Avoiding Surgical Skill Decay: A Systematic Review on the Spacing of Training Sessions

Dario Cecilio-Fernandes, MSc,* Fokie Cnossen, PhD,† Debbie A.D.C. Jaarsma, DVM, PhD,* and René A. Tio, MD, PhD*‡

*Center for Education Development and Research in Health Professions (CEDAR), University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; †Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, Groningen, The Netherlands; and ‡Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

OBJECTIVE: Spreading training sessions over time instead of training in just 1 session leads to an improvement of long-term retention for factual knowledge. However, it is not clear whether this would also apply to surgical skills. Thus, we performed a systematic review to find out whether spacing training sessions would also improve long-term retention of surgical skills.

DESIGN: We searched the Medline, PsycINFO, Embase, Eric, and Web of Science online databases. We only included articles that were randomized trials with a sample of medical trainees acquiring surgical motor skills in which the spacing effect was reported. The quality and bias of the articles were assessed using the Cochrane Collaboration’s risk of bias assessment tool.

RESULTS: With respect to the spacing effect, 1955 articles were retrieved. After removing duplicates and articles that did not meet the inclusion criteria, 11 articles remained. The overall quality of the experiments was "moderate." Trainees in the spaced condition scored higher in a retention test than students in the massed condition.

CONCLUSIONS: Our systematic review showed evidence that spacing training sessions improves long-term surgical skills retention when compared to massed practice. However, the optimal gap between the re-study sessions is unclear.

KEY WORDS: spacing effect, surgical skills, long-term retention, distributed practice, simulation training, medical education

COMPETENCIES: Medical Knowledge, Practice-Based Learning and Improvement

INTRODUCTION

Traditionally, surgical skills have mostly been taught through mentoring and apprenticeship. Recently, McGaghie stated that the underlying assumption of apprenticeship-based clinical training is that students gain competence over time simply by exposing them to patients and experience. He argued that it lacks structured learning objectives, skill practice, and objective assessment with feedback. In the past decades, medical skills training has been shifting toward simulation-based mastery training, and currently it is appreciated that deliberate practice in a simulation lab is a valuable add-on to learning surgical skills. This type of training lays emphasis on achieving defined learning objectives and offers students an opportunity to practice skills without time restrictions. It can be tailored to individual student’s needs concerning skills, knowledge, attitudes, and the decision-making process, which, in turn, allows students to learn at their own pace in a safer, more ethical environment.

Surgical skills training requires a large amount of instructor time, effort, and resources. Furthermore, an acquired surgical skill will decay over time after periods of nonuse, which could potentially be a threat to patient safety. Most skills training sessions focus on student learning rather than long-term retention. Research revealed that some students had not been able to proficiently perform the required skill or 12 months after they had finished their training. These findings imply that practicing until proficiency may not be
enough to guarantee long-term retention. Thus, improving long-term retention of surgical skills becomes crucial to safeguard patient care. Based on cognitive psychology, several guidelines for medical skills training suggest the spacing effect as a way to avoid skills decay. The spacing effect refers to spacing training sessions over time rather than training in just 1 session (massed learning). A comprehensive review that investigated several learning techniques showed that the spacing effect was the most effective strategy for students’ learning when compared to other techniques. Spaced training has been shown to improve long-term knowledge and skills retention, for instance, in tasks concerning verbal recall, English as a second language, computerized spelling, reading, biology, mathematics, medical knowledge, arm movements, command-and-control simulation, and dynamic balance.

The key to improve long-term retention is the time between training sessions, which is known as the intersession interval. The space between the training sessions will determine the retention interval, which is the time between the last training session and the final test. The longer the required retention interval, the longer the intersession intervals should be. A review from the psychology literature suggested that, for the best knowledge retention, the intersession interval should be approximately 10%-15% of the retention interval.

In the medical education literature, some authors recommend spacing the training sessions to increase skills retention, but it remains unclear how often trainees should practice or what the duration of the intervals between the training sessions should be. To optimize skills training and foster retention, we performed a systematic review to answer the following research questions:

1. Is spaced practice better than massed practice for acquiring and retaining surgical skills?
2. If so, what would be the optimal intersession interval?

To answer our research questions, we conducted a systematic review on studies on the spacing effect related to surgical skills retention. We strived to identify underlying theories as well as aspects of the spacing effect that were taken into account in the design of skills training programs.

