



Understanding the effects of structured self-assessment in directed, self-regulated simulation-based training of mastoidectomy: A mixed methods study

Steven Arild Wuyts Andersen ^{a, b, *}, Martin Frenthø ^{a, b}, Mads Guldager ^{a, b}, Mads Sølvsten Sørensen ^a

^a Department of Otorhinolaryngology—Head & Neck Surgery, Rigshospitalet, Blegdamsvej 9, DK-2100, Copenhagen, Denmark

^b The Simulation Centre at Rigshospitalet, Copenhagen Academy for Medical Education and Simulation (CAMES), Centre for HR, the Capital Region of Denmark. Rigshospitalet, Afsnit 5404, Blegdamsvej 9, DK-2100, Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 11 November 2019

Received in revised form

18 December 2019

Accepted 20 December 2019

Keywords:

Temporal bone surgery

Structured self-assessment

Directed

Self-regulated learning

Virtual reality surgical simulation

Technical skills training

Simulation-based training

ABSTRACT

Objective: Self-directed training represents a challenge in simulation-based training as low cognitive effort can occur when learners overrate their own level of performance. This study aims to explore the mechanisms underlying the positive effects of a structured self-assessment intervention during simulation-based training of mastoidectomy.

Methods: A prospective, educational cohort study of a novice training program consisting of directed, self-regulated learning with distributed practice (5x3 procedures) in a virtual reality temporal bone simulator. The intervention consisted of structured self-assessment after each procedure using a rating form supported by small videos. Semi-structured telephone interviews upon completion of training were conducted with 13 out of 15 participants. Interviews were analysed using directed content analysis and triangulated with quantitative data on secondary task reaction time for cognitive load estimation and participants' self-assessment scores.

Results: Six major themes were identified in the interviews: goal-directed behaviour, use of learning supports for scaffolding of the training, cognitive engagement, motivation from self-assessment, self-assessment bias, and feedback on self-assessment (validation). Participants seemed to self-regulate their learning by forming individual sub-goals and strategies within the overall goal of the procedure. They scaffolded their learning through the available learning supports. Finally, structured self-assessment was reported to increase the participants' cognitive engagement, which was further supported by a quantitative increase in cognitive load.

Conclusions: Structured self-assessment in simulation-based surgical training of mastoidectomy seems to promote cognitive engagement and motivation in the learning task and to facilitate self-regulated learning.

© 2019 PLA General Hospital Department of Otolaryngology Head and Neck Surgery. Production and hosting by Elsevier (Singapore) Pte Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Simulation-based training is a well-established learning modality for surgical technical skills acquisition including in temporal bone surgery before further refinement of skills using for example

animal models or training on human cadavers and ultimately, supervised surgery on patients. Often, simulation-based training occurs outside of the clinical learning environment with limited access to surgical tutors for feedback. This allows for training at the convenience of the learner and better supports the repeated and distributed practice beneficial for skills development (Shea et al., 2000). Such self-directed learning can be a necessary condition of simulation-based training of surgical technical skills but can—if not carefully designed—lead to inefficient learning and suboptimal skills.

Self-directed training is a major challenge in temporal bone surgical training as an unsafe approach can be learned and the learning curve can plateau at an unsatisfactory level. Ceased

Abbreviations: CL, Cognitive Load; DSRL, Directed, Self-Regulated Learning; RT, Reaction Time; VR, Virtual Reality.

* Corresponding author. Department of Otorhinolaryngology—Head & Neck Surgery, Rigshospitalet, Blegdamsvej 9, DK-2100, Copenhagen, Denmark.

E-mail address: stevenarild@gmail.com (S.A.W. Andersen).

Peer review under responsibility of PLA General Hospital Department of Otolaryngology Head and Neck Surgery.

<https://doi.org/10.1016/j.joto.2019.12.003>

1672-2930/© 2019 PLA General Hospital Department of Otolaryngology Head and Neck Surgery. Production and hosting by Elsevier (Singapore) Pte Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

cognitive effort (Jowett et al., 2007) and inadequate self-assessment skills (Andersen et al., 2017) during training are among the possible explanations for such learning curve plateaus. In accordance with the theory of deliberate practice, increasing performance and refining skills beyond the learning curve plateau (development of expertise) requires an active and sustained investment of cognitive resources (Ericsson, 2004). According to cognitive load (CL) theory, cognitive resources must be invested in the learning process. CL is a balance between the germane load component used for actual learning, the intrinsic load relating to the task itself, and the extraneous load imposed by for example instructional circumstances (van Merriënboer and Sweller, 2010). Because working memory is limited, a cognitive overload can occur when the combined load of the components exceed the capacity of the learner. In contrast to novices, experts within a domain have developed the necessary psychomotor skills and mental schema for efficient cognitive processing, resulting in less cognitive effort and the ability to handle more complex tasks (Sweller et al., 2011). Cognitive load of a learning task can be estimated in different ways (Naismith and Cavalcanti, 2015) and bring insights into the investment of cognitive resources during simulation-based training.

