Retention of Mastoidectomy Skills After Virtual Reality Simulation Training

Steven Arild Wuyts Andersen, MD; Lars Konge, MD, PhD; Per Cayé-Thomasen, MD, DMSc; Mads Sølvsten Sørensen, MD, DMSc

**Importance** The ultimate goal of surgical training is consolidated skills with a consistently high performance. However, surgical skills are heterogeneously retained and depend on a variety of factors, including the task, cognitive demands, and organization of practice. Virtual reality (VR) simulation is increasingly being used in surgical skills training, including temporal bone surgery, but there is a gap in knowledge on the retention of mastoidectomy skills after VR simulation training.

**Objectives** To determine the retention of mastoidectomy skills after VR simulation training with distributed and massed practice and to investigate participants’ cognitive load during retention procedures.

**Design, Setting, and Participants** A prospective 3-month follow-up study of a VR simulation trial was conducted from February 6 to September 19, 2014, at an academic teaching hospital among 36 medical students: 19 from a cohort trained with distributed practice and 17 from a cohort trained with massed practice.

**Interventions** Participants performed 2 virtual mastoidectomies in a VR simulator a mean of 3.2 months (range, 2.4-5.0 months) after completing initial training with 12 repeated procedures. Practice blocks were spaced apart in time (distributed), or all procedures were performed in 1 day (massed).

**Main Outcomes and Measures** Performance of the virtual mastoidectomy as assessed by 2 masked senior otologists using a modified Welling scale, as well as cognitive load as estimated by reaction time to perform a secondary task.

**Results** Among 36 participants, mastoidectomy final-product skills were largely retained at 3 months (mean change in score, 0.1 points; \( P = .89 \)) regardless of practice schedule, but the group trained with massed practice took more time to complete the task. The performance of the massed practice group increased significantly from the first to the second retention procedure (mean change, 1.8 points; \( P = .001 \)), reflecting that skills were less consolidated. For both groups, increases in reaction times in the secondary task (distributed practice group: mean pretraining relative reaction time, 1.42 [95% CI, 1.37-1.47]; mean end of training relative reaction time, 1.24 [95% CI, 1.16-1.32]; and mean retention relative reaction time, 1.36 [95% CI, 1.30-1.42]; massed practice group: mean pretraining relative reaction time, 1.34 [95% CI, 1.28-1.40]; mean end of training relative reaction time, 1.31 [95% CI, 1.21-1.42]; and mean retention relative reaction time, 1.39 [95% CI, 1.31-1.46]) indicated that cognitive load during the virtual procedures had returned to the pretraining level.

**Conclusions and Relevance** Mastoidectomy skills acquired under time-distributed practice conditions were retained better than skills acquired under massed practice conditions. Complex psychomotor skills should be regularly reinforced to consolidate both motor and cognitive aspects. Virtual reality simulation training provides the opportunity for such repeated training and should be integrated into training curricula.
Surgical training is undergoing a paradigm shift from traditional apprenticeship to increased use of simulation-based training. Patient safety issues, constraints on working hours, and productivity demands contribute to limited training opportunities under the traditional apprenticeship. Still, safe performance of high-risk surgery requires extensive and high-quality training, and the complex psychomotor skills needed to perform surgery must be developed both efficiently and reliably. Virtual reality (VR) simulation-based surgical skills training has been demonstrated in a range of different surgical fields to improve performance and transfer newfound skills to the operating room.

In temporal bone surgery, VR simulation is primarily used to supplement other training modalities, such as cadaveric dissection, and current evidence supports the effectiveness of VR simulation in training of novices to perform mastoidectomy. However, performance during practice is often the only reported outcome in these studies. Nevertheless, measurement of the retention of acquired skills is a better indicator of actual learning than is performance during practice because consolidated skills and consistency of performance are the goals of surgical training. In other words, retention tests “attempt to remove the effects of temporary modulators on performance such as fatigue, and rely only on the retrieval of skills from memory.”

To some extent, complex psychomotor skills acquired in a VR simulation environment seem to be retained for several months. However, surgical skills are retained heterogeneously and depend on the procedure, the task studied, and time elapsed since training. In addition, several other factors affect the retention and transfer of skills training, including deliberate practice, training in a subset of skills relating to the overall task, task variability, and overlearning after reaching proficiency.

