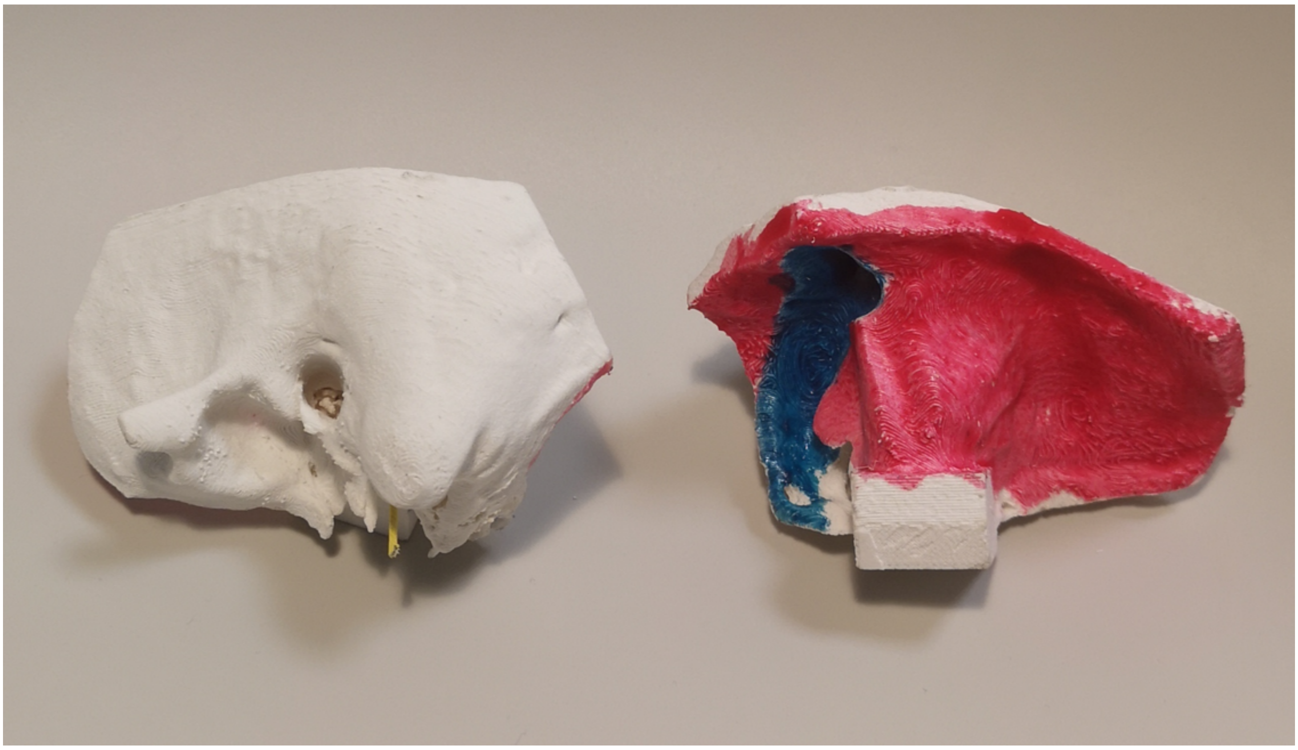
We created a printable file based on a digitized temporal bone model from the OpenEar library.[121] Two 3D-printing experts identified eight inexpensive materials which could mimic bone properties and be used in an extrusion-based printer. An experienced otosurgeon drilled the materials for testing which replicated bone properties most accurately. The model was then printed using a material extrusion printer (Ender-3, Creality, Shenzhen, China) and the optimal print settings were iteratively developed.

During a practical dissection course, a total of 11 ORL residents and attending physicians performed two anatomical mastoidectomies on the 3D-printed models before cadaver dissection. Immediately after completion, the participants filled out a questionnaire on their experience with drilling the 3D-printed models compared with cadaver surgery.

Results

The printable file was generated and made available online for free.[123] Melting was a problem for the majority of the materials during drilling. However, a PLA filament with a high level of chalk (Lay-brick filament, CC-products, Cologne, Germany) was more heat resistant compared to the other materials and created a more realistic drilling experience. For optimizing the number of successful prints we made some minor modifications to the printer: First, we installed a direct drive (Micro Swiss Direct Drive Extruder for Creality Ender3, Mico Swiss, Ramsey, MN, USA) which prevents the filament from breaking during printing and reduces wear of the drive gear. Second, we installed a ruby nozzle that does not get as easily worn out as a standard nozzle. These modifications enhance durability: The ruby nozzle can print >100 models before wearing out *vs* ~5 models for the standard nozzle, and the direct drive needs replacement after estimated 250 prints which is considerably more than the standard drive gear. After printing, the dura, the sigmoid sinus, and the facial nerve had to be manually applied using colored latex and a yellow wire, respectively (Figure 4). The total manufacturing cost of "The Copenhagen Temporal Bone model" was estimated to be ~15 USD per model with a material cost of ~5 USD per model. Participants in the temporal bone course evaluated the model positively. They agreed that the model adequately replicated temporal bone anatomy and had the opinion that it can serve as a good tool for mastoidectomy training.

Figure 4 *–The 3D-printed temporal bone model (Copenhagen Temporal Bone).*



Conclusion

It is feasible to create a cost-effective temporal bone model using entry-level 3D-print technologies. The model was well received by the residents who considered it to be a good training tool for learning the mastoidectomy procedure. Despite using a heat-resistant filament, some melting still occurs during drilling and a few post-processing steps are required for representing key anatomical structures. Nevertheless, this study can serve as a guide for most training departments to start manufacturing their own cost-effective temporal bone training models.

Study III: 3D-printed models for temporal bone training: A validity study

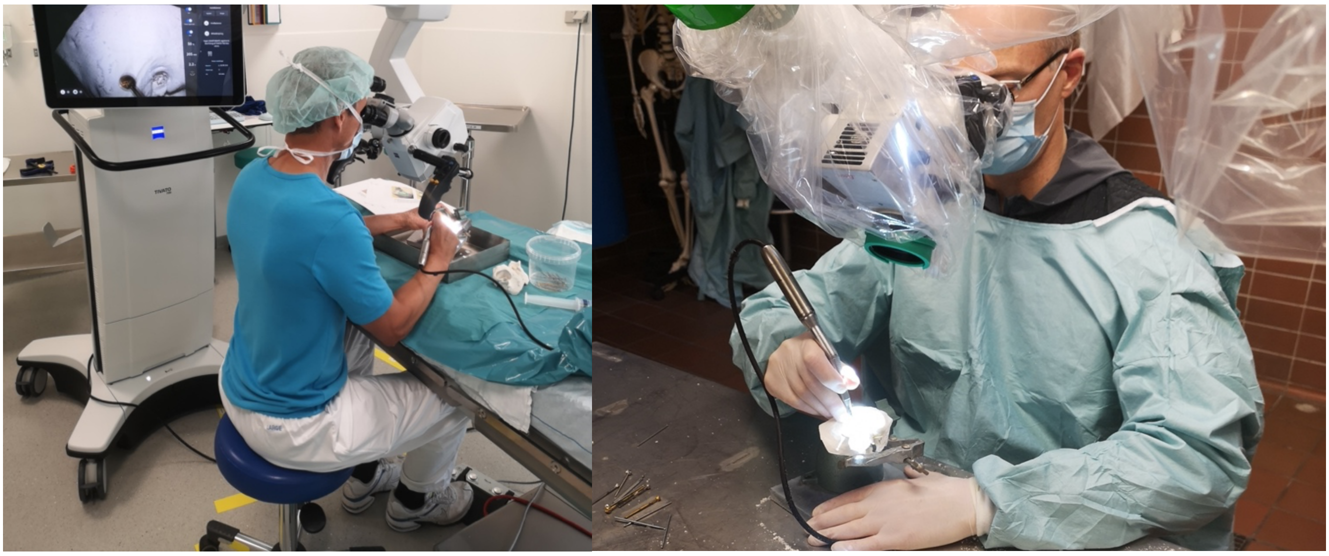
Background

There are several reports on 3D-printed temporal bone models. Nevertheless, due to the use of outdated validity frameworks, strong validity evidence supporting their use in training is still lacking. It is pivotal to explore the validity evidence supporting the 3D-printed models in training. Evaluations should be based on modern validity frameworks, such as Messick's validity framework, if the models are to be integrated in the temporal bone training curricula. We, therefore, aimed to collect validity evidence for using the Copenhagen 3D-printed temporal bone model in mastoidectomy training. Further, we aimed to establish a credible pass/fail level for supporting competency-based training.

Methods

We evaluated the 3D-printed model according to Messick's five sources of validity evidence. Messick's 1st source (*content)* was addressed by having surgical, educational, and technical experts finding the optimal workflow for printing a cost-effective and realistic temporal bone model. For assessing the remaining sources of validity, we collected data on mastoidectomy-performances from eighteen residents and eleven experts (Figure 5). Each participant completed 2–3 procedures and the performances were assessed by three blinded raters using a 25-item modified Welling Scale. To ensure consistency of the *response process* (Messick's 2nd source), one investigator was responsible for data collection.

Figure 5 *– Two expert otosurgeons performing a mastoidectomy on the 3D-printed model.*



Results

Messick's 3rd source (Internal Structure): There was a high internal consistency of the assessment (Cronbach's alpha=0.92) and good inter-rater reliability (Kappa ranges of 0.68–0.70). Using Generalizability theory, we found an overall G-coefficient of 0.91, corresponding to a very high level of reliability. Decision studies demonstrated that one rater assessing two performances, or two raters assessing one performance, would be adequate for reaching a G-coefficient >0.8, which is generally considered sufficient for high-stakes assessment.

