Full citation: 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 Oct 1;42(9):e1188-e1196.
DOI: 10.1097/MAO.0000000000003268

Title: Current evidence for simulation-based training and assessment of myringotomy and ventilation tube insertion: A systematic review

Authors:
Lisette Hvid Hovgaard, MD[1]; Fahd Al-Shahrestani, MD[2]; Steven Arild Wuyts Andersen[1,2,3], MD, PhD.

Affiliations:
1. Copenhagen Academy for Medical Education and Simulation (CAMES), Center for HR & Education, RegionH, Denmark.
2. Department of Otorhinolaryngology-Head & Neck Surgery, Rigshospitalet, Copenhagen, Denmark.
3. Department of Otolaryngology, Nationwide Children's Hospital, and the Ohio State University, Columbus, Ohio, USA.

Correspondence: Lisette Hvid Hovgaard, Copenhagen Academy for Medical Education and Simulation (CAMES), Blegdamsvej 9, DK-2100 Copenhagen, Denmark. E-mail: lisette.hvid.hovgaard@gmail.com

Conflicts of Interest and Source of Funding: Dr. Andersen has received research funding by the Independent Research Fund Denmark (8026-00003B). The authors have no other funding or conflicts of interest to disclose.

ABSTRACT

Objective: Myringotomy and ventilation tube insertion (MT) is a key procedure in otorhinolaryngology and can be trained using simulation models. We aimed to systematically review the literature on models for simulation-based training and assessment of MT and supporting educational evidence.

Databases Reviewed: PubMed, Embase, Cochrane Library, Web of Science, Directory of Open Access Journals (DOAJ).

Methods: Inclusion criteria were MT training and/or skills assessment using all types of training modalities and learners. Studies were divided into 1) descriptive and 2) educational interventional/observational in the analysis. For descriptive studies, we provide an overview of available models including materials and cost. Educational studies were appraised using Kirkpatrick's level of educational outcomes, Messick's framework of validity, and a structured quality assessment tool.

Results: Forty-six studies were included consisting of 21 descriptive studies and 25 educational studies. 31 unique physical and three virtual reality simulation models were identified. The studies report moderate to high realism of the different simulators and trainees and educators perceive them beneficial in training MT skills. Overall, simulation-based training is found to reduce procedure time and errors, and increase performance as measured using different assessment tools. None of the studies used a contemporary validity framework and the current educational evidence is limited.

Conclusion: Numerous simulation models and assessment tools have been described in the literature but educational evidence and systematic implementation into training curricula is scarce. There is especially a need to establish the effect of simulation-based training of MT in transfer to the operating room and on patient outcomes.

INTRODUCTION

In recent years, the traditional paradigm of "see one, do one, teach one" in surgical education has shifted towards competency-based training to ensure patient safety and high-quality patient care.(1) Consequently, contemporary surgical training now requires strong evidence for the chosen training and assessment methods in the surgical curriculum.

Myringotomy is the incision of the tympanic membrane, which is often followed by insertion of a ventilation tube. Myringotomy with ventilation tube insertion (MT) is mainly performed to alleviate secretory otitis media and recurrent acute otitis media in children(2,3) and is the most commonly performed procedure in otorhinolaryngology (ORL) in many countries including the United States, England and Denmark.(4-6) Frequently, MT is one of the first procedures that need to be mastered during ORL residency. However, the procedure can be challenging due to the compound skills required such as hand-eye coordination, a narrow surgical field, and the use of an operating microscope.

Physical and virtual reality (VR) simulation models can be used for surgical skills training so that patients are not used during initial skills acquisition.(5) Generally, simulation-based training (SBT) has demonstrated large and positive effects on knowledge, skills and behaviors in addition to effects on patient-related outcomes.(7) SBT can provide a safe learning environment with a range of difficulty levels and the opportunity for individualized learning.(8,9) One of the educational challenges is to provide meaningful feedback during (formative feedback) and after (summative feedback) each performance. This is typically given during direct supervision but risks being delivered only occasionally and/or in an unstructured way. Structured assessment tools have been introduced into many modern surgical curricula with the Objective Structured Assessment of

Technical Skills (OSATS) being a widely used format and having inspired numerous procedure and context specific assessment tools.(10) The SBT environment facilitates structured skills assessment because of the standardized conditions during practice and also allows for self-directed learning, where the trainee can practice until proficiency.(11)

In a time where many ORL residents are offered less time training in the operating room, SBT allows trainees to practice technical skills such as MT, better preparing them for further clinical training. This raises the question: how are ORL trainees trained in MT and what models for SBT and skills assessment are used? And—importantly—in the era of evidence-based surgical training: What is the supporting educational evidence? To the best of our knowledge, there is currently no systematic review on this topic.

This study aims to systematically review the literature on SBT models and assessment tools for MT and supporting educational evidence. Our goal is to equip ORL program directors and educational stakeholders with an overview of currently available models and assessment tools for SBT and assessment of MT technical skills.

MATERIALS AND METHODS

Our review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.(12)

Search strategy

The searches were conducted in April 2020 and updated in December 2020 using the following databases: PubMed, Embase/Ovid, Cochrane Library, Web of Science, and Directory of Open Access Journals (DOAJ).

The search strategy was a combination of MESH terms within the areas of interest for this review such as otitis media, otolaryngology, middle ear ventilation, otologic surgical procedures, computer simulation and none MESH terms such as otitis, tympanostomy, simulat*, virtual reality and assessment (full search strategy in Supplemental Digital Content I). We chose a broad search strategy as not to miss any relevant studies.