METHODS

We conducted a systematic review using principles of the PRISMA Guidelines and guidelines provided in Medical Education.

Search Strategy and Data Sources

We searched the Medline, PsycINFO, Embase, Eric, and Web of Science online databases in February 2016. No language or other limitations were imposed on the search. We first searched the terms "skill* retention", "skill* acquisition", and "spacing effect". As we noticed that the terms "distributed and retrieval" were often used as synonyms for "spacing and testing", we included these words as key words. The search strategy used for Medline was as follows:

1. "(("skill* retention" OR "skill* development" OR "skill* retrieval" OR "skill* acquisition" OR "skill* retrain")).) AND (distribut* OR spac* OR massed)
2. ("distributed practice") AND skill*
3. ("spacing effect") AND skill*

The search strategy was adapted for the other databases. Subsequently, we hand-searched the reference lists of identified articles for citations of additional relevant articles. Web of Science and Google Scholar were searched for citing articles of all included articles.

Inclusion Criteria

Studies were included if they met the following criteria:

2. Intervention: The intervention had to be on surgical skill acquisition.
3. Comparison: Comparisons had to include at least 2 of the following conditions: control, massed, and spaced.
4. Outcome: Change in surgical task performance as measured by motor skill performance.
5. Study design: Randomized trial.

Study Selection

Two authors (D.C.-F. and R.T.) independently reviewed the titles and abstracts of the retrieved publications. Each article was initially categorized as “maybe” or “excluded” based on the information of the titles and abstracts. If one of the reviewers had classified an article as “maybe,” the full text was retrieved to verify whether the article met the inclusion criteria. In the subsequent stage, the same authors independently reviewed the full articles. All articles that matched the inclusion criteria were included in the review.

Data Extraction

The first author extracted and documented information about the type of task, design of the experiment, participants, groups and practice schedule, length of the retention interval, measures, spacing, and main findings. The other authors verified the retrieved information.

Quality Criteria

We assessed the quality and bias of the articles using the Cochrane Collaboration’s risk of bias assessment tool based on
the sequence generation, allocation concealment, blinding of participants, and outcome assessors and outcome data.25

RESULTS

In total, 1955 articles were retrieved (Embase = 374, Eric = 320, Medline = 354, PsyCINFO = 228, and Web of Science = 679). After removing the duplicates, 1302 articles were identified. Subsequently, 1254 articles were excluded because the titles and abstracts did not meet the inclusion criteria. The main reason for excluding articles was that the investigated skills were not surgical skills (n = 974). We retrieved the remaining 48 full articles for further investigation of which 39 articles were excluded. These 39 articles were excluded because the investigated skill was not surgical (n = 5), there was no comparison between at least 2 of the 3 conditions (n = 23), the participants were not randomly assigned to the groups (n = 9), and when the full article was not available (n = 2). One more article was found through hand-searching for citing articles, and another article was retrieved by NCBI alert. Finally, 11 articles were included. For an overview, see Figure.

Based on the Cochrane Collaboration risk of bias tool,25 we assessed the quality of the 11 articles. The overall quality of the experiments was “moderate.”25 Risk assessment revealed that 7 studies were considered as having unclear bias and 4 studies as having low risk of bias in their methodology for selection, performance, attrition, and detection. The overall risk of bias across study domains is described in Table 1. It was not possible to assess the blinding of participants and instructors or selection bias. Most articles did not address either of these issues; in particular, random sequence generation and allocation concealment were not clearly explained. Random sequence generation assesses whether a study used a randomized sequence of assignments. Allocation concealment assesses whether those sequences are protected and concealed from the participants, assessors, and instructors.

Table 2 displays the following characteristics of the included studies: type of task, design of the experiment, who the participants were, which groups and practice schedules were included, length of the retention interval, measures, the effect of spacing, and also the main findings. The surgical tasks were suturing and knot-tying, a laparoscopic transfer task, end-to-side vascular anastomosis, laparoscopic suturing, and microvascular anastomosis. The complexity and difficulty of the tasks differed. All tasks had an emphasis on motor skills. The goal of all articles was to improve learning and retention of the motor skill of the task.