Other features of efficient self-directed training in simulation-based training of temporal bone surgery must be considered: in general, explicit and specific process goals embedded within supportive and directive instructions (Brydges et al., 2009) are key to scaffolding simulation-based learning experiences. This concept of directed, self-regulated learning (DSRL) has been demonstrated to be superior to self-guided approaches (Brydges et al., 2009) and instructor-regulated learning (Brydges et al., 2012). However, for learners both to self-regulate their learning and to exercise deliberate practice in a self-directed context, accurate self-assessment of skills is pivotal because external feedback is not available for guidance and validation.

This argues the case for addressing both cognitive effort and self-assessment skills in temporal bone surgical training. We recently reported on the effects of structured self-assessment during simulation-based training of mastoidectomy (Andersen et al., 2019): the intervention consisted of structured self-assessment performed after each mastoidectomy procedure in distributed practice using VR simulation. The structured self-assessment intervention was centered on an 8-item rating form for self-assessment of mastoidectomy performance, based on a previously established assessment tool for final-product analysis of mastoidectomy (Andersen et al., 2015). The self-assessment rating form was supported by small videos. Medical students were used in the study as to represent true novices. We found that the intervention modified the learning curve and reduced the number of mastoidectomies until the plateau phase was reached, increased final-product performance, ensured that all participants drilled the minimum volume that experts would, and induced a safer performance with significantly fewer collisions with critical structures within the temporal bone compared with a reference cohort (Andersen et al., 2019).

The previous study only demonstrated that the intervention had a significant and positive effect on mastoidectomy performance with no potential explanations for the reasons behind this. Therefore, we designed the present follow-up study to identify the underlying mechanisms and understand the effects of the structured self-assessment in temporal bone surgical training by individual post-training interviews with the participants. We triangulated these qualitative interview data with quantitative data on cognitive load during training to provide insights into actual cognitive engagement.

2. Material and methods

2.1. Setting and participants

Participants were 15 medical students recruited from the Faculty of Health and Medical Sciences, the University of Copenhagen, Denmark, from both pre-clinical and clinical semesters in March–June 2017. They were novices in relation to the studied procedure since they had no previous exposure to temporal bone surgery or temporal bone simulation. Participation was voluntary and considered an extracurricular activity. No study credit or any other compensation was provided.

2.2. Study design

This qualitative study was part of a prospective study on the effect on performance and simulator metrics of an educational intervention—structured self-assessment—in repeated, distributed, VR simulation training of temporal bone surgery as detailed previously (Andersen et al., 2019). Participants completed five distributed training blocks of three identical mastoidectomy procedures (Fig. 1, Flow chart) in the Visible Ear Simulator (Sørensen et al., 2009)—a freeware temporal bone surgical simulator (available from <http://ves.alexandra.dk>). Participants spent a median of 32.2 min per procedure (range 16.9–93.9 min). A reference cohort of 14 medical students had previously completed an identical training program but had not received the structured self-assessment intervention and could therefore contribute quantitative data for comparison. None of the participants received feedback or information on the simulator-based metrics and also, the intervention cohort did not have access to their own self-

| TRAINING BLOCK 1–5 | | |
|-----------------------|------------------|-----------------------------|
| Baseline series of RT | | |
| PROCEDURE 1 | t=5 min | Procedural series of RT |
| | t=15 min | Procedural series of RT |
| | End-of-procedure | Structured self-assessment* |
| PROCEDURE 2 | t=5 min | Procedural series of RT |
| | t=15 min | Procedural series of RT |
| | End-of-procedure | Structured self-assessment* |
| PROCEDURE 3 | t=5 min | Procedural series of RT |
| | t=15 min | Procedural series of RT |
| | End-of-procedure | Structured self-assessment* |
| Baseline series of RT | | |

*only the self-assessment cohort

Fig. 1. Study flow chart with details on measurements. The training program consisted of five blocks each of three identical procedures. RT: reaction time.

assessment forms from previous procedures.

Participants in the self-assessment cohort were first introduced to the structured self-assessment tool using a 6-min video demonstrating a good and a bad performance in relation to key areas of the mastoidectomy and the 8 items of the self-assessment rating form (Andersen et al., 2019). Each item was rated using a 5-point Likert scale with a numerical score (1–5) as well as descriptive anchors for the two extremes and the middle score. Participants filled in the self-assessment form immediately upon completing each procedure, and to further support the self-assessment, all participants had during the self-rating the option to access: 1) the complete 6-min video used for the introduction, 2) a briefer 1-min summary video, and 3) individual video clips for each of the 8 items. A visual map (printed poster) highlighting the anatomical area of the procedure that was related to the items was placed next to the simulator.