Inclusion and exclusion criteria

We included all types of descriptive and educational studies involving SBT of MT in addition to assessment tools regardless of they were used to assess skills performance on patients or physical/virtual reality (VR) simulation models. All types of participants were eligible including medical students, residents, experienced surgeons, and other health care professionals. Studies were excluded if full texts were not available through university library access, not in English, or were reviews, abstracts with insufficient information, or commentaries/letters.

Study selection

Search results were imported into an online tool for reviews.(13) Two reviewers (LH and SA) independently screened first the title and abstract, followed by screening of full texts of any article that was deemed potentially eligible by either reviewer. Articles identified through reference lists of the included articles as well as excluded reviews were also screened. Any disagreements were resolved by consensus.

Data extraction

Data were extracted using a data extraction form, which was piloted independently by both reviewers on five randomly selected papers. The final form included study information (authors, year, study design, etc.), characteristics of the simulation model (modality, materials), details of training/assessment (participants, type of training and assessment, training outcome and results). Microsoft Excel 2011 (Microsoft Corporation, Washington, USA) was used in the data extraction and management. The data was extracted independently by two authors (LH and SA) and disagreements were resolved by consensus.

Data synthesis, study quality assessment and analysis

Included studies were divided into 1) descriptive studies (i.e. studies describing only the development of a simulation model) and 2) educational studies (i.e. studies describing training, validation, and/or assessment). For the descriptive studies, we considered the reproducibility and cost of the models. For educational studies, two authors (LH and SA) independently used four frameworks/tools (detailed in Table 1) to appraise educational validity and quality: 1) Kirkpatrick's hierarchy of educational outcomes(14); 2) Messick's framework of validity evidence(15); 3) study quality using a quality assessment tool for educational studies developed by Gordon et al.(16); and 4) strength of conclusion as described by Yardley et al.(17) Meta-analysis was not planned due to the research question as well as expected heterogeneity of study design, models, and outcomes.

RESULTS

5,276 studies were screened by title/abstract after automated duplicate removal, hereof were 91 studies considered for full-text screening. Forty-six studies were included (Figure 1, flowchart) and study characteristics are provided in Table 2. Twenty-one studies(18-38) described only development/technical aspects of simulation models and were categorized as “descriptive”. Twenty-five studies(5,39-62) evaluated models for training/educational purposes or assessment of MT skills and were categorized as “educational”. Overall, we found 31 unique physical models and three different VR models.

Descriptive studies

All 21 descriptive studies detailed physical models with the exception of a single VR model.(26) Details on model materials, reproducibility, construction instructions, and cost is found in Table 3 (see Supplemental Digital Content II). We deemed the reproducibility easy for 15 of the models since they used common materials for example a cut cylindrical container covered with a glove placed in an empty Lidocaine bottle.(31) Three models were deemed of medium difficulty to replicate due to specially designed ear models(25,30) or the requirement of a specific wooden box.(20) Three models were considered difficult: One needed a tympanic membrane from a sheep(24), another a special manufactured steel box(21), and the third was a VR simulator.(26) Similarly, the cost for constructing most of the models were considered to be low except for models requiring special components or high-end computer equipment to run VR simulation.

Altogether, numerous papers provide clear instructions on how to build low-cost models out of easily available materials, making them easy to reproduce and potentially implement into training programs. However, supporting evidence of the educational value and use in training/assessment of technical skills in MT for these models is absent from the literature.

Educational studies

Study design & participants

Most of the 25 educational studies (Table 4, see Supplemental Digital Content III) detailed and evaluated physical models (N=18, 72%) with fewer studies on VR models (N=4, 16%). Three studies (12%) included performances on real patients in the OR to explore validity of different methods for technical skills assessment.

Altogether, the studies included 578 participants: 65 (11%) participants were expert otolaryngologists; 223 (39%) represented intermediate trainees such as ORL interns, residents and registrars; and 290 (50%) were novices (medical students and general surgery interns with no prior experience in MT). Consequently, much of the educational evidence supporting SBT of MT is based on novices' experience, which might be different from ORL trainees in relation to prior knowledge, skills and motivation.

Performance assessment

In the studies that included assessment of performance(39-43,45,47,49,51-55,59,60,62), different assessment methods and tools were used. These were mainly Global Rating Scales (GRS) where the rater for example assigns an overall score based on observation of the performance(39,41,42,45,51-55,62); Task-Based/Task-Specific Checklists (TBC/TSC) where the rater for example checks if specific sub-tasks are performed or errors made(41,42,51-53,55,62); and metrics such as time to completion of the procedure.(39-43,45,49,52-55,60) Two studies used completely automated metrics derived from the VR simulators such as procedure time, contact with the external ear canal, number of attempts of tube insertion etc.(47,59)

Two studies aimed primarily to explore the validity of a specific assessment method or tool: One study investigated the validity and reliability of video recordings in assessment of participants' performance using two blinded raters for whom a significant inter-rater correlation was demonstrated(39); the other study aimed to develop and collect validity evidence for a Task-Specific Checklist and Global Rating Scale and found a high inter-rater and intra-rater reliability.(55)

Altogether, several different assessment methods and tools have been used across the educational studies, only few have supporting validity evidence and none include evidence from all five sources in Messick's framework.

Outcomes of simulation-based training

A majority of the survey studies and one observational study had a Kirkpatrick level 1 educational outcome such as perceived realism of the model and/or opinion on its potential use for training.(5,44,46,48,56,57,61) Generally, the different models were perceived to be of moderate to high realism and potentially beneficial to trainees' skills acquisition. The remaining two survey studies included participants' self-assessed effects on knowledge and confidence after training using the models—corresponding to Kirkpatrick level 2a outcomes—and reported favorable responses.(50,58) None of these studies included objective measurements of performance or knowledge tests.