Messick's 4th source (Relationship with other variables): The novice group (residents) achieved a mean performance score of 13.9 points (95 % CI [13.2–14.5]) while the expert group had a score of 23.2 points (95% CI [22.2–24.2]) out of 25 points. This means that there was a difference of 9.3 points between the two groups (95% CI [8.2–10.5], p<0.001), corresponding to 67%.

Messick's 5th source (Consequences of testing): Using the contrasting groups method[82], we established a pass/fail score of 21 out of 25 points on the modified Welling Scale. At this cut-off score, none of the experienced otosurgeons would fail (i.e., zero observed false negatives) and none of the residents would pass (i.e., zero observed false positives).

Conclusion

Validity evidence supports the use of a 3D-printed temporal bone model for mastoidectomy training and assessments using the 25-item modified Welling Scale (WS) are reliable. Further, we have established a pass/fail standard drilling the models of 21 out of 25 WS points. This level can be used to support competency-based training for example guiding when a trainee can progress to the more costly cadaver dissection.

Study IV: Effect of 3D-printed models on Cadaveric Dissection in Temporal Bone Training

Background

3D-printed models could be a well-suited supplement to the scarce available cadaveric temporal bones for training. A prerequisite for 3D-printed models to fill this gap is that skills acquired during training transfer to cadaver surgery. Such *transfer of skills* has been established for mastoidectomy training using VR simulation. This is not the case for 3D-printed models where most evaluations have solely been based on trainee/expert opinion and not outcomes related to actual surgical skills. Consequently, it remains largely unknown whether 3D-printed models can improve surgical performance.[117] This study aims to explore if mastoidectomy skills acquired by training on 3D-printed models transfer to cadaveric dissection performance and how this compares to VR simulation training.

Methods

Eighteen ORL residents received three hours of self-directed mastoidectomy training performing 2–3 mastoidectomies on the Copenhagen 3D-printed temporal bone model (Figure 6). After training, residents independently performed a mastoidectomy on a cadaveric temporal bone (intervention). The 3D-printed models and cadaver dissections were rated by three blinded raters using the 26-items modified Welling Scale for final product assessment.[91,94] Cadaver dissection performances were compared with the performances of 66 ORL residents (historic controls) who had received similar amounts of training using VR simulation before cadaver dissection.

***Figure 6** – Dissection training using 3D-printed models*

Results

The intervention group, training on 3D-printed models, had a mean cadaver dissection performance score of 13.7 points (95% CI, 12.3–15.1) whereas the historic control group, training on a VR simulator, had a mean score of 10.6 points (95% CI, 9.9–11.3). Consequently, the intervention group outperformed the historic controls by 29% during cadaveric dissection (mean difference=3.1 points, [95% CI 1.7–5.0], P<.001). We found a moderate correlation between performance on 3D-printed models and cadaveric dissection performance (r=.49, P<.001).

Conclusion

Mastoidectomy skills acquired during training on 3D-printed models transfer to cadaveric dissection performance. Compared to similar amounts of VR simulation training, 3D-printed models seemed slightly more effective. Overall, 3D-printed temporal bone models are a valuable supplement to cadaveric dissection for basic mastoidectomy skills acquisition.

Study V: Reliability of video-recordings for assessment of mastoidectomy skills in cadaver dissection and training using 3D-printed models

Background

Valid and reliable assessment is pivotal for simulation-based training. Previously, assessors had to be physical present for observing the performance; however, video-based assessment could make remote and asynchronous assessment possible. Reliable assessment often requires more than on assessor[81] but assessors are typical senior consultants with many clinical obligations. By using video-based recordings, multiple assessors can rate the performance at their convenience and without geographical constrains. However, the reliability levels of video-based mastoidectomy assessment are unexplored. In this study, we aim to investigate the reliability of video-based assessment *vs* physical assessments of mastoidectomy performances on 3D-printed models and cadaver dissections.

Methods

At a temporal bone dissection course, eighteen ORL residents independently performed 2–3 mastoidectomies on a 3D-printed temporal bone model followed by one mastoidectomy on a human cadaver. The final dissection results were video-recorded for later evaluation. All cadaver mastoidectomy performances were rated physically immediately after completion by three expert raters. Immediately after the course, the 3D-printed models were also assessed physically. Approximately two weeks after the physical ratings, the same three raters assessed the video-recordings for both cadaver and 3D-printed model drillings.

We used kappa statistics to evaluate agreement among the raters (i.e., inter-rater reliability) and between the same rater's assessment of video and physical performances (i.e., intra-rater reliability).

Further, we used Generalizability theory (G theory) to estimate the relative variance contributions and decision (D) studies to explore the optimal number of performances and raters for a reliable assessment of video- and physical assessments.

Results

Agreement among the raters (i.e., inter-rater reliability) had a mean kappa score of 0.58 (range, 0.53–0.62) for video-based and 0.60 (range, 0.55–0.69) for physical assessment. When comparing video-based and physical ratings for the same raters (i.e., intra-rater reliability), kappa scores had a mean of 0.62 (range, 0.55–0.69).

The overall G coefficient was 0.92 with the largest contribution to the variance (45% for 3D-printed models and 55% for cadaveric dissection) being attributed to the participant's performance (i.e., true score variance). The interaction between the rater and assessment modality contributed with 8.1% variance, meaning that specific raters were affected by the change of modality (i.e., video-based *vs* physical assessment). Performing independent G analysis for video-recordings and physical ratings, we found a G coefficient of 0.86 (video-recordings) and 0.87 (physical ratings) for cadaveric dissection, and 0.85 (video-recordings) and 0.80 (physical ratings) for 3D-printed models. Subsequent D studies found that two raters performing two performances would be sufficient to achieve a G coefficient >0.8 for video-based assessment of 3D-printed models (Figure 7). For physical assessment, a G coefficient >0.8 would require three raters assessing two performances or two raters evaluating three performances (Figure 8). For a single cadaveric dissection performance, both physical and video-ratings would require two raters to achieve a sufficient G coefficient.

***Figure 7** – Decision studies estimating the Generalizability coefficient with an increasing number of observed performances for 1, 2, and 3 raters using video-based assessment.*[124]

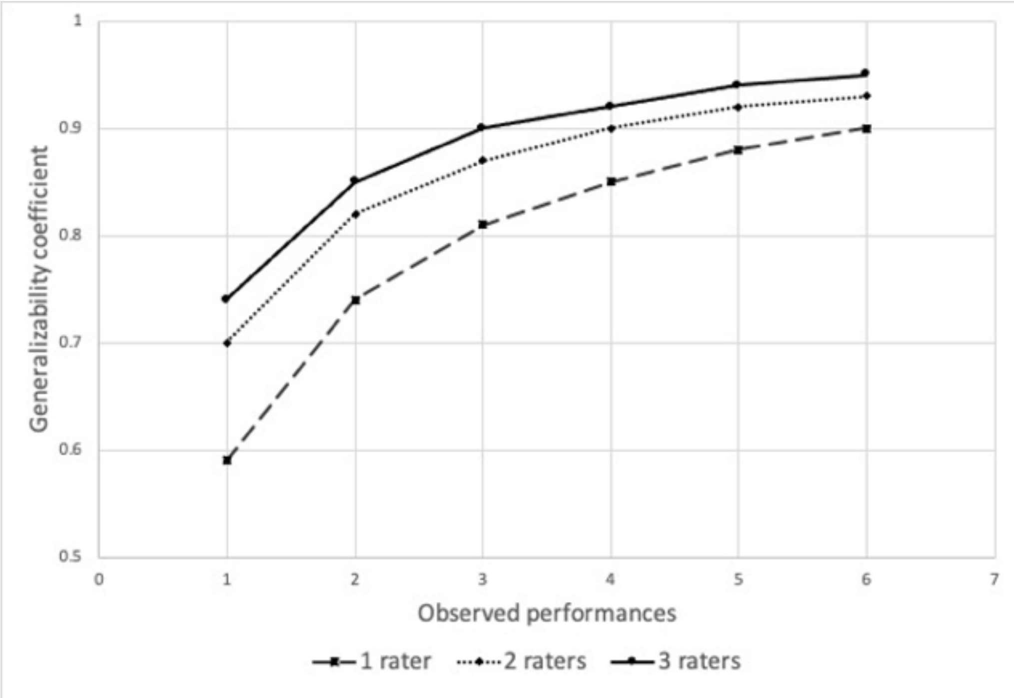
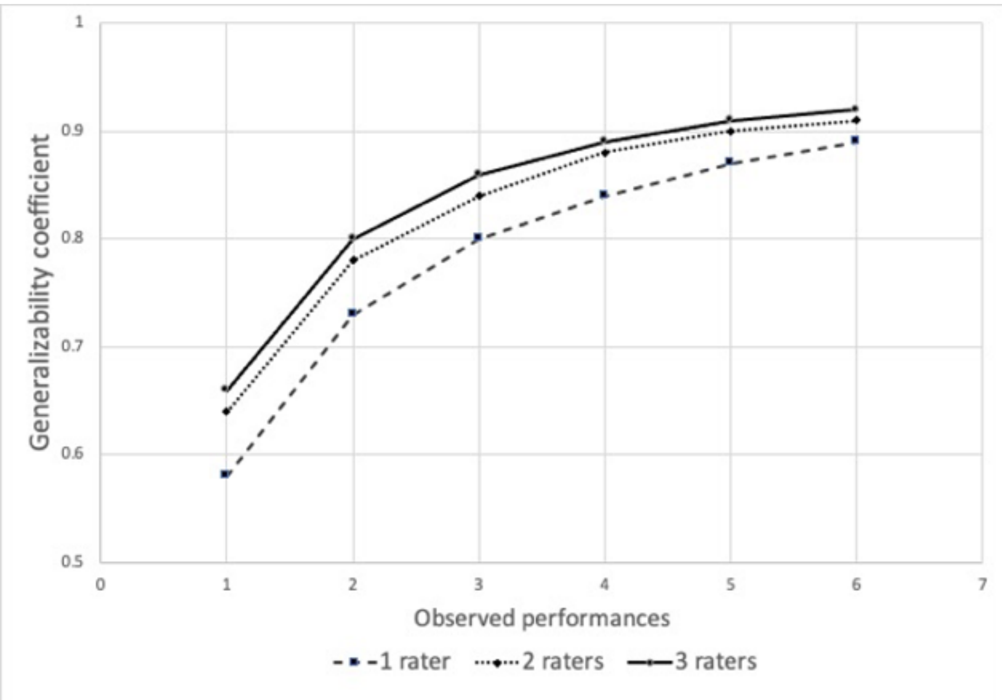


