Full citation: Andersen SAW, Guldager M, Mikkelsen PT, Sørensen MS. The effect of structured self-assessment in virtual reality simulation training of mastoidectomy. *Eur Arch Otorhinolaryngol*. 2019 Dec;276(12):3345-3352.

DOI: 10.1007/s00405-019-05648-6

Title: The effect of structured self-assessment in virtual reality simulation training of mastoidectomy.

Authors and affiliations:

Steven Arild Wuyts Andersen[1,2] (ORCID ID: 0000-0002-3491-9790)

Mads Guldager[1,2]

Peter Trier Mikkelsen[3]

Mads Sølvsten Sørensen[1]

1. Department of Otorhinolaryngology—Head & Neck Surgery, Rigshospitalet, Copenhagen, Denmark.
2. Copenhagen Academy for Medical Education and Simulation (CAMES), Centre for HR, the Capital Region of Denmark.
3. The Alexandra Institute, Aarhus, Denmark.

Correspondence: Steven A. W. Andersen, MD, PhD. Department of Otorhinolaryngology—Head & Neck Surgery, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark. Phone: 0045 20612006. E-mail: stevenarild@gmail.com.

ABSTRACT

Purpose: Virtual reality (VR) simulation surgical skills training is well-established but self-directed practice is often associated with a learning curve plateau. In this study we investigate the effects of structured self-assessment as a means to improve performance in mastoidectomy training.

Methods: The study was a prospective, educational study. Two cohorts of novices (medical students) were recruited for practice of anatomical mastoidectomy in a training program with five distributed training blocks. Fifteen participants performed structured self-assessment after each procedure (intervention cohort). A reference cohort of another 14 participants served as controls. Performances were assessed by two blinded raters using a modified Welling Scale and simulator-recorded metrics.

Results: The self-assessment cohort performed superiorly to the reference cohort (mean difference of final product score 0.87 points, p=0.001) and substantially reduced the number of repetitions needed. The self-assessment cohort also had more passing performances for the combined metrics-based score reflecting increased efficiency. Finally, the self-assessment cohort made fewer collisions compared with the reference cohort especially with the chorda tympani, the facial nerve, the incus, and the malleus.

Conclusions: VR simulation training of surgical skills benefits from having learners perform structured self-assessment following each procedure as this increases performance, accelerates the learning curve thereby reducing time needed for training, and induces a safer performance with fewer collisions with critical structures. Structured self-assessment was in itself not sufficient to counter the learning curve plateau and for continued skills development additional supports for deliberate practice are needed.

Keywords: self-assessment, temporal bone surgery, mastoidectomy, virtual reality surgical simulation, tutoring, simulation-based training.

INTRODUCTION

Virtual reality (VR) simulation is an evidence-based tool for temporal bone surgical skills training and is often used in conjunction with other training modalities such as temporal bone dissection on human cadaveric temporal bones [1]. VR simulation training is increasingly adopted into training curricula as it offers a relatively inexpensive setup compared with maintaining open laboratory facilities for cadaveric dissection [2,3]. Furthermore, skills acquired in VR temporal bone surgical simulation transfers to increased dissection skills [3–6], which allows better use of donated specimens for refinement of skills after initial VR simulation training. In addition, VR simulation allows an unlimited number of bones of different anatomies to be drilled as well as direct access to on-screen learning supports to facilitate trainees' learning. Altogether, VR simulation offers an attractive platform for fundamental temporal bone surgical skills training.

One of the potential advantages but also challenges of VR surgical simulation is that of self-directed training: the trainee can easily practice whenever they like but this is often without the presence of faculty for guidance and feedback. Instead the accessibility of VR simulation better supports distributed practice, which is beneficial for learning and skills consolidation [6,7]. In contrast to faculty-led training during for example temporal bone dissection courses, other mechanisms are needed to ensure development of adequate and safe skills in self-directed simulation-based training. According to the educational framework of directed, self-regulated learning (DSRL) [8,9] such self-directed simulation-based training therefore requires strong learning supports to scaffold learning and the development and progression of skills.

In VR simulation training of mastoidectomy, it has been demonstrated that the learning curve has an initial plateau after 4–9 procedures and that simulator-integrated tutoring can lead to tutoring overreliance. [7–10]. Deliberate practice is needed to overcome this performance plateau, and without direct external feedback during self-directed practice, the learner has to rely on reflection and evaluation of their own performance [11]. However, inadequate self-assessment skills of the learner seems to be a major issue resulting in the early learning curve plateau: the self-determined proficiency level of novices often do not match that of external assessment, i.e. they do not perform as well as they think so themselves [12,13]. Critical to the self-regulation of learning is accurate self-assessment skills [14]. Self-assessment abilities vary with the domain being self-assessed, but for surgical technical skills, some improvement can be seen with repeated practice of the surgical task [15].