Seven studies with different study designs included measurement of the effects of SBT on skills performance(42,43,45,49,52-54), corresponding to Kirkpatrick level 2b outcomes, and a single recent study further examined transfer of skills to the operating room (OR), corresponding to a level

3 outcome.(62) One study found that SBT resulted in a significant improvement in overall performance of ORL trainees and experts on GRS(52), and another study demonstrated significant improvement in medical students' performance using GRS and TBC.(54) Both studies also reported that time to completion of the procedure in SBT decreased significantly: from 180 to 70 seconds and 5 to 2 min, respectively. A third study also reported improvement in ORL residents' GRS score and time to completion after SBT, however, it was not reported whether this was statistically significant.(42)

In three studies, participants were randomized to either SBT or no training.(45,53,62) Two of the studies demonstrated a significant effect of SBT on time to completion of the procedure with average improvements of 1.6 and 2.4 minutes, respectively.(45,53) One of these studies also reported a significant improvement in GRS from 12.9/25 points at baseline to 17.2/25 points after SBT(45) whereas another study found that the intervention group had significant better OSATS global scores (17.4/25 vs. 13.7/25) and OSATS task scores (4.5/5 vs. 3.6/5) in the posttest.(62) The third study demonstrated improvement in both groups' GRS and TSC but without statistically significant difference.(53) One of the studies further investigated transfer of skills acquired in the simulated environment to real-life surgery but found no difference in performance measured using an OSATS instrument between the SBT group and the control group during initial live surgery.(62)

Another randomized study investigated the effect of massed vs. interval training on skills performance and found that interval training was seemingly superior to massed practice in reducing time to procedure completion and errors in the simulation procedure even though this did not achieve statistical significance.(49) Also, one randomized study examined medical students' learning curves using either an endoscope or microscope when performing MT on a simulator but found no significant difference in learning rate (68.6 vs. 78.71 sec) between the two groups.(43)

Of the remaining 8 educational studies with level 2b educational outcomes, five studies compared the skills performance of novices and experts and four of the studies found the experts significantly outperformed the novices on time to completion and errors.(40,41,47,59,60) Finally, three studies assessed novices' and expert' performance in real-life surgery (without prior SBT) and similarly found that experts used significantly less time and achieved a higher performance score.(39,51,55)

Altogether, these studies support that SBT can reduce time to completion of the procedure, reduce errors, and improve total performance score in the simulation environment and that experts perform better than novices. Only a single and very recent study included higher-level educational outcomes (i.e. level 3–transfer to real life performance)(62) whereas none investigated the effects of training on organization (level 4a) or patient outcomes (level 4b).

Validity evidence

The contribution across educational studies to validity evidence of SBT and assessment of MT skills was analyzed according to the five sources of validity in Messick's framework.

Content validity: One study replicated five of the physical models described in other studies(20,23,28,38,53) which were then rated by ORL registrars/consultants in relation to educational quality: All models were found to be below moderate realism and below the expected threshold for recommendation of use in a training curriculum.(5) As described earlier, several other studies also surveyed experts and trainees in relation to model representation and usefulness for training.(40,44,46,57,61) Finally, three studies included content experts in the development of questionnaires or assessment tools.(44,55,61)

Response process: All the educational studies employed methods to eliminate or control potential sources of bias. This included for example providing participants with introduction such as handouts or video demonstrations to the simulator/procedure; standardized time to practice; and blinding of assessors.

Internal structure: Five studies explored aspects of reliability and internal consistency of their assessment(39,44,45,55,57): One study reported a good to excellent inter-rater and intra-rater reliability for a Global Rating Scale(45); another study found a significant correlation between the ratings of two blinded raters for a nine-parameter rating tool(39); and one study investigated the intraclass correlation and found this to be high.(55) The last two studies report a high internal consistency between participant ratings of simulation realism.(44,57)

Relations with other variables: Five studies found that simulation performances of participants with different levels of experience could be discriminated(40,41,47,59,60), supporting that experts still outperform novices in the simulation environment in relation to skills, errors, and time to completion.

Consequences: None of the studies examined for example the impact of test scores on trainees' advancement or set a pass/fail-standard for the procedure that can be used to determine when a trainee is proficient.

Study quality assessment

In our quality assessment, we found a mean score across the educational studies of 13.1 points (lower score indicating less sources of potential bias; minimum score 6, maximum score 18). Only six studies had a total score below the mean.(42,45,47,51,55,62) Only a single study was assessed as having a low risk of bias in relation to educational underpinning as all other studies lacked sufficient details on educational theory of frameworks guiding their study.(51)

Furthermore, most studies lacked a clear description of the curriculum's process/outcome or the educational context such as learner characteristics and consequently only two studies were deemed low risk of bias in relation to curriculum(45,55) and setting.(42,62) Also, only three studies provided a clear description of the specific content of the educational intervention and were deemed low risk of bias in relation to content.(42,55,62) None of the studies scored a low risk of bias in the pedagogical domain since none provided specific details on the pedagogical methods used such as examples of instructions. In contrast, a majority of the studies were deemed to have a low risk of bias in their conclusion meaning that the conclusion mostly reflected the findings of the study and did not overgeneralize findings. This was further corroborated by the "strength of conclusion" scoring tool, where twelve studies scored 3 (i.e. conclusions are probably supported by the results) and eleven studies scored 4 (i.e. results are clear and very likely to be true).