2.3. Semi-structured interviews and analysis

Within 14 days of concluding the training program, all participants in the structured self-assessment cohort were invited by e-mail to an individual follow-up telephone interview. The interviews were semi-structured with sub-themes relating to learning strategies in DSRL and self-assessment both in general and in relation to the use of the specific structured self-assessment intervention (rating form and the supporting videos). The interview guide (Supplemental content, Appendix 1) was reviewed by an experienced educational consultant. The interviews were recorded after verbal agreement by the participant. The interviewer had no knowledge regarding the participants or their performance and the participants did not know the interviewer but were aware of the overall purpose of the interview through the invitation e-mail. The interviewer (SA) is male, holds an MD and a PhD degree, is experienced with medical educational research, and at the time of the study, worked in a combined clinical/postdoctoral research position.

The interviews were transcribed verbatim by a medical secretary and audited by the first author (SA). Participants did not review the transcription of their interviews nor the analysis. For qualitative analysis, the transcripts were structured in a spreadsheet (Microsoft Excel, Microsoft Corporation, Redmond, WA, USA) with each question/response placed in a separate row, with adjacent columns used for interpretation of content and identification of themes. The interviews were analysed using directed content analysis (Hsieh and Shannon, 2005). Two coders (SA and MF) independently coded the interviews and achieved agreement on themes and sub-themes by serial discussion. Adequate saturation was obtained with the number of interviews conducted. Select quotes for the analysis were finally translated by one coder (SA) and verified by the other coder (MF). The qualitative reporting was based on the Consolidated Criteria for Reporting Qualitative Research (COREQ) (Tong et al., 2007).

2.4. Quantitative data

For triangulation with the qualitative data, the mean of the combined self-assessment score (minimum score 8 points, maximum score 40 points) was calculated for each procedure number (with 95% confidence intervals). In addition, cognitive load during each procedure was estimated by the relative increase (to individual baseline) in secondary task reaction time (in response to a sound stimulus) measured (by MG) in repeated series of four (measurement number) at $t = 5$ min and $t = 15$ min using a commercially available external reaction timing device (American Educational Products LLC, Fort Collins, CO, USA), similar to

previously (Andersen et al., 2018). Estimated marginal means were calculated using linear mixed models due to repeated measurements with time ($t = 5/t = 15$ min), measurement number (1–4), and procedure number (1–15), as fixed factors. The final model included procedure number, cohort, procedure number*cohort interaction, and measurement number, as time of measurement was not found to be significant. SPSS version 23 for MAC OSX (IBM, Armonk, NY, USA) was used for the quantitative analyses.

2.5. Ethics

The regional ethics committee for the Capital Region of Denmark deemed this study to be exempt (H-17002805). All participants signed informed consent.

3. Results

The 15 participants had a mean age of 23.8 years (SD 3.2), 11 were women and 4 men, and they had on average studied 6.9 semesters (SD 2.6) in the 12-semester medical program at the University of Copenhagen, Denmark. The intervention and reference cohorts had similar background data (Andersen et al., 2019). Thirteen of the 15 participants who had completed training accepted the invitation for a follow-up telephone interview, the remaining two did for reasons unknown not respond to the invitation for the follow-up interview. The average length of the interviews was 16.2 min (range 10–26 min). All 15 participants contributed to the quantitative data.

3.1. Qualitative data

Six main themes were identified in the interviews (Table 1): A) Goal-directed behaviour, B) Use of learning supports for scaffolding, C) Cognitive engagement, D) Motivation from self-assessment, E) Self-assessment bias, and F) Feedback on self-assessment.

3.1.1. Goal-directed behaviour

Many participants reported that the self-assessment tool brought their attention to specific elements of the procedure: “[...] then you knew the criteria for success. Or what the purpose was, what you needed to get from it [...] when it is considered satisfactory, and when it isn’t” (A, 4th semester student). The structured self-assessment tool seemed to provide directions for the participants on what constituted a good performance and how to achieve excellency: “naturally it gives a focus on performing the specific parts, that is to do it as good as possible, because you know you will then achieve a high score” (B, 9th semester student). Not all participants had a focus on the final product or a “score” (as they were not provided with such a score): for some participants, the self-assessment tool rather seemed to provide a guide that could facilitate a mental scheme of the important sub-goals of the procedure. This they could use to define their own way to the goal and strategies: “unconsciously, I have divided the procedure into smaller areas, that I had to work through. But I haven’t just done it in the order of the steps” (C, 8th semester student).