In the case of temporal bone surgery, the early plateau suggests that repetition in itself is a less efficient approach to increase self-assessment skills. Other approaches are therefore needed to keep novices cognitively investing in further improving their skills. The literature on DSRL emphasizes providing explicit and specific process goals and directive instructions to support the self-guided learning.16 Increasing the learners' awareness of goal-setting and self-assessment, could potentially result in the desired continuous cognitive effort, support deliberate practice, and counter the initial learning curve plateau. Consequently, we constructed a structured tool for self-assessment (an 8-item rating form supported by small videos) to support self-assessment and the formation of process goals during VR simulation training of mastoidectomy. In this study, we wanted to investigate the effects on mastoidectomy final-product performance and simulator metrics of using such a structured self-assessment approach.

MATERIALS AND METHODS

Setting and participants

A cohort of 15 medical students from both pre-clinical and clinical semesters at the University of Copenhagen, Denmark, was recruited for VR simulation training with a learning intervention (structured self-assessment) in March–June 2017. A reference cohort of 14 medical students was recruited for similar VR simulation training in the period Oct. 2016–Jan. 2017. A single participant dropped out of the reference cohort due to lack of time to complete, whereas none dropped out in the intervention cohort (Flow chart, Figure 1). None of the participants had previously received any VR temporal bone surgical simulation training (exclusion criterion), nor had they had any other hands-on experience with temporal bone surgery. Participation was considered a voluntary, non-credit, extracurricular activity. Data collection took place at the Simulation Centre at Rigshospitalet, Copenhagen, Denmark.

Study design

The study was designed as a prospective, educational, cohort study with comparison between an intervention cohort and a reference cohort. All participants provided background data and then received a brief, hands-on introduction to the VR simulator on how to navigate the simulator. Training of both cohorts consisted of distributed practice with five training blocks spaced by at least a week. Each training block consisted of three completely identical procedures (anatomical mastoidectomy and posterior tympanotomy) on the same virtual temporal bone model. Training in the VR simulator was self-directed and participants in both cohorts had access to the on-screen step-by-step guide of the procedure at all times for reference.

The interventional cohort was additionally introduced to the structured self-assessment with a complete 6-minute video demonstrating the excellent and poor performance in relation to key areas of the mastoidectomy. At any point during training or self-assessment they could review this video again or view a brief 1-minute summary video as well as brief videos for each of the individual 8 items on the self-assessment rating form. The self-assessment form (Supplemental material, Appendix 1) was developed specifically for this study and we based it on the modified Welling Scale (WS1) [17]. In contrast to the dichotomous rating of the modified WS1, we allowed each item for self-assessment to be rated using a 5-point Likert Scale with a numerical score (1–5) with descriptive anchors for the two extremes and the middle score. Participants in the intervention cohort were asked to complete a self-assessment form immediately after each procedure while their drilled temporal bone was kept on screen for reference.

VR simulation platform and data sampling

In this study, an experimental version 2.1 of the VR temporal bone surgical simulator, the Visible Ear Simulator, was used [18,19]. The simulator is academic freeware [20] and can be run from PCs with a GeForce GTX graphics card (Nvidia, Santa Clara, CA, USA). The Geomagic Touch haptic device (3D systems Inc., Rock Hill, SC, USA) provides interaction with force feedback. The experimental version of the simulator recorded data on 128 different metrics and derivatives [21]. In addition, simulation final-products were saved for later performance evaluation by two raters (SA and MS) using the modified WS1 [17]. Raters were blinded to participant, intervention, cohort, procedure number, time used, simulator metrics, and the metrics-based score.

Data analysis, outcomes and statistics

Sample size was based on previous studies [7,17]. Analyses were performed in SPSS (SPSS Inc., IL, USA) version 23 for MacOS X. Due to repeated measurements, linear mixed models were used for both final-product performance analysis (procedure number, rater, and cohort as fixed factors) and for the analysis of simulator-gathered metrics (procedure number and cohort as fixed factors). The compound metrics-based score (MBS) was calculated as reported previously [21]. For the learning curve graphs, estimated marginal means with 95 % CI were plotted.

Ethics

The study was deemed exempt by the regional ethics committee for the Capital Region of Denmark (H-17002805). All participants volunteered for the study and signed informed consent before participation.

