Received: 26 March 2023 | Revised: 17 June 2023 | Accepted: 30 June 2023

DOI: 10.1002/lio2.1123

ORIGINAL RESEARCH

Laryngoscope Investigative Otolaryngology

Gathering validity evidence for a 3D-printed simulator for training of myringotomy and ventilation tube insertion

Michael Lüscher MD, PhD[1] | Lars Konge MD, PhD[2] | Peter Tingsgaard MD, PhD[3] | Thomas Qvist Barrett MD[4] | Steven Arild Wuyts Andersen MD, PhD[5]

[1]Sundhedshuset, Aarhus, Denmark

[2]Copenhagen Academy for Medical Education and Simulation (CAMES), Center for Human Resources & Education, Copenhagen, Denmark

[3]ENT Private Clinic, Slagelse, Denmark

[4]ENT Private Clinic, Randers, Denmark

[5]Department of Otorhinolaryngology, Copenhagen Hearing and Balance Center, Rigshospitalet, Copenhagen, Denmark

Correspondence
Michael Lüscher, Sundhedshuset, Park Allé 15, Aarhus, Denmark.
Email: michaellyscher@yahoo.dk

Funding information
Fund for Professional Development in Specialist Practice, Grant/Award Number: A3418

Abstract

Objectives: This study aimed to gather validity evidence according to Messick's framework for a novel 3D-printed simulator for myringotomy with ventilation tube insertion for use in technical skills training of otorhinolaryngology (ORL) residents.

Methods: The study included 15 junior ORL residents (trainees) and 13 experienced teaching otolaryngologists (experts). Experts and trainees first received an identically structured introduction to the procedure, simulator, and simulation setup. Five procedures performed by each participant were video-recorded and ordered randomly for blinded rating by two independent raters. The rating tools used were a global rating scale (GBRS) and a task-specific checklist. Validity evidence was collected according to Messick's framework. Differences in time consumption and performance scores were analyzed. Finally, a pass/fail standard was established using the contrasting groups' method.

Results: Trainees used significantly more time per procedure (109 s, 95% CI: 99–120) than experts (82 s, 95% CI: 71–93; $p < .001$). Adjusted for repetition and rater leniency, experts achieved an average GBRS score of 18.8 (95% CI: 18.3–19.2) out of 20 points, whereas trainees achieved an average of 17.1 points (95% CI: 16.6–17.5; $p < .001$). In contrast to the task-specific checklist, the GBRS score discriminated between repetition number and participant experience. The pass/fail standard for the GBRS was established at 18.4 points.

Conclusion: We established educational validity evidence for a novel 3D-printed model for simulation-based training of ventilation tube insertion and established a reliable pass/fail standard.

Level of Evidence: 1b.

KEYWORDS
education, myringotomy, simulation, training, validity evidence, ventilation tube insertion

1 | INTRODUCTION

Simulation-based training (SBT) has gained a strong foothold in modern health professional training, including in surgery.[1–3] Simulation has many benefits as a learning and assessment environment.[4] Importantly, it provides trainees with a safe environment to practice and refine their skills and obtain competencies without risk to patients.[5,6] Accordingly, SBT has been adopted in most surgical specialties including otorhinolaryngology (ORL) to improve novice surgical skills acquisition.

Myringotomy with ventilation tube insertion is one of the most frequently performed procedures in pediatric ORL.[7–9] For several reasons, myringotomy lends itself excellently to SBT: First, successful myringotomy requires training of otology residents with no or minimal experience in procedures performed under microscope. Second, the procedure may cause iatrogenic harm, such as damage to middle ear structures or the ear drum, or laceration of ear canal skin.[10] Third, the learning curve is steep, and excess time consumption is not uncommon among less experienced practitioners. Simulator training before myringotomy and ventilation tube insertion on patients therefore has potential benefits for patients and trainees alike. In a recent targeted needs assessment study in Denmark, instructors and ORL trainees identified ventilation tube insertion as the most desired skill to train using simulation.[11] Numerous physical simulation models have been described in the literature.[12] Generally, they consist of a cylinder mimicking the ear canal with a synthetic membrane attached representing the ear drum. Other models are more complex and involve, for example, a virtual reality environment.[13,14] However, only a minority of these simulators are underpinned by educational evidence as few have explored validity using a contemporary validity framework.[15] Messick's validity framework considers five evidence sources: content, response process, internal structure, relationships with other variables, and consequences.[15] Structured gathering of validity evidence is needed to support the educational value of implementing SBT and ensure alignment between construct, measurement, and outcome. Furthermore, a need exists to establish a credible pass/fail score underpinning mastery learning in which all trainees continue to practice the procedure until they reach a predefined proficiency level.[16]

This study aimed to gather validity evidence according to Messick's framework for a novel 3D-printed simulator for myringotomy with ventilation tube insertion in the setting of training and assessing junior ORL residents.