3.1.2. Use of learning supports for scaffolding

The supporting videos were not used much during training. In fact, while participants reported to have been introduced to the structured self-assessment using the videos and many that they were reminded (by MG) of their continued availability throughout the study, only few reported to have actually watched any of the videos again after the first training block: “I saw them the very first time [...] but I only used them that first time, to find out what to do” (D, 9th semester student). Only two participants reported watching

Table 1
Examples from qualitative analysis of the interviews.

| Exemplary quote | Paraphrase | Major theme (minor theme) |
|---|--|--|
| "Sometimes it made more sense to take a different approach, because I in some way had done something differently, or how you should describe it. And then I devised the strategy" (C, 8th semester student) | The participant chose a learning strategy to match their own approach. | Goal-directed behaviour (directed, self-regulated learning) |
| "This next time, I will aim at getting the mastoid tip clear or the digastric ridge or similar. In that way, I definitely think it helped me define some sub-goals" (H, 5th semester student) | The participant defined specific areas to improve for the next procedure, and the structured self-assessment tool helped them make such sub-goals. | Goal-directed behavior (defining sub-goals) |
| "He offered it [to watch the videos] every time. He was really good at asking if I wanted to see them. But I thought, I had it under control. I felt that. I didn't feel it was necessary to" (G, 10th semester student) | The participant was offered to review the supporting videos throughout the study, but did not feel it necessary to review them again. | Learning supports and scaffolding (use of videos) |
| "We used them [the videos] the first time. But it was actually only the first time we did it, that we saw them. And also, I think, I saw the one on 'digastric ridge' once at a later point." (A, 4th semester student) | The participant only watched the videos the first time, except for the one specifically on the 'digastric ridge' item. | Learning supports and scaffolding (use of videos) |
| "Especially after the 11th procedure where you have to do it yet another time. You've done it so many times, and then you just [pauses] if you hadn't self-assessed, then I think I would just have stopped and been a little more ... sluggish ... than if not [...] so now that I had to self-assess, you become more ... you want to perform even better." (J, 4th semester student) | With repeated practice the participant would usually have stopped caring about their performance but because of the self-assessment the participant wanted to cognitively invest in continuing to have a good performance. | Cognitive engagement |
| "If you can see that your [self-assessment] score increases each time, then you can see that you are improving, because you have some data that you can relate to" (D, 9th semester student) | The self-assessment score helps the participant realize that their performance is improving because it provides data supporting this. | Motivation from self-assessment (sense-of-improving) |
| "And also, when you get better, then I at least, did not assign much better scores, because I still [pauses] then the goal for what is good just gets raised further." (K, 10th semester student) | As the participant gets more experience, what they consider a good performance changes and the participant does not assign a higher score. | Self-assessment bias (goal-post moving) |
| "If you are a little self-critical, then you easily rate yourself too low [...] and this also results in you getting even more critical as you get better" (E, 5th semester student) | Self-criticism results in the participant scoring their performance lower than they perceived it maybe should have been and with experience, the participant gets more critical about their own performance. | Self-assessment bias (self-criticism) |
| "If you from the first time practice self-directed without knowing how something is performed correctly, then you might do it the wrong way [...] and then you actually get better at doing something the wrong way. Then you need some expert to show how it is done correctly in the first place" (D, 9th semester) | Self-directed practice can lead to the participant learning something incorrectly and reinforcing incorrect behaviour. An expert is needed to demonstrate the correct way before self-directed practice. | Feedback on self-assessment (the need for external validation) |

any of the short videos—the one concerning what they found to be a particularly difficult item, "the digastric ridge": "*The one on 'the digastric ridge', because I was uncertain, exactly what the goal was there. But I did it at a relatively late point, during the fourth drilling, I think*" (E, 5th semester student).

Participants also seemed to self-regulate and scaffolded their learning using the other available learning supports. Several used the visual map that was placed as a poster next to the simulator: "*the picture that was hanging. Definitely. Because there you could see how it's supposed to look [...] I used that a lot as guidance*" (C, 8th semester student). The built-in on-screen guide to the procedures were mostly used for the first few procedures and otherwise used for reference at the end of the procedure or to guide very specific aspects: "*I made more use of the written instructions [...] I often used the types of burrs you could see in the pictures*" (F, 11th semester student). The same pattern was reported for the simulator's transparency function, which enables the trainee to look through the bone and see the critical structures underneath.

3.1.3. Cognitive engagement

The main objective of implementing structured self-assessment was to encourage continued cognitive engagement with the learning task, and indeed several participants reported that they were more careful and meticulous throughout the procedure because they knew that they had to rate themselves afterwards and also with the knowledge that the items for self-assessments were central to a satisfactory procedure: "*you were more careful the next time. Because knowing that you gave yourself a score of 1 [...] the next time you were extra attentive to not repeating it. If I just repeated the procedure again [without self-assessment], I think I would have done more or less the same, making the same error again and again.*"

Whereas you pay more attention when you yourself have assessed whether it was satisfactory or unsatisfactory" (J, 4th semester student).

Another common statement was that the structured self-assessment helped keep focus on the learning task and the important parts of the procedure: "*The fact that you self-assess, it keeps you focused on the important. And it increases your concentration.*" (K, 10th semester student).