There is also evidence that heavier demands on cognitive functions during acquisition of motor skills negatively affect retention. Highly complex motor skills could cause substantial cognitive load owing to the limitations of working memory and thereby inhibit the capacity for learning. Several instructional designs can modify the cognitive load; studies have previously demonstrated that organizing training as distributed practice (practice sessions spaced in time) rather than massed practice (all sessions in 1 day) provides superior learning curves and reduces cognitive load. However, there is a gap in knowledge on whether such an improvement in performance and reduction in cognitive load during performance of the procedure are sustained after the training period.

Based on this finding, we hypothesized that different training strategies affect the retention of surgical motor skills and cognitive load during retention performance. The aims of this study were to determine the retention of mastoidectomy skills after VR simulation training with distributed and massed practice and to investigate the cognitive load in the retention procedures, with the purpose of informing the optimal organization of temporal bone skills training.
For this study, participants who had completed training in the previous study were invited back for retention testing after 3 months. Thirty-six participants accepted the invitation for this follow-up study: 19 of 21 participants in the distributed practice cohort and 17 of 19 participants in the massed practice cohort completed the retention procedures. None of the participants had practiced the procedure in the intervening period. The follow-up retention testing was scheduled at the convenience of the participant and consisted of 2 mastoidectomy procedures identical to the 30-minute procedures in the initial study. During retention testing, participants had access to the standard on-screen instructions and received no other assistance.

**Outcomes**
The virtual mastoidectomy was automatically saved by the simulator every 10 minutes, and performances were later assessed by 2 masked expert raters (P.C.-T. and M.S.S.) using final-product analysis with a modified Welling Scale. In addition, participants were tested on their reaction time while performing a secondary task provided by the simulator at baseline and several times during the procedure to estimate the cognitive load by the increase in reaction time during simulation relative to baseline measurements. The outcomes (final-product performance and relative reaction time for cognitive load estimation) were analyzed as previously described to ensure comparability with previous studies. Supplemental analyses of the volume removed during VR simulation sessions were performed for this study.

**Statistical Analysis**
The mean scores and mean reaction times of the 2 retention procedures (sessions 13 and 14) and the last 2 procedures (end-of-training procedures; sessions 11 and 12) of the initial study were compared. Data were analyzed using SPSS (SPSS, Inc), version 22 for MacOS X with analysis of variances, paired samples 2-tailed t tests, and Pearson r for correlations.

**Results**
Two participants in the distributed practice cohort (10%) and 2 participants in the massed practice cohort (11%) were unavailable for follow-up. Participant characteristics were therefore similar to those reported in the initial study: individuals in the distributed practice group were significantly older, more often male, and had a higher frequency of playing video games than participants in the massed practice group (Table 1). As in the initial study, these factors could not be demonstrated to be associated with the outcomes.

The mean number of days between the end-of-training sessions in the initial study and the retention sessions in this follow-up study were comparable for the 2 practice groups (Table 1). In addition, the number of the days until follow-up was not associated with final-product performance or relative reaction time performance.

For both practice groups, the difference in mean final-product performances of the end-of-training sessions and the retention sessions were not statistically significant (Table 2). The slightly lower performance during retention procedures was related to the anatomical boundaries of the procedure, such as adequately removing cells in the sinodural angle, along the tegmen, and in the mastoid tip; not overexposing the facial nerve; and expanding the facial recess.
We also found that the final-product performance of the massed practice group increased significantly from the first to the second retention procedure (mean change, 1.8 points; \(P = .001\)), whereas the performance of participants in the distributed practice group remained unchanged during the retention procedures (Figure 2). The 2 groups had equal mean retention final-product performances (mean change, 0.1 points; \(P = .89\)). A different pattern was observed for the relative reaction time (Figure 3): both groups had an increase in relative reaction time when comparing retention sessions with end-of-training sessions, even though this increase was only statistically significant for the distributed practice group (\(P < .01\)), and both practice groups had equal mean reaction times in the retention sessions (Table 2).

We performed a supplemental analysis of the total volume removed during the VR simulation sessions to explore whether the fixed 30-minute time frame masked differences between the groups in time needed to complete the task. This analysis demonstrated that the distributed practice group consistently removed more bone than did the massed practice group in both the end-of-training sessions and the retention sessions (eFigure 1 in the Supplement), consistent with the higher final-product performance of the distributed practice group. Also mirroring the final-product performance, a decrease in the total volume removed was found in the first retention session. Finally, the massed practice group removed significantly more bone during the last 10 minutes of the first retention session (session 13) than in their last end-of-training session (session 12) (\(P < .002\)) (eFigure 2 in the Supplement) while retaining total volume removed, reflecting time compensation.