RESULTS

The participants in the intervention and reference cohorts had comparable background and experience such as average computer usage per week, self-rated computer skills, and gaming frequency (Table 1). As per the only exclusion criterion, none of the participants had any experience with temporal bone surgical simulation.

Effects of structured self-assessment on final-product performance

For both cohorts, final-product performance increased with repeated practice. The self-assessment cohort performed superiorly to the reference cohort (linear mixed models, mean difference 0.87 points, p=0.001). The level of performance of the control cohort after 15 repetitions was reached by the intervention cohort after only 8 repetitions (Figure 2). The gap between the learning curves was reduced with repeated practice and started to plateau at a level of about 19 out of 26 points. The self-assessment group used slightly more time for each procedure (mean diff 1.5 min, p<0.01) (Supplemental Figure 1), resulting in the final-product score per time being similar for the two cohorts (Supplemental Figure 2).

For most of the individual items of the modified WS1, the self-assessment cohort significantly outperformed the reference cohort (Supplemental Table 1), but especially relating to the items concerning the outer boundaries of the procedure: exposing the sigmoid sinus while leaving no overhang (item 7), drilling the sinodural angle sharp (item 10) and without cells remaining (item 11), exposing tegmen mastoideum (item 14), exposing the digastric ridge (item 17) and leaving no remaining cells in the mastoid tip (item 19), and finally, exposing the tympanic chorda (item 25). For exposing the attic/tegmen tympani (item 12), the reference cohort performed significantly better than the self-assessment cohort.

Effects of self-assessment on simulator metrics

For the compound metrics-based score (MBS), calculated on the basis of 17 different evidence-based metrics and reflecting mainly efficiency [21], both cohorts performed similarly and had identical learning curves based on estimated marginal means (Figure 3). However, the self-assessment cohort had significantly more performances that passed the proficiency level of 83.6 % by the end of training than the reference cohort. 100 % of the performances by participants in the self-assessment cohort passed the minimum volume to be removed during the procedure (based on the minimum volume drilled by experts), whereas only 93.4 % of the performances by the reference cohort fulfilled this criterion.

For the individual metrics, a number of significant and positive effects of structured self-assessment was found (Table 2 and Supplemental Table 2): the self-assessment cohort removed more bone, drilled less while not being in contact with bone, had a shorter path length, and made fewer collisions with especially the chorda, the facial nerve, the incus, and the malleus. Of equivocal effects, the self-assessment cohort used more force on all types of burrs, and more force on small (0.5–2 mm) and medium (3-4 mm) size burrs, and removed less bone using the small burrs compared with the reference cohort. The largest negative effect of structured self-assessment was on the number of “drill jumps”: the self-assessment cohort had a tendency to drill a little at one place, then jump to another location >5 mm away and drill some more, and then jump on to the next area and so forth. In contrast, experts make fewer jumps [21] most likely because they will finish drilling one sub-goal before moving on to the next sub-goal [22].

DISCUSSION

In this prospective, educational, cohort study on the effect of structured self-assessment on performance in VR simulation training of mastoidectomy, we found that the intervention increased final-product performance and accelerated the learning curve. This resulting in the need for substantially fewer repetitions, corresponding to a reduction in training time of approximately three hours. We further found that structured self-assessment induced learning a safer performance resulting in fewer collisions with critical structures such as the chorda and facial nerve and the ossicles.

We purposely did not design the present study to investigate the accuracy of self-assessment compared with the expert ratings: First of all, the study participants were medical students with no experience in temporal bone surgery. Secondly, the assessment tool—even though based to some degree on assessment items used for the experts' ratings—was designed to enhance learning by cognitive engagement and formation of sub-goals in the novice learners. It is therefore not a surprise that the self-assessment scores correlated only weakly with the scores assigned by the raters (Pearson's r=0.45). A large systematic review concluded that physicians mostly have a limited ability to accurately self-assess [23], and in general, self-assessment has limited use as a measurement of performance in medical educational studies [24].

Even though self-assessment data has limited usefulness as a learning outcome, the use of structured self-assessment as a learning support and a way to improve performance in simulation-based surgical skills training is under-investigated. In a recent study, the mastoidectomy performances of 16 otolaryngology residents on cadavers were recorded, after which participants were randomized to self-assess their video recordings using a structured tool or not before they performed a second procedure at a later time [25]. Self-assessment of the first performance did not improve performance on the second procedure compared with the control group. In contrast, our study had multiple instances of structured self-assessment during distributed training program. However, it remains uninvestigated how structured self-assessment affects different levels of learners and the effect on otorhinolaryngology trainees might be different from what we find for medical students.