3.1.4. Motivation from structured self-assessment

The participants reported a variety of reasons for signing up for the study, however, all but one participant reported some sort of intrinsic motivation such as interest in the field of otorhinolaryngology, most likely attributable to a recruitment bias (volunteering for a lengthy study). The one participant who reported an external motivation, explained that they participated because their study buddy wanted to. Some participants reported that it was difficult to keep up motivation during the identical, repeated procedures over a prolonged period of time. Others, however, found that the structured self-assessment helped them maintain their engagement. They explained that the self-assessment highlighted their improvement over time, and this sense of improvement was found to be motivational and kept them engaged: "*I felt that I improved [pauses] and I noticed this improvement more because I had to evaluate my own performance at the same time*" (K, 10th semester student) and "*it helped me keep my spirit high during the drillings*" (F, 11th semester student).

3.1.5. Self-assessment bias

Most participants found self-assessment to be a natural part of their learning process in general even though it might not always

be explicitly articulated during their studies. Even though this seemed to be a strong feature of their professional identity, most reported that they found structured self-assessment to be difficult. With this specific structured rating form, they reported that it was easy to discriminate between the extreme scores (1 and 5) but they had more difficulty choosing between the middle scores for various reasons that they thought introduced a rating bias: consistently, they would rarely assign their own performance the maximum score because they thought there always was some room for improvement. Self-criticism was also a common theme: “I can actually see that this should be scored at a 4, but I still think it looks like crap” (C, 8th semester student) often accompanied by perfectionism: “it really has to be damn perfect before you give yourself a maximum score” (G, 10th semester student). Participants reported getting more and more critical about their own performance the more experience they got, but at the same time, since they also had an expectation of getting better, this resulted in goal-post moving: “I get more critical now that I have been here 9 times, now it should be better, so now it takes more before it is a 3 or 4” (H, 5th semester student) and a sort of self-fulfilment: “because you expect yourself to improve over time ... so because you have expectations of improving, you also rate your performance higher, even though it wasn't necessarily” (A, 4th semester student).

3.1.6. Feedback on self-assessment

A major challenge for the participants seemed to be finding a source of reference for self-assessment of performance because there were no absolute measures or standards. They had to extract this from the videos demonstrating a good and a bad performance for each item, the pictures in the on-screen step-by-step guide, and the poster, as they had no prior knowledge on the procedure theoretically or from observing it in real-life during clinics or received any feedback on their performance. Primarily, they felt that there was a lot of room for interpretation in the self-assessment and the lack of a strict point of reference seemed to frustrate some participants: “I had trouble assessing what constitutes a small hole and a large hole, and how many [...] I think it's difficult to

grade, because there was not a concrete number. Had there for example been something like, ‘have you made holes corresponding to one third of the area’, some sort of number, you could rate.” (I, 4th semester).

Many participants would have liked external validation against expert opinion and feedback, not necessarily continuously as they also wanted to develop their skills on their own, but initially and again at some later point during the simulation, to assess whether they were on track and in what areas to improve: “you can kind of miss having someone [saying:] this bit is really important, if it is going to be a good performance” (F, 11th semester student).

3.2. Quantitative data

3.2.1. Relative reaction time for CL estimation

Overall, the relative reaction time was 6% higher in the self-assessment cohort than the reference cohort ($p = 0.001$, linear mixed models) even though the two groups had similar relative reaction times during the first procedure (Fig. 2). This could indicate that more cognitive resources were invested in learning in the structured self-assessment cohort. However, the relationship between repeated practice and relative reaction time was not linear but followed a negatively accelerated curve with the difference between the two cohorts being more marked towards the final procedure (Fig. 2). The linear mixed model confirmed a significant interaction between procedure number and cohort ($p < 0.0001$). In other words, the relative reaction time decreased by different rates in the two cohorts (see Fig. 2).

3.2.2. Self-assessment scores

The participants' self-assessment scores increased with repeated practice from 19.2 (95% CI [16.3–22.2]) after the first procedure to 26.2 (95% CI [23.3–29.2]) after the 15th procedure (linear mixed models, $p < 0.002$). The mean self-assessment scores followed a negatively accelerated curve (Fig. 3).