***Figure 8** – Decision studies estimating the Generalizability coefficient with an increasing number of observed performances for 1, 2, and 3 raters using physical assessment.*[124]



Conclusion

In this study, we found that the reliability of mastoidectomy final product assessments was similar across video-based and physical assessments. Rater leniency was high within the same rating modality; however, changing from one assessment method to another affected specific raters more than others. Consequently, a combination of video-based and physical assessments should be avoided. This work supports using video-based assessment in simulation-based training and adds to the current knowledge of reliability in surgical skills assessment.

Discussion

Main findings

In this thesis, we first mapped how 3D-printed temporal bone models are manufactured and how educational evidence supports their use in surgical training (Study I). We found that several prototypes have been described, but information on how best to manufacture these models and knowledge on their educational effectiveness are lacking. Next, we developed a simple and cost-effective method for 3D-printing a temporal bone model based on the OpenEar library, using consumer-grade printers, a chalk-loaded filament, and simple post-processing routines. This enables clinical training departments without technical staff to manufacture their models for training (Study II). We then collected validity evidence according to Messick's framework supporting the use of 3D-printed models for mastoidectomy training; here we established a credible pass/fail score of 21 points (out of 25 points) on a modified Welling Scale using the contrasting groups method (Study III). We then used Generalizability theory to establish that assessments of mastoidectomy performances on the 3D-printed model are reliable. Moreover, we calculated how the number of assessors and performances affected reliability levels (Study III and V). For simulation-based training to be relevant, *transfer of skills* is pivotal and we, therefore, investigated the effect of training using 3D-printed models on cadaveric mastoidectomy performance. We found that skills acquired during 3 hours of self-directed training using 3D-printed models substantially improved trainees' mastoidectomy performance during cadaver surgery (Study IV). Finally, we investigated whether the reliability of video-based assessments was equal to that of physical assessments (Study V). We found that video-based assessment was as reliable as physical assessment, demonstrating that remote assessment is reliable, thereby enabling decentralized training and assessment.

3D-printing and simulation-based training in temporal bone surgery

Since 1997, 37 papers have been published on 3D-printing of temporal bone models. Out of these, 27 papers have been published after 2010, reflecting the technical development that enables the replication of the temporal bone.[117] The first report on 3D-printing a temporal bone model for surgical training was by Begall et al., who created a model using vat photopolymerization.[58] Since then, multiple technical descriptions and studies have been published but the educational evaluation is rare, and the actual training benefits for trainees are largely unknown.[117]

In Study I, we mapped the current evidence on the educational impact of 3D-printed models for temporal bone training through a systematic review of the literature. The search strategy was made in collaboration with an experienced research librarian and our reporting followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines[122] to ensure that all relevant published studies were included. Despite our efforts, we have later identified another study that should have been included in the systematic review: In 2017, Aussedat et al.[125] conducted a prospective study where 17 residents, naïve to the mastoidectomy procedure, performed surgery on a 3D-printed model before cadaveric dissection (intervention). The main outcome was cadaveric dissection performance using a 25-item modified Welling Scale, (i.e., a similar outcome as our Study IV). The residents' cadaveric performances were compared with seventeen residents who received no training before cadaveric dissection (control). Aussedat et al. found that the group that trained on the 3D-printed models had a mean score of 16.8 points compared to 12.4 points in the control group. The study was not included in our search among the 595 studies screened. The reason for this is that the paper does not include any keywords related to the Mesh term "printing, Three-dimensional" or relevant key terms (3D-printing, rapid prototyping, or additive manufacturing). Instead Aussedat et al. refer to their model as a "temporal bone prototype" which was built using stereolithography (a vat photopolymerization process). This illustrates that even when using a systematic approach, conducting a systematic review is not a guarantee for complete coverage of the existing literature. Relevant studies can be overlooked if the terms and nomenclatures are not aligned with those generally used in the field. Comparing the results of Aussedat to Study IV, the intervention group (training on 3D-printed models) and control (training using VR simulation) scored 13.7 and 10.6 out of 25 points, respectively, *vs* 16.8 points (intervention) and 12.4 points (control). This difference in scores could be explained by the available time for cadaver dissections. In Study IV, participants had one hour to complete their mastoidectomy while there was no time limit presented in Aussedat et al. Further, only a single rater assessed performances in Aussedat et al. while three blinded raters assessed performances in Study IV.

Another important difference between the two studies is the use of control groups. In Study IV the control group receives a comparable amount of training using VR simulation, while the control group in Aussedat et al. did not receive any training at all.[117,125] When comparable training tools are available, Cook et al. argue that such intervention *vs* no-intervention studies are "*similar to trials comparing a drug to no intervention when other effective drugs exists*".[126] In mastoidectomy training, the effect of VR simulation is already well-established[84] and we, therefore, decided to compare trainees using 3D-printed models before cadaver surgery, to historic controls who had received VR simulation training. We found that the group using 3D-printed models had a 29% better score than the historic VR controls. However, a direct comparison is limited for several reasons: Even though the historic controls had a comparable amount of training and baseline characteristics, and data were collected at previous runs of the same temporal bone drilling course (the Danish national temporal bone drilling course), the use of historic controls introduces a considerable risk of bias.[127] A randomized, controlled trial would be the preferable methodology for a direct comparison of two educational interventions. However, conducting a randomized, controlled trial was not possible due to the limited number of participants at each training course. Also, it is important to recognize that 3D-printed models and VR simulations have different strengths and weaknesses, supporting different aspects of learning.

Both 3D-printed models and VR simulations offer the possibility of an unlimited number of cases. In addition, VR simulators can offer the trainees learning supports such as integrated tutoring or automated feedback. This enables trainees to train self-directed, without the presence of an instructor.[57,128] Such self-directed learning, where trainees can regulate their own learning experience supported by instructional designs, is not only cost-effective but also beneficial for surgical performance.[129,130] Dynamic learning supports are not as easily implemented in training using 3D-printed models, but the physical models do offer training in microscope and otosurgical drill handling including a more realistic tactile feedback. Deliberate changes in the models could potentially introduce relevant learning supports for specific parts of the procedure. For example, enhancing the color contrast of the dura and sigmoid sinus could be used for identifying and smoothening the sino-dural angle. Also, coloring the septum of Koerner could guide novices when trying to access the mastoid antrum. Nevertheless, more immersive learning features such as gamification are currently reserved for technology-enhanced simulation such as VR simulation. The next step is therefore not necessarily to compare the two modalities but instead to explore the potential additive effects of the two training interventions combined.

Printing process

We designed the 3D-printed model to support learning at the early stages of temporal bone training, and as a supplement to cadaver surgery which has limited availability in training institutions throughout Europe.[10] We, therefore, aimed to create a 3D-printed model which accurately replicates the anatomy of the temporal bone while still keeping manufacturing costs low. We chose to use an extrusion-based printer (i.e., the Ender-3 , Creality, Shenzhen, China, with minor modifications) as this print-technology is easy to use, does not require special laboratory facilities and is low-cost compared to other more advanced print-technologies. The strengths of the current model are 1) The extrusion-based printer is a consumer-grade printer that can easily be set up at local training departments, including smaller institutions without technical backup. 2) Maintenance of the printer is limited and does not require technical support. 3) The manufacturing cost is ~15 USD per model which drastically lowers the financial barrier for repeated practice. Using other and more expensive print technologies, would make it feasible to create a model with a higher level of fidelity than our 3D-printed model. Intuitively, a higher level of model fidelity would result in better learning, but this is not necessarily true.[131–133] Hamstra et al. argue that the term *fidelity* should be completely abandoned as the interpretation of the term is diverse. Instead, *fidelity* should be replaced by more accurate descriptions such as *physical resemblance* and *functional task alignment*.[42] These terms are context-specific and requirements changes with the setting and level of the learner. Our model is aimed at the novice learner and requirements of physical resemblance could therefore be lower than for more advanced learners who would need a more accurate representation—for example, visual cues. Even the physical resemblance of cadaver training has limitations. Due to the lack of blood flow, there is no bleeding during surgery and the colors and textures of many anatomical landmarks change. For example, in real surgery, the vessels in the facial nerve sheet create a textured surface that helps identify the vertical part of the nerve. These vessels are not present in cadavers and training this part of the surgery is not possible. The same applies to our 3D-printed model as the lay-brick filament is not transparent. In contrast, VR simulation provides this and other visual features. Using a more transparent material, such as a natural ABS, could allow for some transparency in the 3D-printed model, and a yellow wire with thin lines of red color, representing vessels around the nerve, could offer the same visual cue as in VR training.