DISCUSSION

In this systematic review, we identified numerous models for SBT of MT in the literature with the oldest model dating back to 1968. Despite many of these being easy and inexpensive to reproduce, there is limited supporting educational evidence for their use in training and assessment. Many studies were purely descriptive and the remaining studies had focus on different educational aspects but were of varying quality. This results in scattered evidence relating to validity and effects of SBT of MT with gaps in knowledge on important educational outcomes such as transfer to real-life performance. Therefore, it is unsurprising that there is also a paucity in descriptions and evaluations of systematic implementation of SBT of MT into ORL training curricula.

A third of the educational studies reported only Kirkpatrick level 1 and 2a outcomes such as participants' perceptions of the models and self-reported effects on knowledge and confidence. However, such subjective measurements are poor educational outcomes as they convey very little on the actual effects of training(63), and self-assessment is well-known to be unreliable.(64) More objective measurements of effects on knowledge/skills (i.e. level 2b) were considered in 15 studies such as performance, errors, and/or time to completion of the procedure during SBT of MT, and all studies report positive effects. This finding is in line with a large meta-analysis of technology-enhanced simulation in health care professions education, which found that SBT had large and positive effects on knowledge, skills, and behaviors compared with no intervention.(7) Only one study investigated the transfer of skills from SBT to the OR (i.e. level 3)(62), highlighting that there still is a major gap in knowledge of whether SBT of MT has effects on transfer of skills to the OR and effects on patient outcomes. However, this is a common problem: A systematic review on SBT within otorhinolaryngology found that none of the 70 included studies on 64 different simulators considered educational outcomes above level 2b.(65)

Further, just one of the educational studies on SBT of MT so far has been comparative(5), meaning that most simulation models and modalities have not been compared. Also, only a single study compared different instructional designs of SBT of MT (massed vs. interval training) and found no statistically significant difference in time to completion or errors.(49) This contrasts other studies on skills acquisition of surgical procedures(66,67) including mastoidectomy(68) where distributed practice consistently has been found to be superior to massed practice. Further, many of the included studies used medical students as participants and findings might therefore not necessarily generalize to ORL because of differences in preexisting knowledge, experience, and motivation.

Valid and reliable assessment is key in establishing effects of training and we found several assessment tools for structured assessment of MT skills performance. These tools have mostly been modeled over established approaches such as Global Rating Scales and Task-Based Checklists. This, however, is no guarantee for their validity as assessment conditions and context matters(69) and specific validity studies are warranted. In MT skills assessment, validity evidence has mostly been gathered for inter-rater reliability, supporting acceptable reliability for moderate stakes assessment (reliability co-efficient between 0.80–0.89,(70)).

The validity evidence for SBT of MT is scattered across several different simulation models and assessment tools. None of the studies used a structured approach to explore validity using contemporary frameworks such as Messick's(15) or Kane's.(71) Unsurprisingly, only a single study therefore considered multiple sources of validity.(55) Systematic reviews from general surgery and ophthalmology similarly found that most educational studies used outdated concepts like concurrent validity rather than contemporary frameworks(72,73), which are now considered standard in medical educational research.(74) In our quality assessment, we further found that many studies missed key details on educational underpinning, curriculum, setting and pedagogical intervention. Two other reviews in simulation training in otolaryngology drew similar conclusions and also found considerable limitations in validation studies' methodology and consequently, the supporting evidence.(65,75)

What are the implications of our findings? Most importantly, we find that there is a need for knowledge and better application of established medical educational research methods in future studies. This will lead to a more solid evidence base for the effects of training and validity of assessment, and in turn propel implementation of SBT into practice. This could for example be how to use and structure SBT of MT in training curricula and define levels of proficiency in SBT which trainees need to document before advancing to supervised surgery in the OR. A key step is documenting transfer of MT skills such as SBT results in an increased initial performance in the OR.

The strength of our study is the systematic approach to reviewing the literature and including papers published since the inception of the databases until present, and using established medical educational frameworks in our appraisal of the current literature. A general limitation relates to publication bias in surgical education: many great and established training initiatives and activities are not validated or published in the scientific literature. Further, because of the heterogeneity of the literature relating to study design, outcomes, and quality prevents strong conclusions and best practice guidelines. Further, most studies were descriptive with no evaluation of the models' educational impact.

CONCLUSION

Altogether, it is reasonable to conclude that SBT models are effective in improving skills performance within the SBT environment compared with no training. However, there is currently no evidence to support one approach or model over another and many different physical and virtual reality simulation models are available for integration into the ORL curriculum. Evidence-based recommendations for SBT and assessment of MT require more studies on instructional design, validity and efficacy of these models.

REFERENCES

1. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. *N Engl J Med*. 2006;355(25):2664-2669. doi:10.1056/NEJMra054785

2. Hellstrom S, Groth A, Jorgensen F, et al. Ventilation tube treatment: a systematic review of the literature. *Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg*. 2011;145(3):383-395. doi:10.1177/0194599811409862

3. Heidemann C.H., Lous J., Berg J., et al. Danish guidelines on management of otitis media in preschool children. *Int J Pediatr Otorhinolaryngol*. 2016;87((Heidemann, Hakonsen, Jakobsen, Skrubeltang, Vind) Danish Health and Medicines Authority, Denmark):154-163. doi:10.1016/j.ijporl.2016.06.003

4. Rosenfeld RM, Schwartz SR, Pynnonen MA, et al. Clinical practice guideline: Tympanostomy tubes in children. *Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg*. 2013;149(1 Suppl):S1-35. doi:10.1177/0194599813487302

5. Mahalingam S., Awad Z., Tolley N.S., Khemani S. Ventilation tube insertion simulation: a literature review and validity assessment of five training models. *Clin Otolaryngol*. 2016;41(4):321-326. doi:10.1111/coa.12543