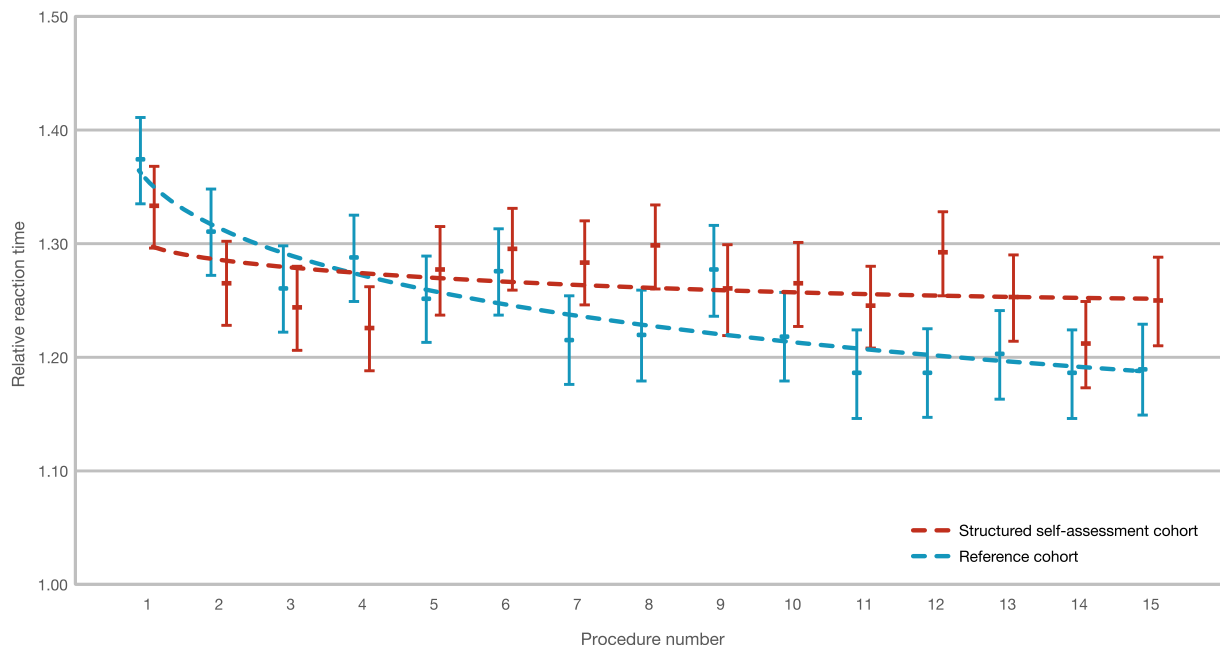


Fig. 2. Means plot of the change in cognitive load as estimated by relative increase in secondary task reaction time compared with baseline. Error bars mark 95% confidence intervals.

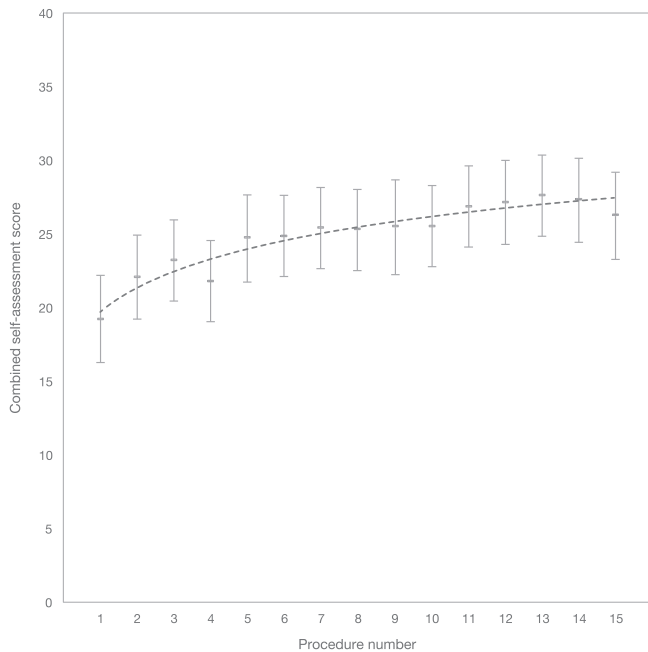


Fig. 3. Mean self-assessment scores. Error bars mark 95% confidence intervals.

4. Discussion

In this study, we explored learner's perception of structured self-assessment in simulation-based training of mastoidectomy through semi-structured telephone interviews to identify and understand the mechanisms of the positive effect on novice performance. We triangulated the qualitative interview data with participants' self-assessment scores and cognitive load estimated by secondary task reaction time measurements during simulation training.

4.1. Synthesis of the major findings

Six major themes relating to structured self-assessment were found. These mainly reflected different aspects of the structured self-assessment and learning processes in a DSRL context. First of all, structured self-assessment was reported to facilitate the formation of specific sub-goals in relation to the mastoidectomy procedure, increasing the learners' cognitive engagement in relation to these sub-goals, and keeping them motivated to continuously refine their performance despite the task being repetitive. This was corroborated by estimation of CL with the reaction time data: the intervention seemed to modify the slope of the CL curves of repeated practice and—despite some fluctuation—overall to induce a higher CL during training compared with the reference cohort. This can potentially be explained by the induction of an increased germane load for the formation of mental schema (van Merriënboer and Sweller, 2009). This would be in agreement with what the participants reported in the qualitative interviews: that structured self-assessment encouraged their cognitive engagement and increased motivation. It seems unlikely that the difference in CL can be explained by the intrinsic load component, as the task was kept identical, nor by the extrinsic load component, as the structured self-assessment was performed after each procedure and the remaining circumstances surrounding the simulation identical to what the reference cohort had experienced. Nonetheless, attributing the effects on CL to an increase in germane load remains speculation as only the total CL was estimated by the secondary

task approach to CL estimation.