In Study I, we found that several 3D-printing technologies can be suitable for printing a realistic temporal bone model. One example of a commercially available model is the Phacon Temporal Bone (Phacon GmbH, Leipzig, Germany). The Phacon model is printed using a Binder Jetting 3D-printer (Z510; 4D-concepts, Gross-Gerau, Germany)[134] and costs ~150 USD. The model is printed in three pieces as unhardened material gets entrapped within the air cells of the mastoid and needs to be removed manually. After inserting a wire for representing the facial nerve and sigmoid sinus, the pieces are glued together and hardened with a polyurethan mixture.[134,135] Wanibuchi et al. and Takahasi et al. report similar problems with the accumulation of residual printing material in the mastoid air cells when using a Powder Bed Fusion and Binder Jetting printer, respectively.[112,136] While Wanibuchi et al. and Phacon solved this problem by printing the model in several individual pieces, Takahasi et al. designed small drainage holes in the model so that the excess material could be removed after printing.[112] Using an extrusion-based printer directly solves this problem, and the model can be printed in one piece without residual material being entrapped. This lowers the need for manual postprocessing which represents a considerable share of the manufacturing cost per model (Study II).[119] However, filaments for extrusion-based printers are thermoplastics which are prone to melting under heat, while thermoset plastics retain their shape after curing. Consequently, melting will occur when drilling models that are printed on extrusion-based printers while this does not apply for printers using thermoset materials (e.g., Binder Jetting and Powder Bed Fusion printers). For addressing this problem, we used a PLA filament with a high load of chalk leading to better resistance from melting due to thermal friction compared with other thermoplastics (Study II). Altogether, this highlights that creating a 3D-printed model for temporal bone training requires careful considerations of technical aspects of manufacturing, price, stakeholders, and targeted level of learners.

Performance assessment and standard setting in mastoidectomy training

In modern surgical education, assessment is pivotal for offering trainees competency-based training, and the dogma *"assessment drives learning"* emphasizes the need for valid and reliable assessment. Traditionally, residents' surgical competence has been assessed by unstructured observations in the clinic. This approach lacks objective measures and a true picture of actual surgical skills is confounded.[137] Structured assessment using dedicated tools addresses this problem. Contrary to other fields, there are several assessment tools supported by modern validity evidence for the mastoidectomy procedure.[84] However, as is the case in the remaining assessment literature, exploration of Messick's 5th source of validity; *Consequences of testing* is lacking.[67,84] Exploring this source includes analyzing the impact of assessment for example by standard setting for pass/fail decisions. Standard setting helps educators create a benchmark for sufficient trainee competency and is therefore key in simulation-based training. Nevertheless, establishing a credible pass/fail level is challenging as surgery is complex and does not only rely on the individual's technical competence alone, but also on non-technical skills. It is therefore imperative to distinguish between surgical proficiency and technical competence, which are often used interchangeably in the literature.[138] Recently, Pietersen et al. conducted a systematic review on standard setting in simulation-based training of surgical procedures.[81] Here they found that several methods are being used for standard setting, but most frequently the standard is set using the mean/median performance of experienced surgeons which is not a recognized method in the literature.[81] In Study III, we proposed a pass/fail score of 21 out of 25 points using the contrasting groups method. If we had used the mean/median expert score, the pass/fail would have been set to 23 points. Using this method for standard setting is problematic as about half of the experts who regularly perform the procedure in real life, and who are considered highly competent, would fail. At the current level (21 out of 25 points), over 2–3 mastoidectomies, all the trainees would fail and all experts would pass. This sets a realistic and achievable standard for mastoidectomy performance on 3D-printed models which is slightly higher, but still comparable to, the pass/fail score of 19.5/26 WS points established in VR temporal bone simulation training.[139]

Another attempt at standard setting in mastoidectomy training used a modified approach of the Angoff method for the Cross-Institutional Mastoidectomy Assessment Tool (CIMAT).[140,141] The Angoff method is an item-based approach, originally created for written assignments, where expert assessors set the standard based on the minimum criteria that would result in a passing performance.[65] This means that the standard is based on a "just adequate" performance. This is a problem because simulation-based training facilitates mastery learning where the goal is to achieve a higher competence level than the minimum expected.[65] In Study III we, therefore, used an examinee-based approach (the contrasting groups method) for setting a standard exceeding the level of "just adequate". However, the contrasting groups method has a major limitation which the item-centered methods such as the Angoff method address. When using a summative score, shortcomings to fulfill the minimum standard of key elements can be missed. A specific example of the mastoidectomy procedure could be damage to the facial nerve. A such mistake would lead to a serious adverse effect and is not compatible with a passing performance. Nevertheless, damaging the facial nerve only results in a 1-point reduction on the Welling Scale and it is still very much achievable to reach the 21 points passing score. This problem can be solved by assigning an independent overall (global) score of "Pass", "Fail", and "Borderline". This was done in the CISAT assessment tool for cochlear implantation developed by our group.[142] Consequently, the pass/fail-level introduced in Study III should not be seen as an absolute threshold for technical competence in mastoidectomy, but instead as a relevant benchmark in the early acquisition of skills.

Generalizability of assessment

In Study III and V, we used Generalizability theory (G theory) for exploring the reliability of ratings on the 3D-printed models. Both studies included performances from eighteen ORL residents who were at the early stage of learning to perform a mastoidectomy. In addition, Study III included performances from eleven experienced otosurgeons. The overall G coefficient was 0.91 in Study III and 0.80 in Study IV. Both levels correspond to a high level of reliability but the relative difference emphasizes the fact that reliability is context specific.[143] A recent study on the reliability of VR mastoidectomy performances found that reliability was highly influenced by the experience of the learner.[144] In Study III, we included both experienced otosurgeons and novices in the reliability analysis. Comparing groups with such different levels of skills will *"erroneously inflate the reliability coefficient*".[145] Consequently, Study III might overestimate the reliability level in the context of novice training. While both studies still demonstrate a high level of reliability this is still a relevant observation. If using 3D-printed models for training other types of learners (e.g., intermediates) the reliability levels can be affected. This could influence the number of observations, or raters, needed for reaching a sufficient level of reliability having consequences for high-stakes assessments such as certification.

Implications for temporal bone training

The findings of this thesis can be used by training departments to develop a competency-based training program for temporal bone surgery using 3D-printed models. In Study II we presented a workflow for printing the 3D-printed model using entry-level and cost-effective print technologies. In every part of the printing process, there are several barriers for training institutions to start their production. Study II addresses this by offering a thorough technical description and recommendations for every aspect of printing an inexpensive training model. The 3D-printed model is based on a digitized temporal bone from the OpenEar library which comprises seven additional digitized bones. The 3D-printing routine can easily be adopted for the additional specimens. This opens the possibility of creating a total of eight different 3D-printed models, offering an even more diverse training opportunity for novices learning to perform a mastoidectomy. Future developments could lead to a pipeline for 3D-printing based on clinical imaging datasets such as CT or CBCT, making it feasible to print patient-specific temporal bone models that can be used for pre-operative rehearsal and surgical planning.

Study IV supports that skills acquired using 3D-printed models transfer to cadaver surgery. Nevertheless, the residents in Study IV did not achieve a WS score reflecting a safe mastoidectomy performance. This emphasizes that the limited training time provided in Study IV is not enough to obtain a sufficient skill level and more practice is needed. The question is: *How much training is required?* The pass/fail score established in Study III could serve as a guideline for this, setting a standard for when a trainee is ready to move on to use cadavers for further skills refinement. This would ensure the most optimal use of the limited cadaver resources.[10] Such "certification" has been successfully implemented in simulation-based training of cataract surgery[73] and could, with further evaluation, become relevant in temporal bone surgery.

Study V has implications for decentralized mastoidectomy training. In general, a high level of reliability is key in competency-based training as the assessment has consequences for the trainees' surgical progression. Reliable video-based assessment enables trainees to receive formative feedback even when an expert assessor is not physical present. This expands training opportunities and convenience.

Strengths and limitations

Designing a study can be challenging and often strengths and limitations mirror the compromise between rigorous methodical design aspirations and what is feasible. In Study I, the systematic approach and evaluation of educational outcomes using contemporary frameworks represent a major strength. However, the modest evidence on optimal printing techniques and educational effectiveness limits conclusions on the best practice. The remaining studies (II–IV) share several strengths. One major strength is that we used residents at the relevant stage of their training. Reliability has shown to be highly context-specific[144] and including residents in the studies support conclusions on reliability levels in the early mastoidectomy training using 3D-printed models. Further, using blinded raters for assessment also represents major strengths.

A general limitation of the studies is the small sample size. We included the drillings of eighteen Danish residents (novices), eleven experienced otosurgeons (experts) and model evaluation by eleven Swiss residents (novices and advanced beginners). It requires a large number of microscopes and drilling equipment for testing the transfer of skills from 3D-printed models to cadaver surgery in a controlled setting. Such a setup is costly and only available at temporal bone drilling courses. Also, the number of otosurgeons who perform mastoidectomies regularly in Denmark is limited.

Due to logistical challenges, we did not succeed at including international otosurgeons in Study IV. Consequently, the included number of participants represents samples of convenience deemed sufficient for the research purpose. Specifically, in Study IV, we chose to use a historical cohort as a control for minimizing the risk of a type 2 error. This does, however, introduce problems related to cohort differences. Even though baseline characteristics are similar, and all groups where completely self-directed when drilling, differences in the group dynamics and natural differences occurring between the groups can affect the internal validity.[127,146]

Another limitation is that we only included technical skills as an outcome in our studies. In reality, other essential skills such as decision-making and communication are fundamental for safe practice. Further, the Welling Scale (WS) assessment tool only concerns the final dissection result and not the drilling process. We solely used the WS in our studies; however, other assessment tools include the dynamic aspects of the procedure which could affect the reliability of the assessment.