The value of self-assessment as an integral part of adult learning was acknowledged by Reznick in his key paper on surgical skills training and testing from 1993 [26], and also later emphasized by Eva and Regehr as an "*important mechanism for ensuring safe and effective performance*" [24]. Our data substantiates that some of the positive effects of using structured self-assessment during training was indeed related to measures of efficiency and especially safety, i.e. avoiding collisions with vital structures. There is little doubt, that accurate self-assessment skills are essential for effective directed, self-regulated learning and should be considered in the design of self-directed learning experiences [14]. This also highlights that increased active participation and a higher meta-cognitive awareness of the objectives of training is better for learning.

Limitations of this study includes the cohort study design and that participants were not randomized for an intervention and control arm. However, participants had similar background data and due to identical recruitment process, simulator-setup, training program, and data collection, we have no reason to believe that this influences results or conclusions. Another limitation is that we used medical students rather than residents, who unfortunately are far too few and often not true novices. With a large number of metrics analyzed (128 in total) there is an inherent risk of mass-significance due to multiple comparisons. We have chosen not to perform adjustments such as Bonferroni correction, as this is very conservative, but have in the results and discussion only put weight on metrics with a substantial statistical significance ($p<0.005$) or multiple sources that indicates the same result. A strength of the study is the repeated measurements design and the combination of both human ratings of performance and simulator-generated metrics. Finally, it should be noted that the intervention consisted of both the use of the structured rating form for self-assessment and supporting small videos to demonstrate key aspects to guide the self-assessment. The effect of one without the other can therefore not be discerned. It is important for self-directed training to be guided and we used the videos to introduce the structured self-assessment and as a reference to illustrate the items of the rating form and differences between an adequate and inadequate performance on the different items. However, besides the mandatory introduction using the videos, the participants self-regulated their subsequent use of the videos and it is our impression that they only rarely re-watched them later in their training. Consequently, we suspect that the videos only contributed little to the effect on the learning curve of the intervention cohort.

The underlying mechanisms for the positive effect of structured self-assessment needs further studies: we implemented the structured self-assessment tool as a means to improve cognitive engagement and self-regulation of learning, but whether this is the actual mechanism of action needs future analysis. The increased number of drill jumps could indicate that structured self-assessment did not cause participants to complete one area of drilling before moving on to the next. Even though structured self-assessment had a number of positive effects, it did not—at least with our current approach—have a continued effect on performance and the learning curve plateaued at a level similar to that of the control cohort. Other methods for improving cognitive engagement and performance still needs exploration. Most likely, external sources of feedback and other learning supports are needed to further encourage deliberate practice [27]. In self-directed VR simulation training, this could for example be the use of simulator-integrated tutoring [7], an instructional design with task- and case variation to gradually increase challenges [28], automated formative or summative feedback based on simulator metrics [29], and/or game elements i.e. gamification [30].

CONCLUSION

VR simulation surgical training often constitutes self-directed training and this requires a strong instructional design and learning supports. This is necessary to scaffold learning and important components are supporting the formation of process goals and self-assessment skills. Structured self-assessment had several positive effects in VR simulation training of temporal bone surgery: first of all, it increased the performance both as assessed by external raters and based on a metrics-based score. Secondly, it accelerated the learning curves and consequently reduced the time needed for training. Finally, structured self-assessment induced a safer performance with significantly fewer collisions with critical structures such as the facial and chorda nerves. The underlying mechanisms for these positive effects remains to be further explored but could possibly be related to an increased cognitive effort during the simulation training. Finally, other learning supports are needed to further increase performance and support deliberate practice to overcome the learning curve plateau.

Funding: The general development of the Visible Ear Simulator software has been financially supported by the Oticon Foundation. The foundation did not play a role in the design or conduct of the study.

Conflict of interest: The authors declare that they have no conflict of interest.