2 | MATERIALS AND METHODS

2.1 | Simulator

Our idea was to provide a portable, standardized, and flexible model for training of myringotomy and ventilation tube insertion that may be distributed to training departments/clinics for convenient decentralized training. We developed a physical simulator based on digital illustrations of the auricle, external ear canal, tympanic cavity, and head using computer-assisted design (CAD) tools. This work was done in collaboration with the Institute of Technology, Aarhus, Denmark. The model was made to resemble the normal human anatomy for the right ear of an approximately 5-year-old child. We made a 3D-printed model and then cast the ear canal and auricle in silicone (Figure 1, left). The tympanic membrane is represented by glued rice paper layers, mimicking a thin translucent tympanic membrane. The cartridge with "tympanic membranes" has five membranes and enables fast change for repeated practice by sliding the cartridge into the simulator (Figure 1, right). The final simulation model requires only standard instruments and an otomicroscope, all of which are readily available at training institutions.

2.2 | Participants

We recruited a group of highly experienced specialists ("experts") and a group of trainees with limited experience ("trainees"). Experts were recruited among the Danish network of teaching ear–nose–throat (ENT) specialists in private practice (a total of 19 trainers were eligible nationwide). They all had years of experience performing tubulation and myringotomy in pediatric patients (>2000 procedures) and served as ENT resident trainers. Trainee participants were recruited among otolaryngology residents in their ENT private practice training rotation (i.e., >3 years of ENT experience). The trainees had limited experience with myringotomy with ventilation tube insertion before their rotation in specialist practice. They were further classified as either novices (<10 procedures) or beginners (11–100 procedures).

2.3 | Intervention—Simulation of myringotomy with ventilation tube insertion

Both experts and trainees first received the same structured introduction to the procedure and simulation setup with videos explaining ear anatomy, use of surgical instruments, the otomicroscope, the video-recording set-up used in the study, and steps for myringotomy and ventilation tube placement as included in the assessment checklist (see later). The content of the instructional videos was based on a consensus obtained among the previously mentioned network of teaching ENT specialists ($N = 15$ contributed at the consensus meeting). The consensus was highly aligned with the procedure as described in Scott-Brown's textbook.[17] After the structured introduction, each participant performed five myringotomies with ventilation tube insertion on the simulator. These myringotomies were all video-recorded and assessed as described below. The study flow chart is presented in Figure 2.

2.4 | Data collection—Video recording

The set-up for video recording required a smartphone camera and a tripod. Participants needed to follow the recording guideline provided

FIGURE 1 Left—Simulator for myringotomy with ventilation tube insertion, consisting of a 3D-printed model with a silicone ear mold. Right—Cartridge for fast change of tympanic membrane.

in the instructional video. This included details on how to position the recording equipment for optimal view and ensuring blinding of the participant for unbiased rating. The smartphone camera recording included the visual field from the otomicroscope projected onto a screen, the participant's lower arm and hands during instrument handling, and the simulator (Figure 1). Participants recorded all five performances as one consecutive video. Subsequently, these videos were edited into five stand-alone videos. The video clips were assigned a random number before being assessed. The procedural time was determined by videoclip length.

2.5 | Data collection—Performance assessment

Two ENT specialists with extensive surgical and teaching expertise were recruited to assess performance. They received rater training from the first and last author. Training included review and discussion of the two assessment tools after which three example videos were rated and discussed for consensus. After rater training, the two raters were given access to the video-recorded performances and individually assessed performances using two assessment tools. An example of a video recording is provided as Supporting Information digital content (Video S1).