Research perspectives

Deciding on the best material for creating a 3D-printed temporal bone model comes with important considerations on price, print technologies, knowledge and available lab facilities. In Study II, we found the Laybrick filament to be cost-effective and have characteristics mimicking the temporal bone. However, the filament still has three major shortcomings: 1) Even though it is more heat resistant than most other filaments, some melting still occurs when drilling; 2) The Laybrick-filament is fragile and breaks easily during printing, which results in a significant number of unsuccessful prints and wear of the 3D-printer nozzle; 3) The material is opaque which differs from real bone which has some degree of transparency. An experienced otosurgeon chose the Laybrick-filament from eight material samples found by 3D-printing experts. We did not perform any systematic characterization of temporal bone properties or the plastics tested. Future work should therefore aim to identify the most suitable plastic for mimicking the temporal bone, while still being compatible with extrusion-based printers, using a more systematic approach and broader scoping of candidate materials.

In our setting, training took place during a single day and participants only drilled 2–3 models each. In VR temporal bone training novices typically reaches a learning plateau after 4–9 procedures depending on the simulator and practice distribution.[147,148] Measuring learning (y-axis) as a function of effort (x-axis, e.g., time, number of repetitions etc.) can be used for mapping when learning is most effective. Such a learning curve can estimate the practice needed and provide a powerful tool for supporting self-regulated learning.[149,150] While learning curves in VR training of mastoidectomy are well-known,[147,148] this does not apply to training using 3D-printing models. Consequently, it is unknown when trainees reach the learning plateau. Gustaffson et al. used the learning curve plateau in simulation-based training of osteosynthesis for setting a pass/fail score.[151] Mapping the learning curves for training using 3D-printed models could therefore be useful for designing a training curriculum supporting the principles of mastery learning.

Finally, there is a gap in knowledge on how 3D-printed models can be used in other aspects of temporal bone surgery. Several studies have reported using 3D-printed models for presurgical rehearsal[17,114,152–157] but no study has provided solid evidence supporting the value of 3D-printed models in presurgical planning and/or rehearsal.[158] Future research should answer key questions on the specific applicability of 3D-printed models for presurgical rehearsal, including intended users, technical requirements and how these models affect surgical outcomes.

Conclusion

3D-printing a cost-effective model for temporal bone training is feasible using inexpensive, consumer-grade 3D-printers. Using extrusion-based print-technologies, we created a 3D-printed model that can provide a realistic drilling experience for novice learners and be printed locally at clinical departments. We found that mastoidectomy skills obtained during training on 3D-printed models transfer to cadaveric dissection, highlighting that 3D-printed models are indeed an effective training tool. According to Messick's framework, we systematically gathered validity evidence supporting the use of the model for mastoidectomy training. We found assessments of the drillings to be highly reliable both when performed physically or based on video-recordings. This broadens the opportunities for competency-based training, testing, and objective feedback. Further, we established a pass/fail score of 21 out of 25 Welling Scale points to support competency-based training and mastery learning.

This thesis bridges essential knowledge gaps in the educational literature for temporal bone training and represents a considerable step towards creating evidence-based training using 3D-printed temporal bone models.

References

1. *World Report on Hearing: Executive Summary. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO*.

2. Chern A, Golub JS. Age-related hearing loss and dementia Alexander. *Alzheimer Dis Assoc Disord*. 2019;33(3):286-290.

3. Gurgel RK, Ward PD, Schwartz S, Norton MC, Foster NL, Tschanz JT. Relationship of Hearing loss and Dementia: a Prospective, Population-based Study. *Otology and neurotology*. 2015;35(5):775-781. doi:10.1097/MAO.0000000000000313.Relationship

4. Rutherford BR, Brewster K, Golub JS, Kim AH, Roose SP. Sensation and Psychiatry : Linking Age-Related Hearing Loss to Late-Life Depression and Cognitive Decline. *American Journal of Psychiatry*. 2018;(March):215-224. doi:10.1176/appi.ajp.2017.17040423

5. Fitts PM, Posner MI. *Human Performance. (Brooks/Cole).*; 1967.

6. Cook DA, Hatala R, Brydges R, et al. Technology-Enhanced Simulation for Health Professions Education A Systematic Review and Meta-analysis. *JAMA J Am Med Assoc*. 2011;306(9):978-988. doi:10.1001/jama.2011.1234

7. Favier V, Ayad T, Blanc F, Fakhry N, Andersen SAW. Use of simulation-based training of surgical technical skills among ENTs: an international YO-IFOS survey. *European Archives of Oto-Rhino-Laryngology*. 2021;278(12):5043-5050. doi:10.1007/s00405-021-06846-x

8. Wiet GJ, Sørensen MS, Andersen SAW. Otologic Skills Training. *Otolaryngol Clin North*. 2017;50(5):933-945. doi:10.1016/j.otc.2017.05.005.Otologic

9. Bashankaev B, Baido S, Wexner SD. Review of available methods of simulation training to facilitate surgical education. *Surg Endosc*. 2011;25(1):28-35. doi:10.1007/s00464-010-1123-x

10. Frithioff A, Sørensen MS, Andersen SAW. European status on temporal bone training: a questionnaire study. *European Archives of Oto-Rhino-Laryngology*. 2018;275(2):357-363. doi:10.1007/s00405-017-4824-0

11. Mills R, Phil M, Eng FRCS, Ed FRCS, Lee P. Review Articles Surgical skills training in middle-ear surgery. 2003;117(October 2002):159-163.

12. Drake-Lee A. Structured training of ENT specialist registrars in the out-patient clinic and theatre. *Clin Otolaryngol Allied Sci*. 2002;27(5):396-402. doi:10.1046/j.1365-2273.2002.00603.x

13. Bhutta MF. A review of simulation platforms in surgery of the temporal bone. *Clinical Otolaryngology*. 2016;41(5):539-545. doi:10.1111/coa.12560

14. Lui JT, Hoy MY. Evaluating the Effect of Virtual Reality Temporal Bone Simulation on Mastoidectomy Performance: A Meta-analysis. *Otolaryngology-Head and Neck Surgery*. 2017;156(6):1018-1024. doi:10.1177/0194599817698440

15. Kashikar TS, Kerwin TF, Moberly AC, Wiet GJ. A Review of Simulation Applications in Temporal Bone Surgery. *Laryngoscope Investigate Otolaryngology*. 2019;4(4):420-424. doi:10.1002/lio2.277

16. Rose AS, Kimbell JS, Webster CE, Harrysson OLA, Formeister EJ, Buchman CA. Multi-material 3D Models for Temporal Bone Surgical Simulation. *Annals of Otology, Rhinology and Laryngology*. 2015;124(7):528-536. doi:10.1177/0003489415570937

17. Mukherjee P, Cheng K, Flanagan S, Greenberg S. Utility of 3D printed temporal bones in pre-surgical planning for complex BoneBridge cases. *Eur Arch Otorhinolaryngol*. 2017;274(8 PG-3021-3028):3021-3028. doi:10.1007/s00405-017-4618-4

18. Mowry S, Jammal H, Myer C, Solares CA, Weinberger P. A Novel Temporal Bone Simulation Model Using 3D Printing Techniques. *Otol Neurotol*. 2015;36:1562-1565.

19. da Cruz M, Francis HW. Face and content validation of a novel three-dimensional printed temporal bone for surgical skills development. *J Laryngol Otol*. 2015;129(Suppl. S3):S23-S29. doi:10.1017/S0022215115001346
20. Bennett M, Warren F, Haynes D. Indications and Technique in Mastoidectomy. *Otolaryngol Clin North Am*. 2006;39(6):1095-1113. doi:10.1016/j.otc.2006.08.012
21. Theissing J, Rettinger G, Werner J. *ENT—Head and Neck Surgery: Essential Procedures*. Thieme; 2010.
22. Martin M, Raz Y. Mastoid surgery. In: *Operative Otolaryngology: Head and Neck Surgery*. Second. Elsevier; 2008:1163-1176. doi:10.1007/BF03047153
23. Hohman MH, Bhama PK, Hadlock TA. Epidemiology of iatrogenic facial nerve injury: A decade of experience. *Laryngoscope*. 2014;124(1):260-265. doi:10.1002/lary.24117
24. Hamdorf JM, Hall JC. Acquiring surgical skills. *British Journal of Surgery*. 2000;87:28-37.
25. Reznick RK, MacRae H. Teaching Surgical Skills — Changes in the Wind. *N Engl J Med*. 2006;355:2664-2669. doi:10.1021/jf052913o
26. Rajaratnam V, Rahman NA, Dong C. Integrating instructional design principles into surgical skills training models: an innovative approach. *Ann R Coll Surg Engl*. 2021;103(10):718-724. doi:10.1308/rcsann.2020.7132
27. Kulasegaram KM, Grierson LEM, Norman GR. The roles of deliberate practice and innate ability in developing expertise: Evidence and implications. *Med Educ*. 2013;47(10):979-989. doi:10.1111/medu.12260
28. Ericsson KA. Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*. 2008;15(11):988-994. doi:10.1111/j.1553-2712.2008.00227.x

29. Ericsson KA. Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Academic Medicine*. 2004;79(Supplement):S70-S81. doi:10.1097/00001888-200410001-00022

30. Mitchell SA, Boyer TJ. *Deliberate Practice in Medical Simulation*.; 2022.

31. Smith S, Lonie J. Mastery learning : how is it helpful ? An analytical review. *Adv in Med Educ and Prac*. 2017;8:269-275.