6. Sundhedsstyrelsen. National klinisk retningslinje for behandling med trommehindedræn – Diagnostik af mellemørebetændelse og indikationer for trommehindedræn mv. blandt børn i aldersgruppen 0-5 år. Published online May 21, 2015. Accessed July 20, 2020. http://www.sst.dk

7. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. 2011;306(9):978-988. doi:10.1001/jama.2011.1234

8. Sachdeva AK, Buyske J, Dunnington GL, et al. A new paradigm for surgical procedural training. *Curr Probl Surg*. 2011;48(12):854-968. doi:10.1067/j.cpsurg.2011.08.003

9. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 2005;27(1):10-28. doi:10.1080/01421590500046924

10. Sundhedsstyrelsen. Målbeskrivelse for speciallægeuddannelsen i Øre-Næse og Halssygdomme. Published online July 2015. Accessed July 20, 2020. http://www.sst.dk

11. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. 2002;236(4):458-463; discussion 463-464. doi:10.1097/00000658-200210000-00008

12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097

13. https://www.covidence.org/. *Covidence*.

14. Kirkpatrick DL. *Evaluating Training Programs the Four Levels*. 3rd ed. Berrett-Koehler; 2006.

15. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, & Joint Committee on Standards for

Educational and Psychological Testing. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.

16. Gordon M, Hill E, Stojan JN, Daniel M. Educational Interventions to Improve Handover in Health Care: An Updated Systematic Review. *Acad Med J Assoc Am Med Coll*. 2018;93(8):1234-1244. doi:10.1097/ACM.0000000000002236

17. 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

18. Abdel-Salam A., Gibb A.G. A grommet trainer. *J Laryngol Otol*. 1981;95(1):115-116. doi:10.1017/S0022215100090484

19. Baer S., Williams H., McCobe A. A model for instruction in myringotomy and grommet insertion. *Clin Otolaryngol Allied Sci*. 1990;15(4):383-384.

20. Carr E.R.M., Benjamin E. "Surgical skills box": A new training aid for surgical trainees. *J Laryngol Otol*. 2006;120(2):133-134. doi:10.1017/S0022215105000150

21. Chan CY, Lau DPC. Simulators and models for laryngeal laser surgery and laser myringotomy. *The Laryngoscope*. 2016;126(9):2089-2091. doi:10.1002/lary.25817

22. Chiesa Estomba C.M., Melendez Garcia J.M., Hamdam Zavarce M.I., Betances Reinoso F.A. The Vigo grommet trainer. *Eur Ann Otorhinolaryngol Head Neck Dis*. 2015;132(1):53-55. doi:10.1016/j.anorl.2014.06.003

23. Duijvestein M, Borgstein J. The Bradford grommet trainer. *Clin Otolaryngol Off J ENT-UK Off J Neth Soc Oto-Rhino-Laryngol Cervico-Facial Surg*. 2006;31(2):163. doi:10.1111/j.1749-4486.2006.01148.x

24. Hantman I. An ear manikin. Teaching and training device. *Arch Otolaryngol Chic Ill 1960*. 1968;88(4):407-412. doi:10.1001/archotol.1968.00770010409013

25. Holt G.R., Parel S.M., Shuler S.L. A model training ear for teaching paracentesis, myringotomy, and insertion of tympanostomy tubes. *Otolaryngol Head Neck Surg*. 1983;91(3):333-335.

26. Huang C, Agrawal SK, Ladak HM. Virtual Reality Simulator for Training in Myringotomy with Tube Placement. *J Med Biol Eng*. 2016;36(2):214-225. doi:10.1007/s40846-016-0124-1

27. Ismail-Koch H, Marshall L, Alabi T, Buckland J. The ear trainer model. *Clin Otolaryngol Off J ENT-UK Off J Neth Soc Oto-Rhino-Laryngol Cervico-Facial Surg*. 2008;33(6):635. doi:10.1111/j.1749-4486.2008.01745.x

28. Jesudason W.V., Smith I. How we do it: The Bradford grommet trainer: A model for training in myringotomy and grommet insertion. *Clin Otolaryngol*. 2005;30(4):371-373. doi:10.1111/j.1365-2273.2005.00978.x

29. Jia W, Al-Omari B. A versatile grommet trainer. *Clin Teach*. 2019;16(1):78-80. doi:10.1111/tct.12805

30. Kaul L. Myringotomy training device. *S D J Med*. 1969;22(6):27-28.

31. Leong A., Kundu S., Martinez-Devesa P., Aldren C. Artificial ear: A training tool for grommet insertion and manual dexterity. *ORL*. 2006;68(2):115-117. doi:10.1159/000091214

32. Neal S.L., Harris J.P., Davidson T.M. Artificial eardrum for instruction in myringotomy and PE tube insertion. *Laryngoscope*. 1985;95(8):1008-1009.

33. Owa A.O., Farrell R.W.R. Simple model for teaching myringotomy and aural ventilation tube insertion. *J Laryngol Otol*. 1998;112(7):642-643.

34. Paraiso R, Downey L. A simple method for teaching myringotomy and ventilation tube insertion. *Ear Nose Throat J*. 2011;90(3):94-96. doi:10.1177/014556131109000302

35. Qureishi A, Moualed D. The Northampton microsuction and grommet trainer. *Clin Otolaryngol Off J ENT-UK Off J Neth Soc Oto-Rhino-Laryngol Cervico-Facial Surg*. 2014;39(6):400-401. doi:10.1111/coa.12272

36. Singh PK, De M, Vaughan-Jones R. A model for training in grommet insertion. *Ann R Coll Surg Engl*. 2005;87(4):287-288.