The tools were adopted from Malekzadeh et al.[18] and consisted of a (1) global rating scale (GBRS) and a (2) task-specific checklist. The original GBRS considered the following items: Time and motions, Instrument handling, Knowledge of instruments, Flow of operation, and Communication skills. Since the procedure simulated in our study involved no assistant, we omitted assessment of Communication skills. Each of the four remaining items was rated on an anchored five-point Likert scale for a total maximum of 20 points. The task-specific checklist comprised five tasks rated as "done" (1 point) or "not done" (0 points): Visualizes tympanic membrane, performs myringotomy in appropriate quadrant, suctions middle ear and not directly on tympanic membrane, inserts tube through myringotomy incision, and suctions ear tube. Subsequently, the raters made a global pass/fail decision.

2.6 | Analysis—Validity evidence according to Messick's framework

We assessed validity evidence for procedural competency in surgery for ventilation tube insertion using the simulator in accordance with Messick's framework and adopting the following five sources of validity[15]:

1. **Content.** We obtained consensus among a group of experienced ORL surgeons who assessed the elements and content of the simulator, the instructions, and the simulation-based test.
2. **Response process.** Identical and uniform instruction of each participant was achieved using identical instructional videos. Bias in the rating process was eliminated by blinding expert raters to the participants. The raters had no information about the ratings done by the other rater.
3. **Internal structure.** The internal consistency of the simulator items for each repetition of the procedure was assessed using Cronbach's alpha. A coefficient >0.8 is considered acceptable for moderate stakes tests.[19]
4. **Relationships with other variables.** We compared the groups' scores to establish whether simulator test scores may discriminate between different experience levels.
5. **Consequences.** A pass/fail standard was established based on the GBRS score using the contrasting groups' method. This method considers intersection between the distribution of the novice and experts to have as few false positives (passed novice surgeons) and as few false negatives (failed experienced surgeons) as possible, respectively.[20]