### Table 1. Participant Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Distributed</th>
<th>Massed</th>
<th>(P) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention-tested/initially trained, No. (%)</td>
<td>19/21 (90)</td>
<td>17/19 (89)</td>
<td></td>
</tr>
<tr>
<td>Age, mean, y</td>
<td>25.2</td>
<td>23.4</td>
<td>.01</td>
</tr>
<tr>
<td>Sex, No. (%)</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Male</td>
<td>11 (58)</td>
<td>4 (24)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8 (42)</td>
<td>13 (76)</td>
<td></td>
</tr>
<tr>
<td>Semesters of study, mean, No.</td>
<td>8.8</td>
<td>7.5</td>
<td>.12</td>
</tr>
<tr>
<td>Any previous VR simulation experience, No. (%)</td>
<td>6 (32)</td>
<td>5 (29)</td>
<td>.84</td>
</tr>
<tr>
<td>Gaming frequency*</td>
<td>2.3</td>
<td>1.5</td>
<td>.02</td>
</tr>
<tr>
<td>Computer usage, h/wk</td>
<td>18</td>
<td>20</td>
<td>.64</td>
</tr>
<tr>
<td>Time between last training and retention sessions, mean, mo (range)</td>
<td>3.3 (2.4-5.0)</td>
<td>3.1 (2.6-4.1)</td>
<td>.35</td>
</tr>
</tbody>
</table>

Abbreviation: VR, virtual reality.

* Scored on a Likert-like scale from 1 to 5, where 1 is never and 5 is daily.

### Table 2. Final-Product Score and Relative Reaction Time of the 2 Practice Groups in the Last Sessions of Initial Training and in the Retention Sessions

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean (95% CI)</th>
<th>(P) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Final-product score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distributed practice</td>
<td>15.1 (14.2-16.0)</td>
<td>.20</td>
</tr>
<tr>
<td>Massed practice</td>
<td>13.2 (12.4-14.0)</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Relative reaction time</strong></td>
<td></td>
<td>.90</td>
</tr>
<tr>
<td>Distributed practice</td>
<td>1.24 (1.16-1.32)</td>
<td>.01</td>
</tr>
<tr>
<td>Massed practice</td>
<td>1.31 (1.21-1.42)</td>
<td>.25</td>
</tr>
</tbody>
</table>

Mean final-product performance of the distributed and massed practice groups in the last 2 sessions of training and in the retention sessions. Bars indicate 95% CIs. Dotted lines indicate that 3 months have passed between measurements.

Relative reaction time on the secondary task of the distributed and massed practice groups in the last 2 sessions of training and in the retention sessions. Bars indicate 95% CIs. Dotted lines indicate that 3 months have passed between measurements.
Retention of Mastoidectomy Skills After VR Simulation Training

Discussion

In this follow-up study on the retention of mastoidectomy training in a VR simulator with distributed and massed practice of the procedure, we found that, regardless of the organization of training, final-product performance did not deteriorate significantly during a 3-month period in which the participants did not practice their skills. In contrast to this finding, the cognitive load estimated by reaction time measurement had returned to pretrained levels (distributed practice group: mean pretrained relative reaction time, 1.42 [95% CI, 1.37-1.47]; and massed practice group: mean pretrained relative reaction time, 1.34 [95% CI, 1.28-1.40]). Moreover, the skills of the massed practice group were less consolidated, and the participants in this group seemed to use more time within the allowed timeframe during retention testing to achieve a similar performance. During the retention procedures, the participants in general had a poorer performance compared with end-of-training procedures in adequately defining the outer boundaries of the procedure, often violating the facial nerve, and not exposing the facial recess sufficiently, suggesting that these items could be emphasized in future instructions.