37. Tuaño-Uy MJS, Martinez NV. Model Myringotomy Practice Set: A do-it-yourself and inexpensive alternative. *Philipp J Otolaryngol Head Neck Surg*. 2008;23(1):31-34. doi:10.32412/pjohns.v23i1.771

38. Walker T, Duvvi S, Kumar BN. The Wigan grommet trainer. *Clin Otolaryngol Off J ENT-UK Off J Neth Soc Oto-Rhino-Laryngol Cervico-Facial Surg*. 2006;31(4):349-350. doi:10.1111/j.1749-4486.2006.01268.x

39. Bowles P.F.D., Harries M., Young P., Das P., Saunders N., Fleming J.C. A validation study on the use of intra-operative video recording as an objective assessment tool for core ENT surgery. *Clin Otolaryngol*. 2014;39(2):102-107. doi:10.1111/coa.12240

40. Chan AL, Carrillo RJD, Ong KC. Simulation Platform for Myringotomy with Ventilation Tube Insertion in Adult Ears. *Philipp J Otolaryngol Head Neck Surg*. 2018;33(2):14-20. doi:10.32412/pjohns.v33i2.259

41. Clark M, Westerberg B, Mitchell J. Development and validation of a low-cost microsurgery Ear Trainer for low-resource settings. *J Laryngol Otol*. 2016;130(10):954-961. doi:10.1017/S0022215116008811

42. Clark MPA, Nakku D, Westerberg BD. An endoscopic Ear Trainer for the low-resource setting. *J Laryngol Otol*. 2019;133(7):571-574. doi:10.1017/S0022215119001257

43. Denton O, Daglish A, Smallman L, Fishpool S. Investigating the learning curve in endoscopic compared to microscopic myringotomy and ventilation tube insertion. *J Laryngol Otol*. 2020;134(6):497-500. doi:10.1017/S0022215120001188

44. Ho A.K., Alsaffar H., Doyle P.C., Ladak H.M., Agrawal S.K. Virtual reality myringotomy simulation with real-time deformation: Development and validity testing. *Laryngoscope*. 2012;122(8):1844-1851. doi:10.1002/lary.23361

45. Hong P, Webb A, Corsten G, et al. An anatomically sound surgical simulation model for myringotomy and tympanostomy tube insertion. *Int J Pediatr Otorhinolaryngol*. 2014;78(3):522-529. doi:10.1016/j.ijporl.2013.12.036

46. Huang C., Cheng H., Bureau Y., Agrawal S.K., Ladak H.M. Face and content validity of a virtual-reality simulator for myringotomy with tube placement. *J Otolaryngol - Head Neck Surg*. 2015;44(1):40. doi:10.1186/s40463-015-0094-2

47. Huang C., Cheng H., Bureau Y., Ladak H.M., Agrawal S.K. Automated Metrics in a Virtual-Reality Myringotomy Simulator: Development and Construct Validity. *Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol*. 2018;39(7):e601-e608. doi:10.1097/MAO.0000000000001867

48. Ilgner J, Park JJ-H, Westhofen M. Teaching microsurgery to undergraduate medical students by means of high-definition stereo video microscopy - the Aachen “skills lab” experience. In: Woods, AJ and Holliman, NS and Merritt, JO, ed. *STEREOSCOPIC DISPLAYS AND APPLICATIONS XIX*. Vol 6803. Proceedings of SPIE. SPIE; IS&T, Soc Imaging Sci & Technol; IMAX; REAL D; Christie Digital Syst; JVC Professional; DepthQ Stereoscop; Kerner Opt LLC; 3D Biz Ex; MTBS3D; Veritas Visus; Stereo World; 2008. doi:10.1117/12.763867

49. Kesser B.W., Hallman M., Murphy L., Tillar M., Keeley M., Peirce S. Interval vs massed training: How best do we teach surgery? *Otolaryngol - Head Neck Surg U S*. 2014;150(1):61-67. doi:10.1177/0194599813513712

50. Kovatch KJ, Wertz AP, Carle TR, et al. Optimal Timing of Entry-Level Otolaryngology Simulation. *OTO Open*. 2019;3(2):2473974X19845851. doi:10.1177/2473974X19845851

51. Labbe M, Young M, Mascarella M, Husein M, Doyle PC, Nguyen LHP. How Consistent Is Competent? Examining Variance in Psychomotor Skills Assessment. *Acad Med J Assoc Am Med Coll*. Published online September 10, 2019. doi:10.1097/ACM.0000000000002985

52. Luu K., Straatman L., Nakku D., Westerberg B., Carter N., Clark M. Evaluation of a low-fidelity ear surgery simulator in a low-resource setting. *J Laryngol Otol*. 2017;131(11):1010-1016. doi:10.1017/S002221511700216X

53. 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(7):1410-1415. doi:10.1002/lary.21801

54. Molin N, Chiu J, Liba B, Isaacson G. Low cost, easy-to-replicate myringotomy tube insertion simulation model. *Int J Pediatr Otorhinolaryngol*. 2020;131:109847. doi:10.1016/j.ijporl.2019.109847

55. Schwartz J., Costescu A., Mascarella M.A., et al. Objective assessment of Myringotomy and tympanostomy tube insertion: A prospective single-blinded validation study. *Laryngoscope*. 2016;126(9):2140-2146. doi:10.1002/lary.25746

56. Shenton C, Aucott W. How I do it: a simulator of the ear for developing otomicroscopy skills during the coronavirus disease 2019 pandemic. *J Laryngol Otol*. Published online October 21, 2020:1-4. doi:10.1017/S0022215120002261