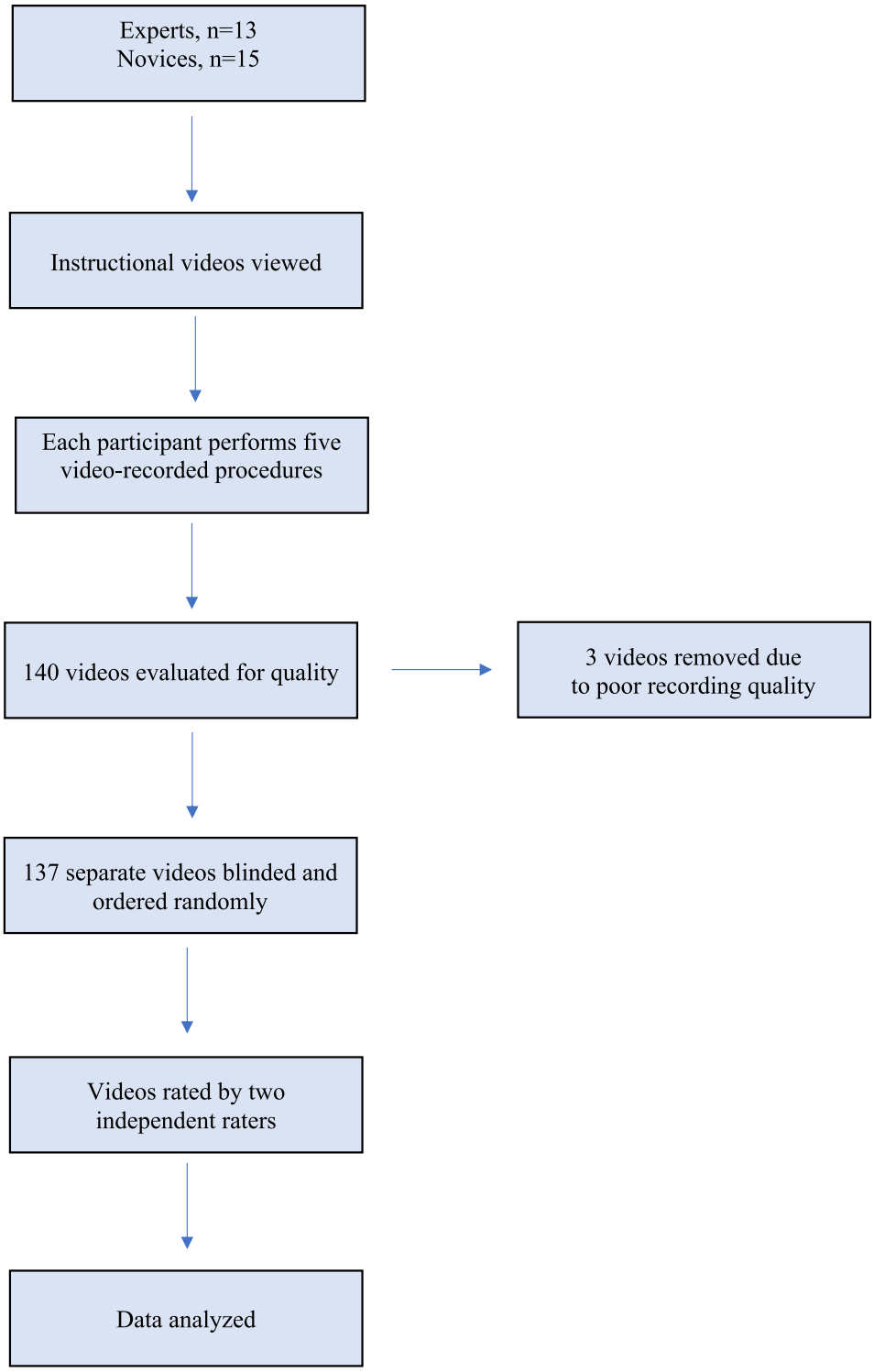

FIGURE 2 Study flow chart.

2.7 | Sample size and statistics

Our sample size calculations were based on data presented by Malekzadeh et al.,[18] suggesting that 12 participants would be needed in each group to detect a significant inter-group difference, assuming the same effect size as found for their educational intervention.

We analyzed data in SPSS statistics version 28 (SPSS Inc, IBM Corp., Armonk, NY). For data analysis, we mainly used linear mixed models as outlined by Leppink to account for repeated measurements (multiple performances by each participant, multiple raters).[21] The final models included level (novice or expert) or experience (0–10, 11–100, or > 2000 real-life procedures), procedure number (1–5), and rater as fixed effect. No interactions were found. Estimated marginal

means are reported. An Excel spreadsheet was used to contrast group calculations.[20]

2.8 | Ethics

Ethical board approval is not required for educational studies under the Danish research provisions. This study complied with the Helsinki Declaration. All participants were informed and provided their written consent for participation.

3 | RESULTS

3.1 | Participants and background

The study participants were 15 junior ORL residents and 13 experienced teaching otolaryngologists in specialist practice (19 experts were eligible, 13 accepted participation). The participants had no previous exposure to the simulator. The trainees were 29–40 years of age (median: 33 years), nine were male; six, female. The experts were 40–65 years of age (median: 53 years), 12 were male; one, female. All 28 participants completed and recorded five simulator procedures, producing a total of 140 video clips. Three recordings were excluded due to poor quality (overexposed), leaving 137 videos for evaluation.

3.2 | Internal consistency

Cronbach's alpha of the GRBS in our context was found to be 0.90, which is considered a very high internal consistency. In contrast, Cronbach's alpha for the task-based checklist was very low—0.34.

3.3 | Procedure time

The trainees used mean 109 s per procedure (95% CI: 99–120); the experts, 82 s (95% CI: 71–93). This 27-s difference was statistically significant ($p < .001$). No statistically significant effect was recorded of procedure number, meaning that trainees and experts did not reduce their procedure time with repetition within the first five procedures. We further found a significant effect of experience within the trainee group ($p < 0.001$): novices (i.e., participants who had performed <10 real-life procedures) used mean 117 s per procedure (95% CI: 105–129); beginners (11–100 real-life procedures), mean 87 s per procedure (95% CI: 70–105).

3.4 | Global rating scale

Adjusted for rater leniency, the experts achieved an average score of 18.8 points (95% CI: 18.3–19.2) out of 20 points on the GBRS,