32. Block J, Burns R. Mastery learning. *Rev Res Educ*. 1976;4:3-49.

33. Felix HM, Schertzer K. *Mastery Learning in Medical Simulation*.; 2022.

34. Kirkpatrick DL, Kirkpatrick JD. *Evaluating Training Programs: The Four Levels.* 1st ed. Berrett-Koehler; 1994.

35. Hammick M. Interprofessional education : evidence from the past to guide the future. *Med Teach*. 2000;22(5):461-468.

36. Cook DA, Reed DA. Appraising the Quality of Medical Education Research Methods: The Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale-Education. *Academic Medicine*. 2015;90(8):1067-1076. doi:10.1097/ACM.0000000000000786

37. Sharma R, Gordon M, Dharamsi S, Gibbs T. Systematic reviews in medical education: A practical approach: AMEE Guide 94. *Med Teach*. 2015;37(2):108-124. doi:10.3109/0142159X.2014.970996

38. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association Between Funding and Quality of Published Medical Education Research. *JAMA*. 2007;298(9):1002-1009. https://jamanetwork.com/

39. Allen LM, Hay M, Palermo C. Evaluation in health professions education—Is measuring outcomes enough? *Med Educ*. 2022;56(1):127-136. doi:10.1111/medu.14654

40. Yardley S, Dornan T. Kirkpatrick's levels and education "evidence." *Med Educ*. 2012;46(1):97-106. doi:10.1111/j.1365-2923.2011.04076.x

41. Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A systematic review of skills transfer after surgical simulation training. *Ann Surg*. 2008;248(2):166-179. doi:10.1097/SLA.0b013e318176bf24

42. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. *Academic Medicine*. 2014;89(3):387-392. doi:10.1097/ACM.0000000000000130

43. Dawe SR, Pena GN, Windsor JA, et al. Systematic review of skills transfer after surgical simulation-based training. *British Journal of Surgery*. 2014;101(9):1063-1076. doi:10.1002/bjs.9482

44. Andersen SAW, Foghsgaard S, Cayé-Thomasen P, Sørensen MS. The Effect of a Distributed Virtual Reality Simulation Training Program on Dissection Mastoidectomy Performance. *Otology and neurotology*. 2018;39:1277-1284. doi:10.1097/MAO.0000000000002031

45. George AP, De R. Review of temporal bone dissection teaching: How it was, is and will be. *Journal of Laryngology and Otology*. 2010;124(2):119-125. doi:10.1017/S0022215109991617

46. Fennessy BG, O'Sullivan P. Establishing a temporal bone laboratory: Considerations for ENT specialist training. *Ir J Med Sci*. 2009;178(4):393-395. doi:10.1007/s11845-009-0373-x

47. Frendø M, Thingaard E, Konge L, Sølvsten M, Steven S. Decentralized virtual reality mastoidectomy simulation training : a prospective , mixed - methods study. *European Archives of Oto-Rhino-Laryngology*. 2019;276(10):2783-2789. doi:10.1007/s00405-019-05572-9

48. Javia L, Deutsch ES. A systematic review of simulators in otolaryngology. *Otolaryngology - Head and Neck Surgery*. 2012;147(6):999-1011. doi:10.1177/0194599812462007

49. Wiet GJ, Stredney D, Kerwin T, et al. Virtual Temporal Bone Dissection System : Development and Testing. *Laryngoscope*. 2012;122:S1-S12. doi:10.1002/lary.22499

50. 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. doi:10.1097/MLG.0b013e3181671b15

51. Chan S, Li P, Lee DH, Salisbury JK, Blevins NH. *A Virtual Surgical Environment for Rehearsal of Tympanomastoidectomy*.

52. Arora A, Khemani S, Tolley N, et al. Face and content validation of a virtual reality temporal bone simulator. *Otolaryngology - Head and Neck Surgery*. 2012;146(3):497-503. doi:10.1177/0194599811427385

53. 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. *Otology and neurotology*. Published online 2009:484-487.

54. Compton EC, Agrawal SK, Ladak HM, et al. Assessment of a virtual reality temporal bone surgical simulator: A national face and content validity study. *Journal of Otolaryngology - Head and Neck Surgery*. 2020;49(1). doi:10.1186/s40463-020-00411-y

55. Davaris M, Wijewickrema S, Zhou Y, et al. The Importance of Automated Real-Time Performance Feedback in Virtual Reality Temporal Bone Surgery Training BT - Artificial Intelligence in Education. In: Isotani S, Millán E, Ogan A, Hastings P, McLaren B, Luckin R, eds. Springer International Publishing; 2019:96-109.

56. Andersen SAW, Mikkelsen PT, Sørensen MS. The Effect of Simulator-Integrated Tutoring for Guidance in Virtual Reality Simulation Training. *Simulation in Healthcare*. 2020;15(3):147-153. doi:10.1097/SIH.0000000000000414
57. Frithioff A, Frendo M, von Buchwald JH, Mikkelsen PT, Sorensen MS, Andersen SAW. Automated summative feedback improves performance and retention in simulation training of mastoidectomy: a randomised controlled trial. *Journal of Laryngology and Otology*. 2022;136(1):29-36. doi:10.1017/S0022215121003352
58. Begall K, Vorwerk U. Artificial petrous bone produced by stereolithography for microsurgical dissecting exercises. *ORL*. 1998;60:241-245. doi:10.1159/000027604
59. Chen G, Jiang M, Mansour K, Chuen J, Amott D. Three-dimensional printing as a tool in otolaryngology training : a systematic review. *J Laryngol Otol*. 2020;134:14-19.
60. Crafts TD, Ellsperman SE, Wannemuehler TJ, Bellicchi TD, Shipchandler TZ, Mantravadi A v. Three-Dimensional Printing and Its Applications in Otorhinolaryngology– Head and Neck Surgery. *Otolaryngol Head Neck Surg*. 2017;156(6):999-1010. doi:10.1177/0194599816678372
61. Reznick RK. Teaching and Testing Technical Skills. *Surgical Education*. 1993;165(March).
62. Szalay D, Macrae H, Regehr G, Reznick R. *Using Operative Outcome to Assess Technical Skill*. Vol 180.; 2000.
63. Shepard LA. The Role of Assessment in a Learning Culture. *Educational Researcher*. 2000;29(7):4-14. doi:10.3102/0013189X029007004
64. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Advances in Simulation*. 2016;1(1):1-12. doi:10.1186/s41077-016-0033-y
65. Downing SM, Yudkowsky R. *Assessment in Health Professions Education*. Routledge; 2009.

66. Downing SM. Validity: On the meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837. doi:10.1046/j.1365-2923.2003.01594.x

67. Borgersen NJ, Naur TMH, Sørensen SMD, et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. *Ann Surg*. 2018;267(6):1063-1068. doi:10.1097/SLA.0000000000002652

68. Ghaderi I, Manji F, Soo Park Y, et al. Technical skills assessment toolbox a review using the unitary framework of validity. *Ann Surg*. 2015;261(2):251-262. doi:10.1097/SLA.0000000000000520

69. Messick S. *Validity. In R. L. Linn (Ed.)*. 3rd ed. American Council on education and Macmillan; 1989.

70. Cook DA, Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments : Theory and Application. *American Journal of Medicine*. 2006;119:166.e7-166.e16. doi:10.1016/j.amjmed.2005.10.036

71. Testing JC on the S for E and. *Standards for Educational and Psychological Testing*.; 1999.

72. Schuwirth LWT, Van Der Vleuten CPM. General overview of the theories used in assessment: AMEE Guide No. 57. *Med Teach*. 2011;33(10):783-797. doi:10.3109/0142159X.2011.611022

73. Thomsen ASS, Kiilgaard JF, Kjærbo H, La Cour M, Konge L. Simulation-based certification for cataract surgery. *Acta Ophthalmol*. 2015;93(5):416-421. doi:10.1111/aos.12691

74. Bloch R, Norman G. Generalizability theory for the perplexed : A practical introduction and guide : AMEE Guide No . 68. *Med Teach*. 2012;34:960-992. doi:10.3109/0142159X.2012.703791

75. Bilgic E, Watanabe Y, Mckendy KM, Ito Y. Reliable Assessment of Performance in Surgery : A Practical Approach to Generalizability Theory. *J Surg Educ*. 2015;72(5):774-775. doi:10.1016/j.jsurg.2015.04.020
76. de Montbrun S, Satterthwaite L, Grantcharov TP. Setting pass scores for assessment of technical performance by surgical trainees. *British Journal of Surgery*. 2016;103(3):300-306. doi:10.1002/bjs.10047
77. Goldenberg MG, Garbens A, Szasz P, Hauer T, Grantcharov TP. Systematic review to establish absolute standards for technical performance in surgery. *British Journal of Surgery*. 2017;104(1):13-21. doi:10.1002/bjs.10313
78. Ben-David M. AMEE Guide No. 18: Standard setting in student assessment. *Med Teach*. 2000;22(2):120-130.
79. Yudkowsky R, Park YS, Lineberry M, Knox A, Ritter EM. Setting Mastery Learning Standards. *Academic Medicine*. 2015;90(11):1495-1500. doi:10.1097/ACM.0000000000000887
80. McKinley DW, Norcini JJ. How to set standards on performance-based examinations: AMEE Guide No. 85. *Med Teach*. 2014;36(2):97-110. doi:10.3109/0142159X.2013.853119
81. Pietersen PI, Bjerrum F, Tolsgaard MG, Konge L, Andersen SAW. Standard Setting in Simulation-based Training of Surgical Procedures: A Systematic Review. *Ann Surg*. 2022;275(5):872-882. doi:10.1097/SLA.0000000000005209
82. Jørgensen M, Konge L, Subhi Y. Contrasting groups ' standard setting for consequences analysis in validity studies : reporting considerations. *Advances in simulation*. 2018;3(5):1-7.
83. Cook DA, Lineberry M. Consequences validity evidence: Evaluating the impact of educational assessments. *Academic Medicine*. 2016;91(6):785-795. doi:10.1097/ACM.0000000000001114