**Is Spaced Practice Better Than Massed Practice for Surgical Tasks?**

In 5 studies, massed practice was compared to spaced practice.26,28,32,35,36 The intersession interval ranged from 5 minutes to 1 week. The retention interval ranged from 5 minutes to 1 year (Table 2). In all of these studies, spaced practice outperformed massed practice at the final measurement.

**What Is the Optimal Intersession Interval?**

Different intersession intervals were compared in 6 studies.28-31,33,36 The length of the intervals ranged from 5 minutes to 1 month. The lengths of the intersession interval that were compared differed per study: weekly versus daily, 1 additional training versus the same amount of time spread over 5 sessions, 1 hour per week training for 3 weeks versus 6 weeks, weekly versus monthly.

The retention interval of the included studies ranged from 5 minutes to 7 months. In none of the studies a significant effect on retention was found between the lengths of intersession intervals, except in a study conducted by De Win et al.28 who did find differences in retention between various intersession intervals: 6 months after the last training session, the group with 1 session per day outperformed the other groups (Table 2).

**Underlying Theory and Implementation of the Spacing Effect**

Although all 11 studies addressed the spacing effect, in only 2 of them26,27 the term “spacing effect” was mentioned. In 2 of
the studies, no theory related to the spacing effect or distributed practice was explicitly mentioned.31,34 The 7 remaining studies referred to the term “distributed practice,”28-30,32,33,35,36 but only in 3 of them a definition was provided.29,32,36Nine studies26-31,33,35,36 were based on previous studies stating that distributed practice improves long-term retention compared to massed practice. Furthermore, several explanations for using distributed practice were given. Two studies referred to psychological theory,26,29 and 3 other constructs were cited as follows: (1) deliberate practice,26,33 (2) overlearning,26,33 and (3) ongoing training.30,34

The way spaced practice was implemented differed per study. However, none of the studies referred to the spacing effect in the design of the training program. Besides, none of the training schedules was based on literature.

DISCUSSION

From the literature, we know that spacing the study sessions improves the long-term retention of factual knowledge. However, it is not clear whether surgical skills’ long-term retention would benefit from spacing the training sessions. Thus, the purpose of our systematic review was to find experimental studies on the spacing effect in acquiring and retaining a surgical skill. Our results showed that students who practiced under the spaced condition scored higher on a retention test than those who practiced under the massed condition. This finding is in concordance with educational and psychological literature, which also found that the spacing effect increases long-term retention.37 However, the optimal training schedule remained unclear.

To determine the optimal gap between the training sessions, it might be beneficial to take the complexity of the tasks into account. Previous research has shown that for simple motor tasks, the intersession interval should be shorter, whereas for complex motor tasks, the intersession interval should be longer.38 However, our review demonstrated that even suturing, which is considered a simple motor task in surgery, benefits from spacing the training sessions.

An interesting question is, therefore, why complex tasks profit more from the spacing effect. The answer may lie in the nature of the knowledge that is required for a particular task. Cognitive psychology distinguishes knowledge between declarative and procedural, where procedural knowledge refers to “knowing how,” and declarative knowledge refers to “knowing what,” i.e., knowledge about facts.39 Interestingly, it is assumed that declarative knowledge will be forgotten over time if not used, whereas procedural knowledge does not show such decay.40 This implies that tasks that rely on declarative knowledge profit from spacing because it counteracts the decay that would otherwise occur, while tasks relying on procedural knowledge only do not show decay, and hence do not profit from repeated practice sessions.