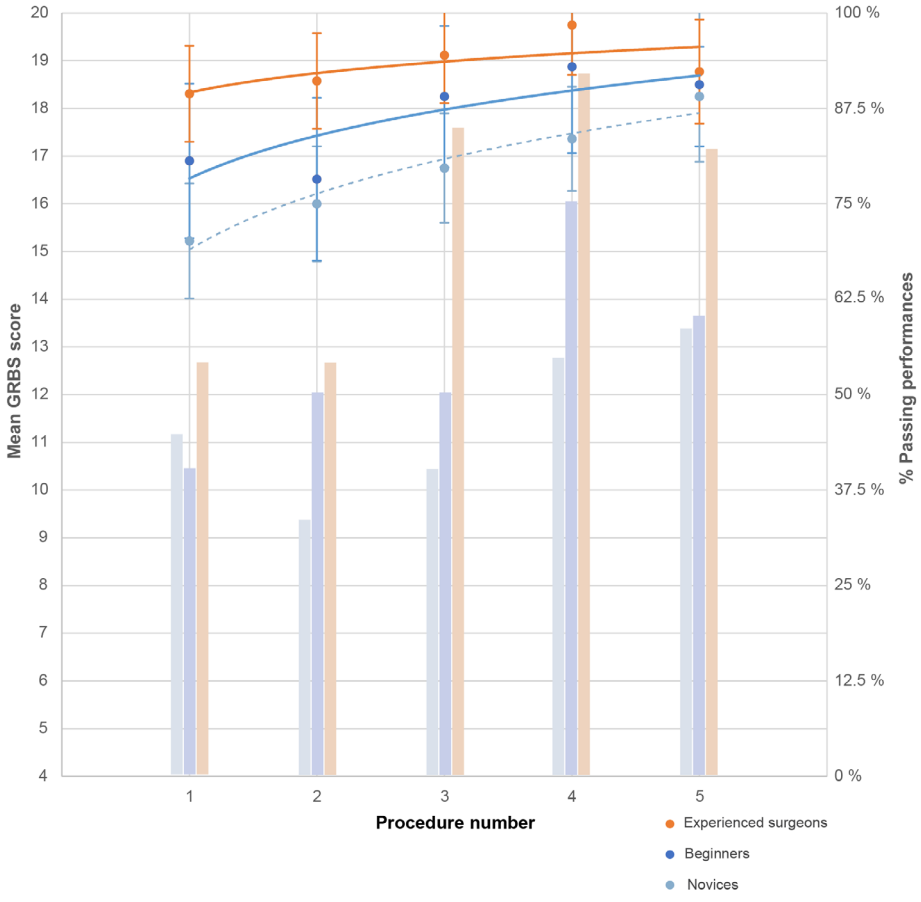

FIGURE 3 Mean GBRS and passing performance. Top: means plot of global rating scale scores (estimated marginal means after adjustment for experience, procedure number, and rater leniency) with 95% confidence intervals based on linear mixed models of total score. Bottom: Bar charts of % of passing performances for each group using the standard pass/fail level set at 18.4 point using the contrasting groups' method. GBRS, global rating scale.

whereas the trainees achieved an average score of 17.1 points (95% CI: 16.6–17.5). This difference was statistically significant ($p < .001$). Furthermore, we found a significant effect of procedure number, meaning that in both participant groups, performance improved with repetition at least for the first three procedures. We further found a significant effect of real-life experience in the trainee group ($p = 0.02$) with novices (<10 procedures) achieving an average score of 16.8 points (95% CI: 16.2–17.3) and beginners (11–100 procedures) achieving an average score of 17.8 points (95% CI: 17.1–18.6). A means plot of performance for each procedure number according to experience level is provided in Figure 3 (top). Altogether, the GBRS score used in the context of the simulation finely discriminated between performance number, participant level, and participant experience.

3.5 | Checklist score

For the task-based checklist score, no significant difference was found between experts and trainees ($p = 0.54$). Thus, the checklist score used in the context of our simulator failed to discriminate between the performances of experts and trainees.

3.6 | Overall pass/fail

Adjusted for rater leniency, the experts had a significantly higher average passing performance in 96.3% of the procedures, whereas the trainees had a passing performance in 86.3% of procedures ($p < .003$).