84. Sethia R, Kerwin TF, Wiet GJ. Performance Assessment for Mastoidectomy: State of the Art Review. *Otolaryngology - Head and Neck Surgery (United States)*. 2017;156(1):61-69. doi:10.1177/0194599816670886
85. Stavrakas M, Menexes G, Triaridis S, Bamidis P, Constantinidis J, Karkos PD. Objective structured assessment of technical skill in temporal bone dissection: Validation of a novel tool. *Journal of Laryngology and Otology*. 2021;135(6):518-528. doi:10.1017/S0022215121001201
86. 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. doi:10.1177/0163278708324444
87. Talks BJ, Lamtara J, Wijewickrema S, Gerard JM, Mitchell-Innes AM, O'Leary S. The Melbourne Mastoidectomy Scale: Validation of an end-product dissection scale for cortical mastoidectomy. *Clinical Otolaryngology*. 2020;45(5):746-753. doi:10.1111/coa.13569
88. Zirkle M, Taplin MA, Anthony R, Dubrowski A. Objective Assessment of Temporal Bone Drilling Skills. *Annals of Otology, Rhinology and Laryngology*. 2007;116(11):793-798.
89. Pisa J, Gousseau M, Mowat S, Westerberg B, Unger B, Hochman JB. Simplified Summative Temporal Bone Dissection Scale Demonstrates Equivalence to Existing Measures. *Annals of Otology, Rhinology and Laryngology*. 2018;127(1):51-58. doi:10.1177/0003489417745090
90. Hochman JB, Pisa J, Singh S, Gousseau M, Unger B. Comparison of Summative Temporal Bone Dissection Scales Demonstrate Equivalence. *Int Arch Otorhinolaryngol*. Published online 2022:556-560. doi:10.1055/s-0041-1740162
91. Butler NN, Wiet GJ. Reliability of the Welling Scale (WS1) for Rating Temporal Bone Dissection Performance. *Laryngoscope*. 2007;117:1803-1808. doi:10.1097/MLG.0b013e31811edd7a

92. Francis HW, Masood H, Chaudhry KN, et al. *Objective Assessment of Mastoidectomy Skills in the Operating Room*.; 2010.

93. Ilgen JS, Ma IWY, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ*. 2015;49(2):161-173. doi:10.1111/medu.12621

94. Andersen SAW, Cayé-Thomasen P, Sørensen MS. Mastoidectomy performance assessment of virtual simulation training using final-product analysis. *Laryngoscope*. 2015;125(2):431-435. doi:10.1002/lary.24838

95. Wohlers T, Gornet T. History of Additive Manufacturing. *Wohlers report*. Published online 2016:1-24. doi:10.4018/978-1-5225-2289-8.ch001

96. Su A, Al'Aref SJ. History of 3D printing. In: *3D Printing Applications in Cardiovascular Medicine*. Elsevier Inc.; 2018:1-10. doi:10.1016/B978-0-12-803917-5.00001-8

97. Langridge B, Cantab BA, Momin S, et al. Systematic Review of the Use of 3-Dimensional Printing in Surgical Teaching and Assessment. *J Surg Educ*. 2018;75(1):209-221.

98. Horn TJ, Harrysson OLA. Overview of current additive manufacturing technologies and selected applications. *Sci Prog*. 2012;95(3):255-282. doi:10.3184/003685012X13420984463047

99. Garcia J, Yang Z, Mongrain R, Leask RL, Lachapelle K. 3D printing materials and their use in medical education : a review of current technology and trends for the future. *BMJ Stel*. 2018;4:27-40. doi:10.1136/bmjstel-2017-000234

100. Ngo TD, Kashani A, Imbalzano G, Nguyen KTQ, Hui D. Additive manufacturing (3D printing): A review of materials, methods, applications and challenges. *Compos B Eng*. 2018;143:172-196. doi:10.1016/j.compositesb.2018.02.012

101. N. Turner B, Strong R, A. Gold S. A review of melt extrusion additive manufacturing processes: I. Process design and modeling. *Rapid Prototyp J*. 2014;20(3):192-204. doi:10.1108/RPJ-01-2013-0012

102. Dey A, Yodo N. A systematic survey of FDM process parameter optimization and their influence on part characteristics. *Journal of Manufacturing and Materials Processing*. 2019;3(3). doi:10.3390/jmmp3030064

103. Oleff A, Küster B, Stonis M, Overmeyer L. Process monitoring for material extrusion additive manufacturing: a state-of-the-art review. *Progress in Additive Manufacturing*. 2021;6(4):705-730. doi:10.1007/s40964-021-00192-4

104. Otton JM, Birbara NS, Hussain T, Greil G, Foley TA, Pather N. 3D printing from cardiovascular CT: A practical guide and review. *Cardiovasc Diagn Ther*. 2017;7(5):507-526. doi:10.21037/cdt.2017.01.12

105. Kaye R, Goldstein T, Zeltsman D, Grande DA, Smith LP. Three dimensional printing : A review on the utility within medicine and otolaryngology. *Int J Pediatr Otorhinolaryngol*. 2016;89:145-148. doi:10.1016/j.ijporl.2016.08.007

106. Leung G, Pickett AT, Bartellas M, et al. Systematic review and meta-analysis of 3D-printing in otolaryngology education. *Int J Pediatr Otorhinolaryngol*. 2022;155(February):111083. doi:10.1016/j.ijporl.2022.111083

107. Freiser ME, Ghodadra A, Mccall AA, Shaffer AD, Magnetta M, Jabbour N. Operable, Low-Cost, High-Resolution Patient-Specific 3D Printed Temporal Bones for Surgical Simulation and Evaluation. *Annals of Otology, Rhinology and Laryngology*. Published online 2021:1-8. doi:10.1177/0003489421993733

108. Mooney MA, Cavallo C, Zhou JJ, et al. Three-Dimensional Printed Models for Lateral Skull Base Surgical Training : Anatomy and Simulation of the Transtemporal Approaches. *Oper Neurosurg*. 2019;18(2):193-201. doi:10.1093/ons/opz120

109. Ospina J, Fandino A, Hernandez S, et al. 3D-printed pediatric temporal bone models for surgical training: A patient-specific and low-cost alternative. *J 3D Print Med*. 2019;3(3 PG-135-143):135-143. doi:http://dx.doi.org/10.2217/3dp-2019-0011

110. Longfield EA, Brickman TM, Jeyakumar A. 3D Printed Pediatric Temporal Bone : A Novel Training Model. *Otol Neurotol*. 2015;36:793-795.

111. Bone TM, Mowry SE. Content Validity of Temporal Bone Models Printed Via Inexpensive Methods and Materials. *Otol Neurotol*. 2016;37:1183-1188. doi:10.1097/MAO.0000000000001131

112. Takahasi K, Morita Y, Ohshima S, et al. Creating an Optimal 3D Printed Model for Temporal Bone Dissection Training. *Annals of Otology, Rhinology and Laryngology*. 2017;126(7):530-536. doi:10.1177/0003489417705395

113. Haffner M, Quinn A, TY H, EB S, Steele T. Optimization of 3D Print Material for the Recreation of Patient-Specific Temporal Bone Models. *Ann Otol Rhinol Laryngol*. 2018;127(5):338-343. doi:10.1177/0003489418764987

114. Rose AS, Webster CE, Harrysson OLA, Formeister EJ, Rawal RB, Iseli CE. Pre-operative simulation of pediatric mastoid surgery with 3D-printed temporal bone models. *Int J Pediatr Otorhinolaryngol*. 2015;79(5):740-744. doi:10.1016/j.ijporl.2015.03.004

115. Unger B, Kraut J, Rhodes C, Hochman J. Design and Validation of 3D Printed Complex Bone Models with Internal Anatomic Fidelity for Surgical Training and Rehearsal. *Stud Health Technol Inform*. 2014;196(PG-439-45):439-445.

116. Bento RF, Rocha BA, Freitas E, Balsalobre F. Otobone ® : Three-dimensional printed Temporal Bone Biomodel for Simulation of Surgical Procedures. *Int Arch Otorhinolaryngol*. 2019;23(4):451-454.

117. Frithioff A, Frendø M, Pedersen DB, Sørensen MS, Wuyts Andersen SA. 3D-Printed Models for Temporal Bone Surgical Training: A Systematic Review. *Otolaryngology - Head and Neck Surgery*. 2021;165(5):617-625. doi:10.1177/0194599821993384

118. Mowry SE, Jabbour N, Rose AS, et al. Multi-institutional Comparison of Temporal Bone Models: A Collaboration of the AAO-HNSF 3D-Printed Temporal Bone Working Group. *Otolaryngology - Head and Neck Surgery*. 2020;Epub ahead. doi:10.1177/0194599820960474

119. Ravi P, Burch MB, Farahani S, et al. Utility and Costs During the Initial Year of 3D Printing in an Academic Hospital. *J Am Coll Radiol*. 2023;20(2):193-204. doi:10.1016/j.jacr.2022.07.001

120. Sieber D, Erfurt P, John S, et al. The openEar library of 3D models of the human temporal bone based on computed tomography and micro-slicing. *Sci Data*. 2019;6:1-9. doi:10.1038/sdata.2018.297

121. Sieber D, Andersen SAW, Sørensen MS, Mikkelsen PT. OpenEar image data enables case variation in high fidelity virtual reality ear surgery. *Otol Neurotol*. 2021;[published. doi:10.1097/MAO.0000000000003175

122. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses : The PRISMA Statement. *PLoS Med*. 2009;6(7). doi:10.1371/journal.pmed.1000097

123. https://www.rigshospitalet.dk/temporal-bone-imaging-and-simulation-research-group.