57. Sowerby L.J., Rehal G., Husein M., Doyle P.C., Agrawal S., Ladak H.M. Development and face validity testing of a three-dimensional myringotomy simulator with haptic feedback. *J Otolaryngol - Head Neck Surg*. 2010;39(2):122-129. doi:10.2310/7070.2009.090079

58. Sparks D, Kavanagh KR, Vargas JA, Valdez TA. 3D printed myringotomy and tube simulation as an introduction to otolaryngology for medical students. *Int J Pediatr Otorhinolaryngol*. 2020;128:109730. doi:10.1016/j.ijporl.2019.109730

59. 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(3):343-345. doi:10.1177/0194599820901684

60. Volsky P.G., Hughley B.B., Peirce S.M., Kesser B.W. Construct validity of a simulator for myringotomy with ventilation tube insertion. *Otolaryngol Head Neck Surg*. 2009;141(5):603. doi:10.1016/j.otohns.2009.07.015

61. Wheeler B., Doyle P.C., Chandarana S., Agrawal S., Husein M., Ladak H.M. Interactive computer-based simulator for training in blade navigation and targeting in myringotomy. *Comput Methods Programs Biomed*. 2010;98(2):130-139. doi:10.1016/j.cmpb.2009.09.010

62. 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. doi:10.1177/0194599820920833

63. Norman G. Data dredging, salami-slicing, and other successful strategies to ensure rejection: twelve tips on how to not get your paper published. *Adv Health Sci Educ Theory Pract*. 2014;19(1):1-5. doi:10.1007/s10459-014-9494-8

64. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med J Assoc Am Med Coll*. 2005;80(10 Suppl):S46-54. doi:10.1097/00001888-200510001-00015

65. Musbahi O., Aydin A., Al Omran Y., Skilbeck C.J., Ahmed K. Current Status of Simulation in Otolaryngology: A Systematic Review. *J Surg Educ*. 2017;74(2):203-215. doi:10.1016/j.jsurg.2016.09.007

66. Moulton C-AE, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Ann Surg*. 2006;244(3):400-409. doi:10.1097/01.sla.0000234808.85789.6a

67. Mackay S, Morgan P, Datta V, Chang A, Darzi A. Practice distribution in procedural skills training: a randomized controlled trial. *Surg Endosc*. 2002;16(6):957-961. doi:10.1007/s00464-001-9132-4

68. 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

69. 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. *Acad Med J Assoc Am Med Coll*. 2020;95(12):1929-1936. doi:10.1097/ACM.0000000000003550

70. Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ*. 2004;38(9):1006-1012. doi:10.1111/j.1365-2929.2004.01932.x

71. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560-575. doi:10.1111/medu.12678

72. Ghaderi I, Manji F, Park YS, 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

73. Thomsen ASS, Subhi Y, Kiilgaard JF, la Cour M, Konge L. Update on simulation-based surgical training and assessment in ophthalmology: a systematic review. *Ophthalmology*. 2015;122(6):1111-1130.e1. doi:10.1016/j.ophtha.2015.02.028

74. 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. doi:10.1007/s10459-013-9458-4

75. Arora A., Lau L.Y.M., Awad Z., Darzi A., Singh A., Tolley N. Virtual reality simulation training in Otolaryngology. *Int J Surg*. 2014;12(2):87-94. doi:10.1016/j.ijsu.2013.11.007