In VR laparoscopic simulation skills training, participants' performance deteriorated in the immediate period following training, but no further skills were lost when retention was tested at a mean of 7 months after training. Similarly, novices retained skills in another laparoscopic simulator for 6 months. However, at 18 months, skills had returned to pretrained levels. A limitation to our study is, therefore, that retention was tested only at a relative early point after training (3 months), which might explain why final-product skills at performing mastoidectomy were largely retained in our follow-up study. Also, the participants had access to the simulator's built-in onscreen instructions on the procedure so as to have similar and comparable conditions during training and retention procedures and support self-directed practice with directed, self-regulated learning. Nonetheless, this access would also help increase performance during the retention procedures and compensate for possible differences between the 2 practice groups.

Other limitations to our study are the small sample size and a nonrandomized study design. Sample size calculations for learning curves are not well defined; for the original study, we aimed at having the number of participants in each practice group be similar to that in other studies. Based on the data on the end-of-training sessions and the included number of participants, a change in final-product score of 2.5 points would be needed to find a statistically significant difference between performance in the end-of-training and retention procedures. A type 2 error is therefore a possibility, and our study could be underpowered to detect smaller changes in performance between the end-of-training and retention sessions.

In the previously mentioned studies on VR laparoscopic simulation training, practice was organized in a distributed schedule. The retention of surgical skills in distributed vs massed practice has been studied for physical simulation models in surgery: distributed practice groups significantly outperformed massed practice groups when tested for retention at 1 month or 1 year. Surgical skills learned under distributed practice settings are therefore suggested to be more robust. Although we found that final-product mastoidectomy skills were retained regardless of practice organization, our supplemental analyses substantiate that time compensation was a factor: only about 5% to 15% of the total volume was removed during the last 10 minutes of the procedure in end-of-training and retention procedures for both groups except for the first retention procedure of the massed practice group (session 13). This finding corroborates that the improvement gained in time to completion during repeated training was not retained in the massed practice group and explains why final-product performance did not deteriorate markedly. When considering both the final-product performance and time to completion, our findings support that distributed practice is superior to massed practice for retention of mastoidectomy skills.

In this study, we performed retention testing using 2 repetitions to reveal retention and not refamiliarization with the simulator, which led to another interesting finding supporting the case for distributed practice being a superior method: the performance of the massed practice group increased significantly from the first to the second retention procedure. Also considering that the performance at the end of initial training was significantly lower for the massed practice group than for the distributed practice group, this finding suggests that the massed practice group still had potential for additional learning, whereas the distributed practice group had already reached an initial plateau during training and did not improve further during the retention testing. A similar pattern was found in a study on VR laparoscopic simulation: one group that had not trained repeatedly to a consistent performance in initial training also improved during retention testing, indicating that some degree of overlearning is beneficial for retention. Even though time spacing of practice is essential for learning, we found that it is possible to continue learning even after a considerable period of nonpractice. This finding is in agreement with a study on VR simulation training of endoscopic sinus surgery in which novices resumed to follow their learning curves after 11 to 60 days of not training.

In a study exploring the retention of skills at analyzing electrocardiograph results (a mainly cognitive skill) following a massed practice training course, approximately half of the performance gained during the course was lost after 2 weeks. In contrast to this finding, motor skills are consistently found to be less susceptible to decay over longer periods of time than are cognitive skills, and basic motor skills in VR laparoscopic simulation are better retained than are complex motor skills that placed heavier cognitive demands. In our initial study, we found that cognitive load decreased with repeated and distributed practice and not with massed practice. In the present study, we also measured retention of the performance on a secondary reaction time test. The relative reaction time reflects the cognitive load during the procedure, and we found that the cognitive load during the retention procedure had returned almost to the level of the first procedure. In agreement with current knowledge, this finding suggests that the reduction in cognitive demands with repeated practice of complex psychomotor skills is not retained as reliably as the acquired motor skills. This finding could have implica-
tions for surgical skills training, such as mastoidectomy training, because the aspect of cognitive learning also should be considered. Training toward cognitive automaticity of a surgical procedure requires substantially more training than training toward simulator proficiency alone.28

Conclusions

Mastoidectomy skills were largely retained at 3 months after self-directed VR simulation training when practice was organized with time distribution between practice sessions. The learning curve could, however, be resumed for the massed practice group because they had not reached their full learning potential during initial training. For both practice groups, the cognitive load during the retention procedures returned to the level of the first procedure. This finding substantiates that cognitive skills deteriorate more rapidly and that this factor should be considered in the organization of surgical skills training. Surgical skills should be reinforced regularly with a frequency that is sufficient to maintain acquired motor as well as cognitive skills.

REFERENCES