124. Frithioff A, Frendø M. Are Video Recordings Reliable for Assessing Surgical Performance ? A Prospective Reliability Study Using Generalizability Theory. *Simul Healthc*. 2022;online ahe:1-7. doi:10.1097/SIH.0000000000000672

125. Aussedat C, Venail F, Nguyen Y, Lescanne E, Marx M, Bakhos D. Usefulness of temporal bone prototype for drilling training: A prospective study. *Clinical Otolaryngology*. 2017;42(6):1200-1205. doi:10.1111/coa.12846

126. Cook DA, Beckman TJ. Reflections on experimental research in medical education. *Adv in Health Sci Educ*. 2010;15:455-464. doi:10.1007/s10459-008-9117-3

127. Flannelly KJ, Flannelly LT, Jankowski KRB. Threats to the Internal Validity of Experimental and Quasi-Experimental Research in Healthcare. *J Health Care Chaplain*. 2018;24(3):107-130. doi:10.1080/08854726.2017.1421019

128. Wijewickrema S, Talks BJ, Lamtara J, Gerard JM, O'Leary S. Automated assessment of cortical mastoidectomy performance in virtual reality. *Clinical Otolaryngology*. 2021;46(5):961-968. doi:10.1111/coa.13760

129. Brydges R, Nair P, Ma I, Shanks D, Hatala R. Directed self-regulated learning versus instructor- regulated learning in simulation training. *Med Educ*. 2012;46:648-656. doi:10.1111/j.1365-2923.2012.04268.x

130. Frendø M, Konge L, Caye-Thomasen P, Sørensen MS, Andersen SAW. Decentralized Virtual Reality Training of Mastoidectomy Improves Cadaver Dissection Performance : A Prospective , Controlled Cohort Study. *Otol Neurotol*. 2020;41(4):476-481. doi:10.1097/MAO.0000000000002541

131. Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ*. 2012;46(7):636-647. doi:10.1111/j.1365-2923.2012.04243.x

132. Frithioff A, Frendø M, Mikkelsen PT, Sørensen MS, Andersen SAW. Ultra-high-fidelity virtual reality mastoidectomy simulation training : a randomized , controlled trial. *European Archives of Oto-Rhino-Laryngology*. 2020;277(5):1335-1341. doi:10.1007/s00405-020-05858-3

133. Grober ED, Hamstra SJ, Wanzel KR, et al. The Educational Impact of Bench Model Fidelity on the Acquisition of Technical Skill The Use of Clinically Relevant Outcome Measures. *Ann Surg*. 2004;240(2):374-381. doi:10.1097/01.sla.0000133346.07434.30

134. Roosli C, Sim JH, Mockel H, Mokosch M, Probst R. An Artificial Temporal Bone as a Training Tool for Cochlear Implantation. *Otology and neurotology*. 2013;34(PG-1048-1051):1048-1051. NS -

135. Chien WW, da Cruz MJ, Francis HW. Validation of a 3D-printed human temporal bone model for otology surgical skill training. *World J Otorhinolaryngol Head Neck Surg*. 2021;7(2):88-93. doi:10.1016/j.wjorl.2020.12.004

136. Wanibuchi M, Noshiro S, Sugino T, et al. Training for Skull Base Surgery with a Colored Temporal Bone Model Created by Three-Dimensional Printing Technology. *World Neurosurg*. 2016;91:66-72. doi:10.1016/j.wneu.2016.03.084

137. Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. *BMJ*. 1999;318:887-888.

138. Szasz P, Louridas M, Harris KA, Aggarwal R, Grantcharov TP. Assessing technical competence in surgical trainees: A systematic review. *Ann Surg*. 2015;261(6):1046-1055. doi:10.1097/SLA.0000000000000866

139. Andersen SAW, Mikkelsen PT, Sørensen MS. Expert Sampling of VR Simulator Metrics for Automated Assessment of Mastoidectomy Performance. *Laryngoscope*. 2019;129:2170-2177. doi:10.1002/lary.27798

140. Dinah W, Wiet G, Welling B, Kerwin T, Stredney D. Creating a Cross-Institutional Grading Scale for Temporal Bone Dissection. *Laryngoscope*. 2010;120(7):1422-1427. doi:10.1053/j.gastro.2011.12.001.Dysregulation

141. Kerwin T, Wiet G, Hittle B, et al. Standard Setting of Competency in Mastoidectomy for the Cross-Institutional Mastoidectomy Assessment Tool. *Annals of Otology, Rhinology and Laryngology*. 2020;129(4):340-346. doi:10.1177/0003489419889376

142. Frendø M, Frithioff A, Konge L, et al. Assessing competence in cochlear implant surgery using the newly developed Cochlear Implant Surgery Assessment Tool. *European Archives of Oto-Rhino-Laryngology*. Published online 2021. doi:10.1007/s00405-021-06632-9

143. Andersen SAW, Nayahangang L, Park YS, Konge L. Use of Generalizability Theory for Exploring Reliability of and Sources of Variance in Assessment of Technical Skills: A Systematic Review and Meta-Analysis. *Academic Medicine*. 2021;(Online ahead of print). doi:10.1097/ACM.0000000000004150

144. Andersen SAW, Park YS, Sørensen MS, Konge L. Reliable Assessment of Surgical Technical Skills Is Dependent on Context : An Exploration of Different Variables Using Generalizability Theory. *Academic Medicine*. 2020;95(12):1929-1936. doi:10.1097/ACM.0000000000003550

145. Cook DA. Much ado about differences: why expert-novice comparisons add little to the validity argument. *Advances in Health Sciences Education*. 2015;20(3):829-834. doi:10.1007/s10459-014-9551-3

146. Trognon AM, Batt M. Group Dynamics and Learning. In: *Encyclopedia of the Sciences of Learning*. Springer US; 2012:1388-1391. doi:10.1007/978-1-4419-1428-6_1863

147. Andersen SAW, Konge L, Cayé-Thomasen P, Sørensen MS. Learning curves of virtual mastoidectomy in distributed and massed practice. *JAMA Otolaryngol Head Neck Surg*. 2015;141(10):913-918. doi:10.1001/jamaoto.2015.1563

148. 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. *Journal of Laryngology and Otology*. 2012;126(7):663-669. doi:10.1017/S0022215112000734

149. Pusic M v., Boutis K, Hatala R, Cook DA. Learning Curves in Health Professions Education. *Academic Medicine*. 2015;90(8):1034-1042. doi:10.1097/ACM.0000000000000681

150. Howard NM, Cook DA, Hatala R, Pusic M v. Learning Curves in Health Professions Education Simulation Research: A Systematic Review. *Simul Healthc*. 2021;16(2):128-135. doi:10.1097/SIH.0000000000000477

151. Gustafsson A, Pedersen P, Rømer TB, Viberg B, Palm H, Konge L. Hip-fracture osteosynthesis training: exploring learning curves and setting proficiency standards. *Acta Orthop*. 2019;90(4):348-353. doi:10.1080/17453674.2019.1607111

152. Löppönen H, Holma T, Sorri M, et al. Computed tomography data based rapid prototyping model of the temporal bone before cochlear implant surgery. *Acta Otolaryngol*. 1997;529:47-49. doi:10.3109/00016489709124077

153. Suzuki M, Ogawa Y, Hasegawa T, et al. Simultaneous, laser-sintered, three-dimensional modelling of bony structures and soft tissue for surgical navigation of extended cholesteatoma. *Journal of Laryngology and Otology*. 2010;124(5):564-568. doi:10.1017/S0022215109991393

154. Suzuki M, Hagiwara A, Kawaguchi S, Ono H. Application of a rapid-prototyped temporal bone model for surgical planning. *Acta Otolaryngol*. 2005;125(1):29-32. doi:10.1080/00016480410016874

155. Ritacco LE, di Lella F, Mancino A, Gonzalez Bernaldo De Quiros F, Boccio C, Milano FE. 3D Printed Models and Navigation for Skull Base Surgery: Case Report and Virtual Validation. Published online 2015. doi:10.1007/s00405

156. della Volpe A, de Lucia A, Ippolito V, et al. Use of a 3D reconstruction model in a patient with severe atresia auris for optimal placement of Bonebridge transcutaneous bone conduction implant. *European Archives of Oto-Rhino-Laryngology*. 2021;278(9):3559-3564. doi:10.1007/s00405-020-06528-0

157. Andrews JC, Mankovich NJ, Anzai Y, Lufkin RB. Stereolithographic model construction from CT for assessment and surgical planning in congenital aural atresia. *Am J Otol*. 1994;15(3):335-339.

158. Omari A, Frendø M, Sørensen MS, Andersen SAW, Frithioff A. The cutting edge of customized surgery: 3D-printed models for patient-specific interventions in otology and auricular management—a systematic review. *European Archives of Oto-Rhino-Laryngology*. 2022;279(7):3269-3288. doi:10.1007/s00405-022-07291-0

Appendix:

The 26-item modified Welling Scale

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

Only assessed in cadaver dissection performances