REFERENCES

1. Frithioff A, Sørensen MS, Andersen SAW. European status on temporal bone training: a questionnaire study. *Eur Arch Otorhinolaryngol*. 2018;275(2):357–363. doi: 10.1007/s00405-017-4824-0

2. 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

3. Andersen SAW, Foghsgaard S, Konge L, Caye-Thomasen P, Sørensen MS. The effect of self-directed virtual reality simulation on dissection training performance in mastoidectomy. *Laryngoscope*. 2016;126(8):1883–1888. doi: 10.1002/lary.25710

4. Wiet GJ, Stredney D, Kerwin T, et al. Virtual temporal bone dissection system: OSU virtual temporal bone system. *Laryngoscope*. 2012;122(S1):S1–S12. doi: 10.1002/lary.22499

5. Zhao YC, Kennedy G, Yukawa K, Pyman B, Stephen O'Leary. Can Virtual reality simulator be used as a training aid to improve cadaver temporal bone dissection? Results of a randomized blinded control trial. *Laryngoscope*. 2011;121(4):831–837. doi: 10.1002/lary.21287

6. Andersen SAW, Foghsgaard S, Caye-Thomasen P, Sørensen MS. The Effect of a Distributed Virtual Reality Simulation Training Program on Dissection Mastoidectomy Performance. *Otol Neurotol*. 2018;39(10):1277–1284. doi: 10.1097/MAO.0000000000002031

7. Andersen SA, Konge L, Caye-Thomasen P, Sørensen MS. Learning Curves of Virtual Mastoidectomy in Distributed and Massed Practice. *JAMA Otolaryngol Head Neck Surg*. 2015;141(10):913–918. doi: 10.1001/jamaoto.2015.1563

8. Brydges R, Nair P, Ma I, Shanks D, Hatala R. Directed self-regulated learning versus instructor-regulated learning in simulation training. *Med Educ*. 2012;46(7):648–656. doi: 10.1111/j.1365-2923.2012.04268.x

9. Brydges R, Manzone J, Shanks D, et al. Self-regulated learning in simulation-based training: a systematic review and meta-analysis. *Med Educ*. 2015;49(4):368–378. doi: 10.1111/medu.12649

10. Nash R, Sykes R, Majithia A, Arora A, Singh A, Khemani S. Objective assessment of learning curves for the Voxel-Man TempoSurg temporal bone surgery computer simulator. *J Laryngol Otol*. 2012;126(07):663–669. doi: 10.1017/S0022215112000734

11. Ericsson KA. Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Acad Med*. 2004;79(Supplement):S70–S81.

12. Jowett N, LeBlanc V, Xeroulis G, MacRae H, Dubrowski A. Surgical skill acquisition with self-directed practice using computer-based video training. *Am J Surg* 2007;193:237–242. doi: 10.1016/j.amjsurg.2006.11.003

13. Andersen SA, Konge L, Mikkelsen PT, Cayé-Thomasen P, Sørensen MS. Mapping the plateau of novices in virtual reality simulation training of mastoidectomy. *Laryngoscope*. 2017;127(4):907–914. doi: 10.1002/lary.26000

14. Brydges R, Dubrowski A, Regehr G. A new concept of unsupervised learning: directed self-guided learning in the health professions. *Acad Med.* 2010;85:S49–S55. doi: 10.1097/ACM.0b013e3181ed4c96

15. MacDonald J, Williams RG, Rogers DA. Self-assessment in simulation-based surgical skills training. *Am J Surg*. 2003;185:319–322. doi: 10.1016/s0002-9610(02)01420-4

16. Brydges R, Carnahan H, Safir O, Dubrowski A. How effective is self-guided learning of clinical technical skills? It's all about process. *Med Educ*. 2009;43:507–515. doi: 10.1111/j.1365-2923.2009.03329.x

17. Andersen SA, 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

18. Trier P, Noe KØ, Sørensen MS, Mosegaard J. The visible ear surgery simulator. *Stud Health Technol Inform*. 2008;132:523–525.

19. Sørensen MS, Mosegaard J, Trier P. The visible ear simulator: a public PC application for GPU-accelerated haptic 3D simulation of ear surgery based on the visible ear data. *Otol Neurotol*. 2009;30(4):484–487. doi: 10.1097/MAO.0b013e3181a5299b

20. The Visible Ear Simulator. Available from http://ves.alexandra.dk. [Last accessed 6 September 2019].

21. Andersen SA, Mikkelsen PT, Sørensen MS. Expert sampling of VR simulator metrics for automated assessment of mastoidectomy performance. *Laryngoscope.* 2019;129(9):2170-2177.. doi: 10.1002/lary.27798

22. Sewell C, Morris D, Blevins NH, Barbagli F, Salisbury K. Evaluating drilling and suctioning technique in a mastoidectomy simulator. *Stud Health Technol Inform.* 2007;125:427-432.

23. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of Physician Self-assessment Compared With Observed Measures of Competence. *JAMA*. 2006;296(9):1094–1102. doi: 10.1001/jama.296.9.1094

24. Eva KW, Regehr G. Self-Assessment in the Health Professions: A Reformulation and Research Agenda. *Acad Med*. 2005;80(Supplement):S46–S54.