The interviews also confirmed that participants self-regulated their learning experience in a directed fashion: they seemed not only to define their own way and strategies to obtain the designated goal (i.e. completion of the mastoidectomy procedure) but to scaffold their learning and use the different available tools such as the supporting videos, the visual map poster, the on-screen step-by-step guide built into the simulation software, and the bone transparency function, as it matched their specific needs and individual preferences. Such behaviour is desirable in the context of DSRL, where the instructional design and learning supports should facilitate a learning experience where the learner self-regulates, chooses individual learning strategies, and defines own goals and sub-goals within the overall aim. A possible explanation for the positive effects of DSRL (Brydges et al., 2009) could be that it promotes deep learning processes and strategies. The structured self-assessment seemed to result in favorable learning behaviors for DSRL, but the relative contributions to this compared with the standard simulation-based training of mastoidectomy received by the reference cohort remains unknown.

Finally, despite their apprehension of self-assessment and the different sources of rating bias mentioned in the interviews such as never using the maximum score, self-criticism and perfectionism, expectation of improvement, and goal-post moving, the mean self-assessment score followed a traditional, negatively accelerated learning curve. However, the learning curve seemingly plateaus at a low score and there is likely a ceiling effect (Munz et al., 2004) attributable to the rating bias. That accuracy of self-assessment is limited is well-established in the literature (Davis et al., 2006), and its value as an learning outcome in itself is minimal (Ewa and Regehr, 2005).

4.2. Strengths and limitations

The main limitation of the present study is the generalizability: we have studied structured self-assessment in the context of the mastoidectomy procedure that requires complex psychomotor skills with little clinical relevance to participants and of no consequence to their future training. In contrast, the intrinsic motivation for other surgical tasks and technical skills and for performing self-assessment could be different. The qualitative analysis, however, adds new knowledge and understanding of the mechanisms of structured self-assessment during simulation-based training of surgical technical skills. The supplemental quantitative data for triangulation with the qualitative data were also found to corroborate statements from the interviews, supporting the conclusions using different sources of data. A two week period between finishing the distributed training program and the telephone interviews was planned to allow the participants time to reflect on their learning experience. However, this inherently has the potential to introduce a recall bias.

4.3. Comparison with other studies

In contrast to other domains such as professional competency, few studies have investigated the use of self-assessment (mainly of watching videos) as a means to improving surgical technical skills performance (Jamshidi et al., 2009; Ganni et al., 2018; and Jethwa et al., 2018), and none in the context of extended, distributed and repeated practice or temporal bone surgical training. Eva and Regehr suggest that there is a close relationship between self-assessment, self-monitoring, and reflective activities (Eva and Regehr, 2005). Even though the body of literature on reflective practice is substantial, the role of reflection in enhancing learning remains under-investigated (Mann et al., 2007). When considering

structured self-assessment as a way to integrate reflection-in-practice, our current study contributes both quantitative and qualitative insights into the effects of reflection activities in enhancing learning in simulation-based training of temporal bone surgery.

4.4. Future directions

The interviews suggest directions for future improvements of structured self-assessment in general: additional framing and anchoring could potentially reduce relativism, self-criticism, and goal-post moving; furthermore, external validation, computer-generated metrics or other sources of feedback could potentially be valuable in improving the accuracy of self-assessment of mastoidectomy performance and further refine the formation of sub-goals and mental-schema in relation to the procedure. The structured self-assessment tool could also be used to support learning on other training models such as plaster/plastic temporal bones as well as cadaveric temporal bones. However, structured self-assessment cannot stand alone and needs to be integrated with other learning supports in the training curriculum and should not be used for decisions on progression to supervised surgery in the clinic.

5. Conclusions

Structured self-assessment in simulation-based training of temporal bone surgery seemed to promote cognitive engagement and motivation in the learning task and to facilitate self-regulated learning including the formation of procedural sub-goals, use of individualized learning strategies, and scaffolding using the range of learning supports. Triangulation of the qualitative interview data with quantitative data estimating CL during the mastoidectomy simulations demonstrated that the structured self-assessment intervention increased the CL and modified the CL learning curve compared to a reference cohort. Finally, participants reported a number of biases in their self-assessment such as rarely using the maximum score for each item, self-criticism and perfectionism, expectation of improvement, and goal-post moving. This could indeed lead to the ceiling effect found in the learning curve of the average combined self-assessment score. Participants desired external validation of their mastoidectomy performance in order to confirm the accuracy of their self-assessment. Altogether, structured self-assessment is a valuable support in simulation-based training of temporal bone surgery.