FIGURE LEGENDS

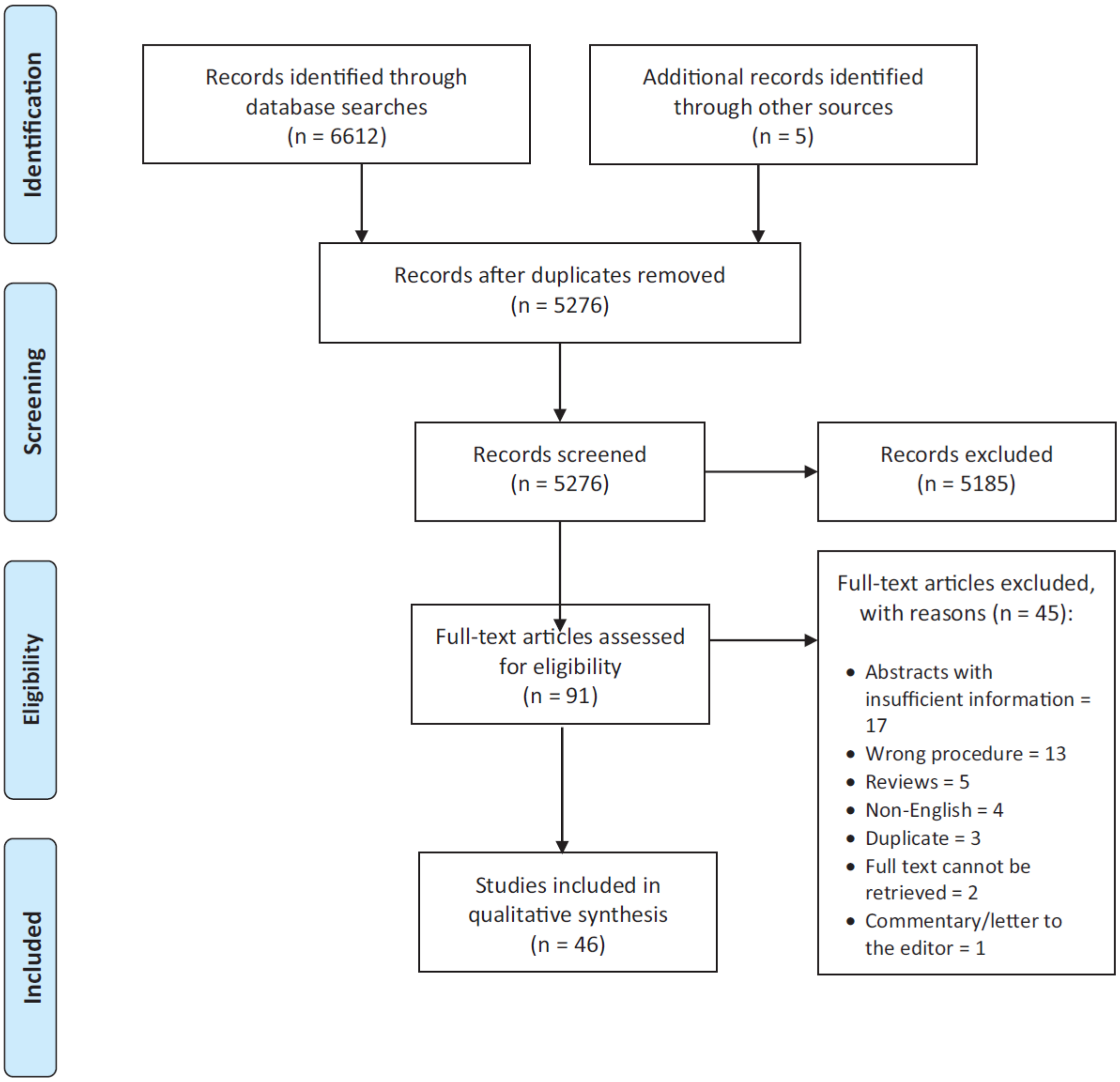

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.

Table 1. The frameworks used to assess the quality and validity evidence of the included educational studies.

Framework	**Description**	**Categories**	**Definition and examples**	**Score**
Kirkpatrick's hierarchy (14)	The impact of the educational interventions	Level 1 - participation	The participant's view on the learning intervention, usually evaluated by surveys.	n/a
		Level 2a – modification of attitudes/perceptions	Changes in the participant's attitude or perception of the intervention such as self-assessment of confidence level in a procedure.	
		Level 2b – Modification of knowledge/skills	The acquisition of knowledge and/or skills for example testing the effect of simulation training on actual skills performance or on a knowledge-test.	
		Level 3 – Behavioral change	Transfer of skills for example from simulation to the operating room (OR).	
		Level 4a – change in organizational practice	Changes in the organization for example following the implementation of an educational curriculum.	
		Level 4b – benefits to patients/clients	Effects on patients' outcome that results from the educational intervention for example a reduction in complication rates after a training program has been introduced.	
Messick's Framework of validity (15)	Validity of assessment tools	Content	The relevance of test content when compared to the construct it is intended to measure i.e. ensuring that the assessment is representative.	Each category is scored from 1–5 points.
		Response process	Elimination or control of potential sources of bias i.e. standardized	

		Internal structure Relations to other variables Consequences	written instructions, quality control of scoring etc. The reproducibility or generalizability of the assessment tool i.e. reliability analysis. The correlation with external factors, i.e. if test scores can discriminate between different levels of experience. The intended and unintended effects of the test i.e. reasonableness of method to establish a pass/fail-standard.	
Quality assessment tool (16)	The quality of methodology based on the risk of bias in 6 different educational sources	Educational underpinning Curriculum Setting Pedagogical Content Conclusion	The use of an educational framework or mentioning of the educational theory underpinning the intervention. Description of the training program, the training process/outcomes and/or assessment design. The educational settings in which the intervention was conducted and characteristics of the relevant participants. The use of pedagogical methods to deliver the intervention to the participants for example instruction/guidance. Details on the materials that is included in the intervention – handouts, questionnaire etc. The degree the conclusion reflects the findings of the study.	Each category is scored 1–3, with: Low risk = 1 Unclear risk = 2 High risk = 3

Strength of conclusion (17)	The trustworthiness of the findings		1: No clear conclusions can be drawn, not significant. 2: Results ambiguous, but there appears to be a trend. 3: Conclusions are probably supported by the results. 4: Results are clear and very likely to be true. 5: Results are unequivocal.	1–5

Table 2. Characteristics of the included studies

Study characteristics	No. of studies	No. of participants
All studies	46	578
Study design		
Descriptive	21	0
Single group pre-post	2	28
Interventional	1	10
Randomized	5	219
Observational	9	155
Survey study	8	166
*Participants**		
Medical students	8	270
General surgery interns	1	20
ORL interns/residents/registrars	18	223
Otolaryngologists (experts)	14	65
Training modality		
Physical model	37	475
Technology-enhanced model (VR models)	6	52
Real-life surgery	3	51
Technical skill		
Myringotomy and ventilation tube insertion	41	549
Only myringotomy	5	29
Outcomes (Kirkpatrick's hierarchy)		
Not applicable (model description only)	21	0
Level 1 - Participation	7	83
Level 2a - Modification of attitudes/perceptions	2	103
Level 2b - Modification of knowledge/skills	15	329
Level 3 - Behavioral change	1	63
Level 4a - Changes in organizational practice	0	0
Level 4b - Benefits to patients	0	0

*Not all studies included participants and some studies included more than one group of learners.

SUPPLEMENTAL DIGITAL CONTENT

Supplemental Digital Content I.doc

Supplemental Digital Content II.doc

Supplemental Digital Content III.doc

Supplemental Digital Content IIII.doc