25. Jethwa AR, Perdoni CJ, Kelly EA, Yueh B, Levine SC, Adams ME. Randomized Controlled Pilot Study of Video Self-assessment for Resident Mastoidectomy Training. *OTO Open*. 2018;2(2):2473974X1877041-X1877045. doi: 10.1177/2473974X18770417

26. Reznick RK. Teaching and testing technical skills. *Am J Surg*. 1993 Mar;165(3):358–61. doi: 10.1016/s0002-9610(05)80843-8

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

28. Piromchai P, Ioannou I, Wijewickrema S, et al. Effects of anatomical variation on trainee performance in a virtual reality temporal bone surgery simulator. *J Laryngol Otol*. 2017;131(S1):S29–S35. doi: 10.1017/S0022215116009233

29. Al-Shahrestani F, Sørensen MS, Andersen SAW. Performance metrics in mastoidectomy training: a systematic review. *Eur Arch Otorhinolaryngol*. 2019 Mar;276(3):657–664. doi: 10.1007/s00405-018-05265-9

30. Rutledge C, Walsh CM, Swinger N, et al. Gamification in Action: Theoretical and Practical Considerations for Medical Educators. *Acad Med*. 2018;93(7):1014–1020. doi: 10.1097/ACM.0000000000002183

FIGURE LEGENDS

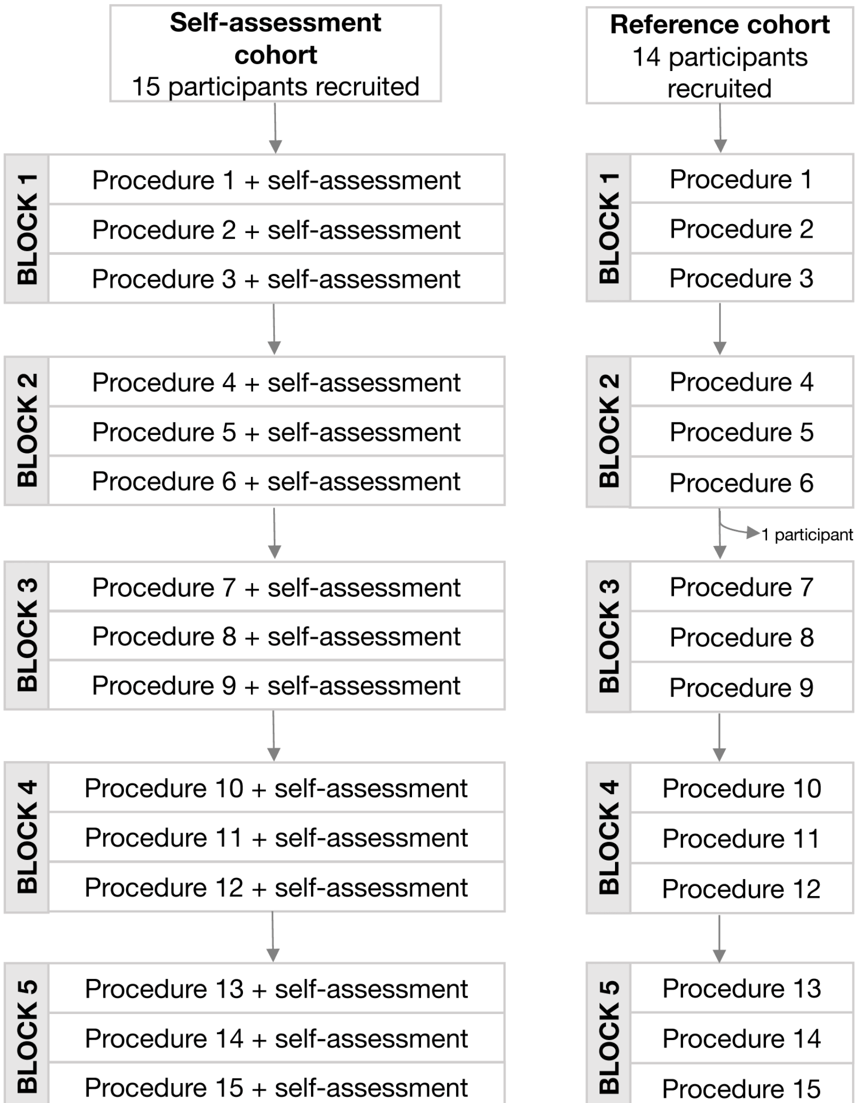

Figure 1. Study flow chart.