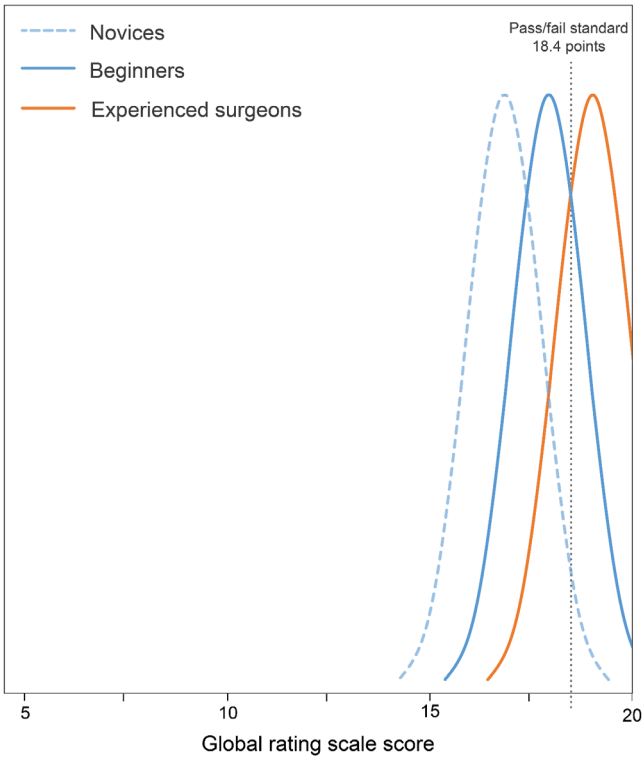

FIGURE 4 Contrasting groups for standard setting.

3.7 | Standard setting

Using the contrasting groups' method, we calculated a pass/fail standard level of 18.4 points as the intersection between the score distribution of the beginners (10–100 procedures) and the experts (>2000 procedures). This level would have a theoretical false positive rate of 25.3% (i.e., beginners passing), and a theoretical false negative rate of 25.9% (i.e., experts failing). The percentage of passing performances for the standard level set according to experience is found in Figure 4 (bottom). The average absolute agreement between the pass/fail assigned by raters and by using the standard setting method was 67.5%.

4 | DISCUSSION

In this prospective study, we gathered validity evidence for a novel 3D-printed tubulation simulator and investigated trainees' and experienced surgeons' performance. The main finding was that performance assessment using the simulator statistically significantly discriminated between different experience levels (novice, beginner, expert). Furthermore, we established evidence for other domains of validity according to Messick's framework, including setting a pass/fail level.

Using the GBRS score proposed by Malakzadeh et al.[18] adapted for use in our simulation context, we discriminated between performances based on surgical experience. We found that both trainees and experienced participants demonstrated a learning curve described by increasing performance quality with repetition. That a learning curve in SBT is found is unsurprising[22] and was also observed in most other relevant studies[23]: trainees need to learn both the procedure and the simulation equipment and therefore demonstrate a prolonged learning curve, whereas experienced surgeons' learning curve may be explained by their need to learn how to use the simulator and adjust to any differences between the simulation environment and real-life surgery.

We also found that the fewer procedures trainees had performed, the more they benefited from SBT on our simulator as we found an effect of real-life experience in the trainee group with beginners (11–100 procedures) achieving a significantly higher score than novices (<10 procedures). This finding is in line with a common negatively accelerated learning curve and the results of Kovatch et al.[24] who described that sub-interns gained the most from SBT in terms of self-reported knowledge and confidence with myringotomy and ventilation tube insertion.

Other simulation models for tubulation appear not to have achieved wide-spread implementation into ORL training curricula. One reason may be that current simulators do not adequately simulate anatomical structures and human tissues.[10,25] In a large study by Wiet et al.,[26] several barriers were encountered to using simulation in ventilation tube insertion in a nation-wide scale: First, the low-cost simulation model may not have been a sufficient training model despite previous validation. Second, differences in the degree or quality of teaching may potentially create variance within raters and institutions.

The OSATS model for assessment was used successfully in various surgical training settings, specifically those involving bench models or individual training stations.[27–29] Prior research has found that OSATS may be used to reliably assess and measure technical skills acquisition.[30] Similarly to OSATS-type tools for technical skills assessment, we assessed participants' performance using the global rating scale by Malekzadeh et al.[18] We found that the OSATS-type GBRS was very useful for discrimination between relevant experience levels. In contrast, the task-based checklist score used in our simulation context failed to discriminate between experts' and trainees' performance and further had a very low internal consistency. This finding is in agreement with Malekzadeh et al.,[18] who also found the global rating scale to have discriminative validity, whereas a checklist score did not. Generally, GBRS are favored over task-based checklists for their granularity.[30] However, it was not the primary scope of our investigation to compare the two assessment tools, and the poor performance of the task-based check list, could be due to other factors in our study, for example, sample size and level of difficulty of the simulator.