Declaration of competing interest

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. The authors have no conflicts of interest to disclose.

Acknowledgements

The authors would like to acknowledge Dr. Ebbe Thinggaard for his insights into directed content analysis, educational consultant

Kirstine Dahl for reviewing the interview guide, medical secretary Lena Boye who was hired to transcribe the interviews, and computer scientist Peter Trier Mikkelsen for his important contribution of programming the Visible Ear Simulator software.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.joto.2019.12.003>.

References

- Andersen, S.A.W., Caye-Thomasen, P., Sørensen, M.S., 2015. Mastoidectomy performance assessment of virtual simulation training using final-product analysis. *The Laryngoscope* 125, 431–435.
- Andersen, S.A.W., Konge, L., Mikkelsen, P.T., Caye-Thomasen, P., Sørensen, M.S., 2017. Mapping the plateau of novices in virtual reality simulation training of mastoidectomy. *The Laryngoscope* 127, 907–914.
- Andersen, S.A.W., Konge, L., Sørensen, M.S., 2018. The effect of distributed virtual reality simulation training on cognitive load during subsequent dissection training. *Med. Teach.* 40, 684–689.
- Andersen, S.A.W., Guldager, M., Mikkelsen, P.M., Sørensen, M.S., 2019. The effect of structured self-assessment in virtual reality simulation training of mastoidectomy. *Eur. Arch. Oto-Rhino-Laryngol.* 276, 3345–3352.
- Brydges, R., Carnahan, H., Safir, O., Dubrowski, A., 2009. How effective is self-guided learning of clinical technical skills? It's all about process. *Med. Educ.* 43, 507–515.
- Brydges, R., Nair, P., Ma, I., Shanks, D., Hatala, R., 2012. Directed self-regulated learning versus instructor-regulated learning in simulation training. *Med. Educ.* 46, 648–656.
- Davis, D.A., Mazmanian, P.E., Fordis, M., Van Harrison, R., Thorpe, K.E., Perrier, L., 2006. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *J. Am. Med. Assoc.* 296, 1094–1102.
- Ericsson, K.A., 2004. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad. Med.* 79 (Suppl. ment), S70–S81.
- Eva, K.W., Regehr, G., 2005. Self-assessment in the health professions: a reformulation and research agenda. *Acad. Med.* 80 (Suppl. ment), S46–S54.
- Ganni, S., Botden, S.M.B.I., Schaap, D.P., Verhoeven, B.H., Goossens, R.H.M., Jakimowicz, J.J., 2018. "Reflection-Before-Practice" improves self-assessment and end-performance in laparoscopic surgical skills training. *J. Surg. Educ.* 75, 527–533.
- Hsieh, H.-F., Shannon, S.E., 2005. Three approaches to qualitative content analysis. *Qual. Health Res.* 15, 1277–1288.
- Jamshidi, R., LaMasters, T., Eisenberg, D., Duh, Q.Y., Curet, M., 2009. Video self-assessment augments development of videoscopic suturing skill. *J. Am. Coll. Surg.* 209, 622–625.
- Jethwa, A.R., Perdoni, C.J., Kelly, E.A., Yueh, B., Levine, S.C., Adams, M.E., 2018. Randomized controlled pilot study of video self-assessment for resident mastoidectomy training. *OTO Open* 2, 2473974X1877041-X1877045.
- Jowett, N., LeBlanc, V., Xeroulis, G., MacRae, H., Dubrowski, A., 2007. Surgical skill acquisition with self-directed practice using computer-based video training. *Am. J. Surg.* 193, 237–242.
- Mann, K., Gordon, J., MacLeod, A., 2007. Reflection and reflective practice in health professions education: a systematic review. *Adv. Health Sci. Educ.* 14, 595–621.
- van Merriënboer, J.J.G., Sweller, J., 2010. Cognitive load theory in health professional education: design principles and strategies. *Med. Educ.* 44, 85–93.
- Munz, Y., Moorthy, K., Bann, S., Shah, J., Ivanova, S., Darzi, S.A., 2004. Ceiling effect in technical skills of surgical residents. *Am. J. Surg.* 188, 294–300.
- Naismith, L.M., Cavalcanti, R.B., 2015. Validity of cognitive load measures in simulation-based training. *Acad. Med.* 90, S24–S35.
- Shea, C.H., Lai, Q., Black, C., Park, J.-H., 2000. Spacing practice sessions across days benefits the learning of motor skills. *Hum. Mov. Sci.* 19, 737–760.
- Sweller, J., Ayres, P., Kalyuga, S., 2011. *Cognitive Load Theory*. Springer, New York, 2011.
- Sørensen, M.S., Mosegaard, J., Trier, P., 2009. 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.* 30, 484–487.
- Tong, A., Sainsbury, P., Craig, J., 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int. J. Qual. Health Care* 19, 349–357.