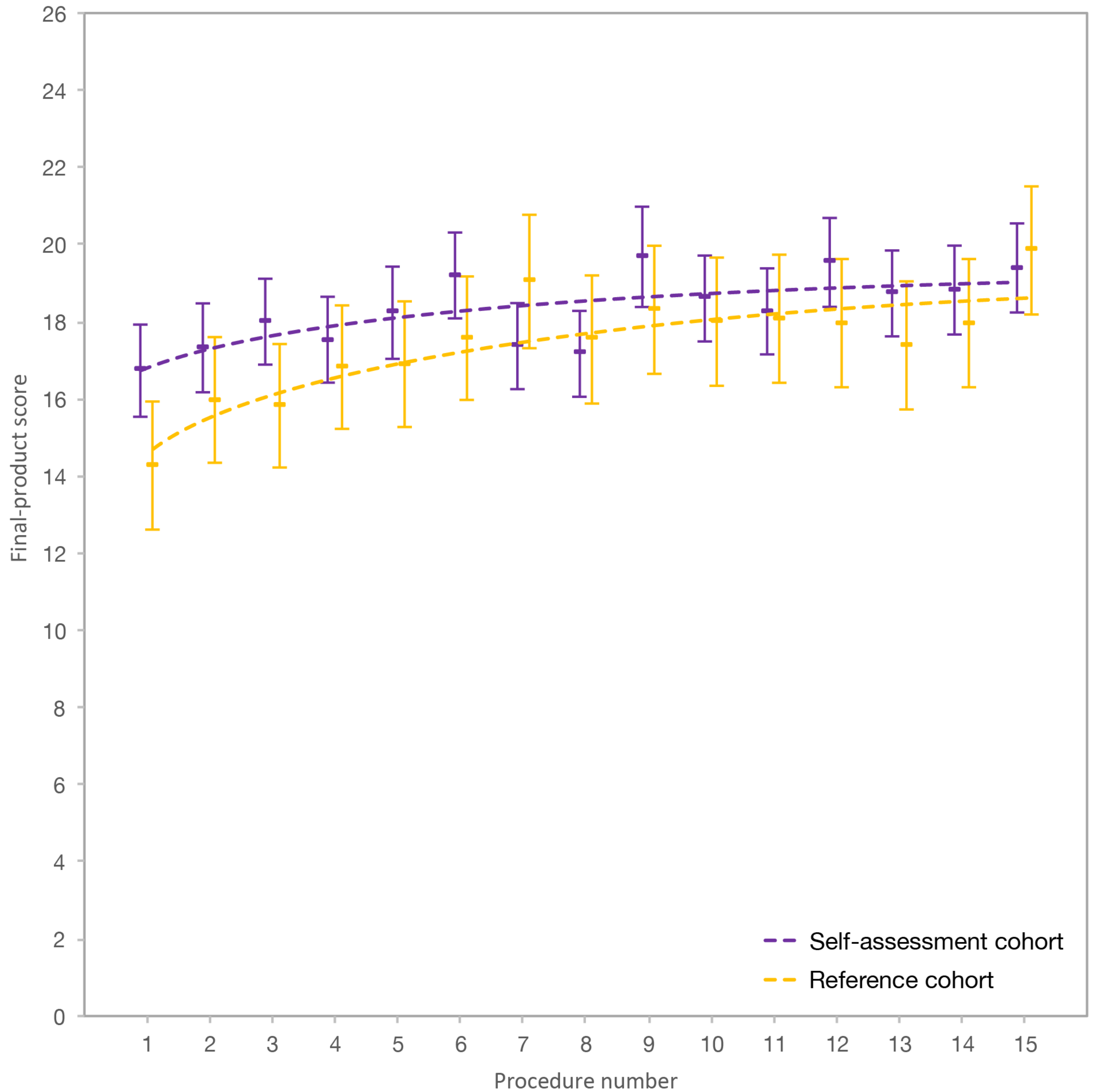

Figure 2. Means plot with learning curves for final-product performance (bars indicate 95 % confidence interval) for the self-assessment and reference cohorts.