The use of a standard setting refers to the process employed to establish the cut-off between pass and fail of a performance or a competent and a non-competent learner. The contrasting groups' standard setting method is commonly used in surgical technical skills training.[31] We established a pass/fail standard for GBRS of 18.4 points out of 20 possible points using this method and, as expected, found that more performance attained passing level with repetition and experience level. Furthermore, we found a moderate absolute agreement between this pass/fail level and the pass/fail assigned by the raters. In general, we found a good agreement between raters.

Altogether, this pass/fail score suggests a level of performance that may be implemented in future proficiency-based simulation training.

A strength of our study was that participants were recruited among target learners (junior ORL residents) and teachers. Another strength is that we used a contemporary validity framework to structure the gathering of educational evidence. A limitation of our study is that our approach did not ensure perfect participant blinding. We used a smart phone camera including the visual field from the otomicroscope projected onto a screen, the operator's lower arm and hands, and the simulator. It cannot be ruled out that raters could identify novices as younger individuals; however, the youngest expert was approximately of same age as the oldest novice, which reduces bias.

Our study has several implications for the implementation of SBT of myringotomy with ventilation tube. The 3D-printed model enables decentralized training, and assessment may be reliably done by video recording performance. The established standard may possibly be used for proficiency-based training, meaning that trainees need to consistently demonstrate performance at the defined level in the simulator before continuing supervised training on patients.[32] However, a need exists to establish the effect of SBT using the simulator on live surgery performance (so-called transfer).[33] This is important to determine if appropriate tubulation simulator training may lower time consumption, secure optimal ventilation tube placement, and reduce the complication rate in real life.

The simulator was designed to imitate real-life surgery. It incorporates important components of the procedure, including accurate head positioning, appropriate instrument selection, handling the ventilation tubes, microscope adjustments, and realistic haptic feedback related to the eardrum. However, certain aspects inherent in real-life procedures were not entirely addressed. Potential refinements to the simulator involve the integration of different ear canal anatomies, middle ear fluid, and the simulation of complications, such as hemorrhage from the eardrum and middle ear. This is under consideration for future improvements of the simulator.

5 | CONCLUSIONS

We established educational validity evidence for a novel 3D-printed model for SBT of ventilation tube insertion and established a credible pass/fail standard. This performance level may be employed in proficiency-based training. The simulator- and simulation-based assessment may be embedded into the ORL surgical training curriculum as initial training before the procedure is performed under supervision on patients.

FUNDING INFORMATION

This project was supported by a grant from the Fund for Professional Development in Specialist Practice.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Michael Lüscher https://orcid.org/0000-0003-2054-1130
Steven Arild Wuyts Andersen https://orcid.org/0000-0002-3491-9790

REFERENCES

1. Carraccio C, Englander R, Van Melle E, et al. Advancing competency-based medical education: a charter for clinician-educators. *Acad Med J Assoc Am Med Coll*. 2016;91(5):645-649.
2. Shah N, Desai C, Jorwekar G, Badyal D, Singh T. Competency-based medical education: an overview and application in pharmacology. *Indian J Pharmacol*. 2016;48(Suppl 1):S5-S9.
3. Ten Cate O. Competency-based postgraduate medical education: past, present and future. *GMS J Med Educ*. 2017;34(5):Doc69.
4. Bjerrum F, Sorensen JL, Thinggaard E, Strandbygaard J, Konge L. Implementation of a cross-specialty training program in basic laparoscopy. *JSLS*. 2015;19(4):e2015.00059.
5. Bewley WL, O'Neil HF. Evaluation of medical simulations. *Mil Med*. 2013;178(10 Suppl):64-75.
6. Nayahangan LJ, Albrecht-Beste E, Konge L, et al. Consensus on technical procedures in radiology to include in simulation-based training for residents: a European-wide needs assessment. *Eur Radiol*. 2021;31(1):171-180.
7. Siggaard LD, Barrett TQ, Lüscher M, et al. Parent satisfaction and symptom relief in children with otitis media undergoing tympanostomy tube insertion. *Dan Med J*. 2019;66(9):A5567.