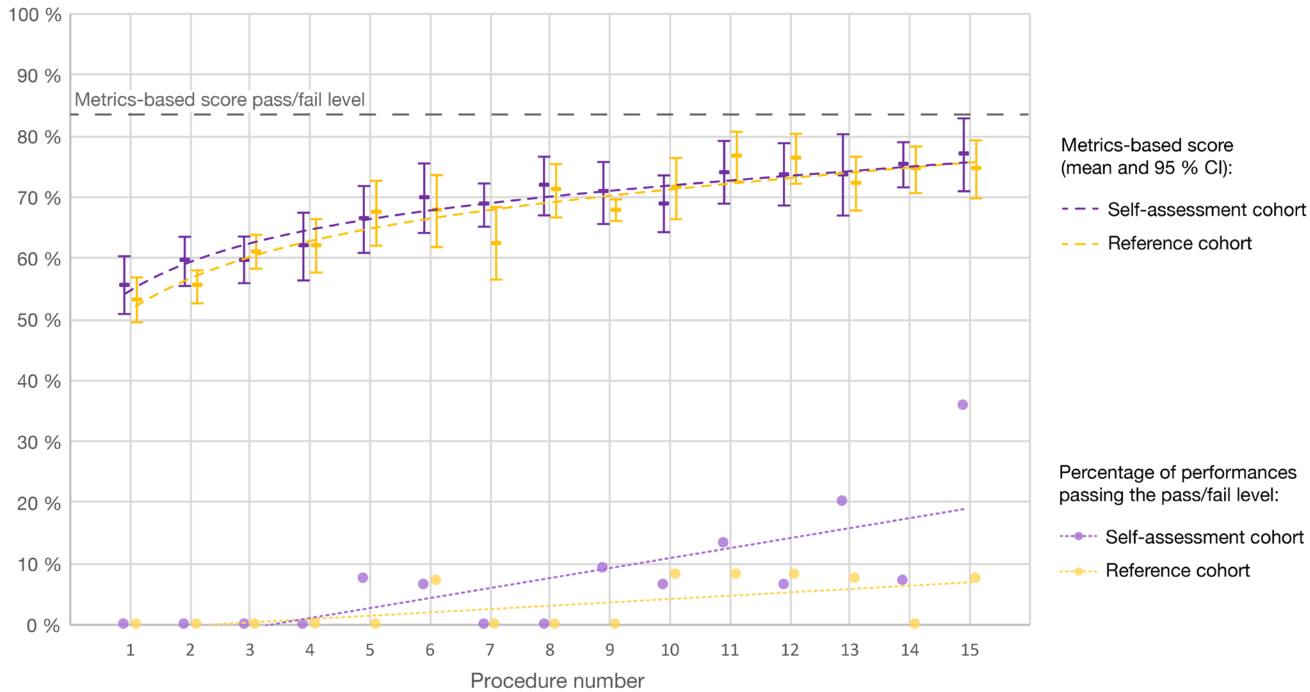

Figure 3. Means plot of learning curves for the metrics-based score (MBS) and percentage of performances passing the pass/fail level for the self-assessment and reference cohorts.

TABLES

Table I. Participant characteristics

	Self-assessment cohort	Reference cohort
	Count (%) / Mean (SD)	Count (%) / Mean (SD)
Number of participants	15	14
Age	23.8 (3.2)	23.5 (2.8)
Sex		
Female	11 (73 %)	10 (71 %)
Male	4 (27 %)	4 (29 %)
Semesters of study	6.9 (2.6)	6.1 (3.0)
Prior experience with the Visible Ear Simulator, hours	0.0 (0.0)	0.0 (0.0)
Average computer usage, hours/week	21.3 (16.2)	22.4 (9.0)
Self-rated computer skills (1–7 Likert scale)	4.5 (1.0)	4.4 (0.9)
Gaming frequency (1–5 Likert scale)	3.1 (1.6)	3.2 (1.3)

Table 2. Overview of statistically significant main effects on individual metrics of structured self-assessment.

Positive effects	Equivocal effects	Negative effects
Larger volume removed (+4.4 %, p<0.01) Smaller percentage of time not drilling (-3.7 %, p<0.0001) and less time drilling but not in contact with bone (-15.4 %, p<0.001) Shorter total path length (-48.4 %, p<0.0001) Fewer collisions with vital anatomical structures: chorda tympani (-29.7 %, p<0.0001); facial nerve (-27.2 %, p=0.0002); incus (-41.9 %, p<0.0001); malleus (-57.1 %, p=0.004)	Higher average force on burrs (sharp burrs: +5.5 %, p=0.007); coarse diamond burrs +8.3 %, p=0.001; fine diamond burrs +9.8 %, p=0.004) Higher average force on small size burrs (0.5–2 mm, +6.5 %, p=0.003) and medium size burrs (3–4 mm, +6.5 %, p=0.0001) Smaller volume removed using small size burrs (0.5-3 mm) (-26.7 %, p=0.0005)	More drill jumps (+38.8 %, p<0.0001)