8. Rosenfeld RM, Schwartz SR, Pynnonen MA, et al. Clinical practice guideline: tympanostomy tubes in children. *Otolaryngol Head Neck Surg*. 2013;149(1 Suppl):S1-S35.
9. National clinical guideline tympanostomy tube treatment. Danish Health Authority. 2015 Accessed October 4, 2021. https://www.sst.dk/da/udgivelser/2015/~/media/B23EC346DEC1408C8419725D96A76C0A.ashx
10. Montague ML, Lee MS, Hussain SS. Human error identification: an analysis of myringotomy and ventilation tube insertion. *Arch Otolaryngol Head Neck Surg*. 2004;130(10):1153-1157.
11. Lüscher M, Nayahangan LJ, Thrane JF, Danstrup CS, Konge L, Andersen SAW. Simulation-based surgical training needs in otorhinolaryngology. *Dan Med J*. 2022;69(4):A08210630.
12. Hovgaard LH, Al-Shahrestani F, Andersen SAW. Current evidence for simulation-based training and assessment of myringotomy and ventilation tube insertion: a systematic review. *Otol Neurotol*. 2021;42(9):e1188-e1196.
13. Vaitaitis VJ, Dunham ME, Kwon Y-C, et al. A surgical simulator for tympanostomy tube insertion incorporating capacitive sensing technology to track instrument placement. *Otolaryngol Head Neck Surg*. 2020;162:343-345.
14. Ho AK, Alsaffar H, Doyle PC, Ladak HM, Agrawal SK. Virtual reality myringotomy simulation with real-time deformation: development and validity testing. *Laryngoscope*. 2012;122:1844-1851.
15. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract*. 2014;19(2):233-250.
16. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med*. 2013;88:1178-1186.
17. Watkingson J, Clarke R, eds. *Scott-Brown's Otorhinolaryngology and Head and Neck Surgery*. 8th ed. CRC Press; 2018.
18. Malekzadeh S, Hanna G, Wilson B, Pehlivanova M, Milmoe G. A model for training and evaluation of myringotomy and tube placement skills. *Laryngoscope*. 2011;121:1410-1415.
19. Yudkowsky R, Park YO, Downing SM, eds. *Assessment in Health Professions Education*. 2nd ed. Routledge; 2019.
20. Jørgensen M, Konge L, Subhi Y. Contrasting groups' standard setting for consequences analysis in validity studies: reporting considerations. *Adv Simul (Lond)*. 2018;3:5.
21. Leppink J. Data analysis in medical education research: a multilevel perspective. *Perspect Med Educ*. 2015;4(1):14-24.
22. Pusic MV, Boutis K, Hatala R, Cook DA. Learning curves in health professions education. *Acad Med*. 2015;90(8):1034-1042.
23. Howard NM, Cook DA, Hatala R, Pusic MV. Learning curves in health professions education simulation research: a systematic review. *Simul Healthc*. 2021;16(2):128-135.
24. Kovatch KJ, Wertz AP, Carle TR, et al. Optimal timing of entry-level otolaryngology simulation. *OTO Open*. 2019;3(2):2473974X19845851.
25. Mahalingam S, Awad Z, Tolley NS, Khemani S. Ventilation tube insertion simulation: a literature review and validity assessment of five training models. *Clin Otolaryngol*. 2016;41(4):321-326.
26. Wiet GJ, Deutsch ES, Malekzadeh S, et al. SimTube: a National Simulation Training and research project. *Otolaryngol Head Neck Surg*. 2020;163(3):522-530.
27. Moorthy K, Munz Y, Sarker S, Darzi A. Objective assessment of technical skills in surgery. *BMJ*. 2003;327:1032-1037.
28. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.
29. Jakubowski L, Leader BA, Ishman SL, et al. Validation of an objective assessment tool for tonsillectomy in otolaryngology resident training. *Laryngoscope*. 2021;131(2):E359-E366.
30. Reznick RK, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg*. 1996;172:226-230.
31. 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.
32. McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. *Med Educ*. 2014;48:375-385.
33. Stefanidis D, Arora S, Parrack DM, et al. Association for Surgical Education Simulation Committee. Research priorities in surgical simulation for the 21st century. *Am J Surg*. 2012;203(1):49-53.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lüscher M, Konge L, Tingsgaard P, Barrett TQ, Andersen SAW. Gathering validity evidence for a 3D-printed simulator for training of myringotomy and ventilation tube insertion. *Laryngoscope Investigative Otolaryngology*. 2023;1-8. doi:10.1002/lio2.1123