### TABLE 1. Risk of Bias of Included Studies

<table>
<thead>
<tr>
<th>Reference</th>
<th>Overall Bias</th>
<th>Random Sequence Generation (Selection Bias)</th>
<th>Allocation Concealment (Selection Bias)</th>
<th>Blinding of Participants and Personnel (Performance Bias)</th>
<th>Blinding of Outcome Assessors (Detection Bias)</th>
<th>Incomplete Outcome Data (Attrition Bias)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spruit et al.26</td>
<td>Unclear risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Akdemir et al.27</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>De Win et al.28</td>
<td>Unclear risk</td>
<td>High risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Willis et al.29</td>
<td>Unclear risk</td>
<td>High risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Uninlear</td>
<td>Low risk</td>
</tr>
<tr>
<td>Van Bruwaene et al.30</td>
<td>Unclear risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Robinson et al.31</td>
<td>Unclear risk</td>
<td>High risk</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>Uninlear</td>
</tr>
<tr>
<td>Gallagher et al.32</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Uninlear</td>
</tr>
<tr>
<td>Mitchell et al.33</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Stefanidis et al.34</td>
<td>Unclear risk</td>
<td>Low risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
</tr>
<tr>
<td>Moulton et al.35</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Mackay et al.36</td>
<td>Unclear risk</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Low risk</td>
<td>Uninlear</td>
<td>Low risk</td>
</tr>
<tr>
<td>Reference</td>
<td>Skill</td>
<td>Research Design</td>
<td>Participants</td>
<td>Groups/Practice Schedules</td>
<td>Retention Interval</td>
<td>Measures</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------</td>
<td>-------------------</td>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Spruit et al.26</td>
<td>Laparoscopic</td>
<td>Randomized</td>
<td>41 medical students</td>
<td>Massed: practice in 1 day (n = 21)</td>
<td>2 weeks and 1 year</td>
<td>Completion time and accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Spaced: practice once a week for 3 wk (n = 20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akdemir et al27</td>
<td>Basic laparoscopic (salpingectomy)</td>
<td>Randomized</td>
<td>22 gynecology residents</td>
<td>Training: After the basic training, 1 h/wk for 4 wk Control: No training after the basic training</td>
<td>5 weeks</td>
<td>Time, economy of movement scores, and error scores.</td>
</tr>
<tr>
<td>De Win et al.28</td>
<td>Laparoscopic knots</td>
<td>Randomized</td>
<td>145 Medical students</td>
<td>All groups participated on 6 sessions training. Group 1 = 3 sessions daily (n = 22). Group 2 = 2 sessions daily (n = 25). Group 3 = 1 session per day (n = 21). Group 4 = 1 session on alternative days (n = 24). Group 5 = 1 session weekly (n = 26). Group 6 = 1 session weekly with “deliberate practice” between sessions. (n = 27).</td>
<td>1 and 6 mo</td>
<td>The cumulative time to approximate the skin edges adequately was used as measurement. 3 validated video-trainer tasks— Southwestern Drills— checkerboard, bean drop, and running string.</td>
</tr>
<tr>
<td>Willis et al.29</td>
<td>Laparoscopic transfer task</td>
<td>Randomized</td>
<td>75 preclinical medical students</td>
<td>Mass practice = 3 training separated by 5 min breaks. Similar surgical exercise = 3 session separated by a similar task Dissimilar surgical exercise = 3 training</td>
<td>Mass, similar and dissimilar = after the nontarget task</td>
<td>Pretests and posttests consisted in 1 peg transfer trial.</td>
</tr>
<tr>
<td>Reference</td>
<td>Skill</td>
<td>Research Design</td>
<td>Participants</td>
<td>Groups/Practice Schedules</td>
<td>Retention Interval</td>
<td>Measures</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>----------------</td>
<td>--------------</td>
<td>---------------------------</td>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Van Bruwaene et al. 30</td>
<td>Suturing on a box trainer and on a cadaver porcine Nissen model.</td>
<td>Randomized controlled</td>
<td>39 medical students</td>
<td>Group 1: without additional training (n = 9) &lt;br&gt; Group 2: 1 supervised training session (150 min) after 2.5 mo (n = 10). &lt;br&gt; Group 3: 5 monthly unsupervised training sessions of 30 min on a box trainer (n = 10). &lt;br&gt; Group 4: 5 monthly unsupervised training sessions of 30 min on Lap Mentor (n = 10).</td>
<td>5 mo</td>
<td>Retention testing included suturing on a box trainer and on a cadaver porcine Nissen model.</td>
</tr>
<tr>
<td>Robinson et al. 31</td>
<td>End-to-side vascular anastomosis</td>
<td>Randomized controlled</td>
<td>37 junior residents</td>
<td>Short course was 3 wk with 1 h teaching per week (n = 18). &lt;br&gt; Long was 6 wk long with 1 h teaching per week (n = 19).</td>
<td>1 and 16 weeks</td>
<td>Knowledge and technical proficiency were measured with a standard 50-point vascular skills assessment (SVSA).</td>
</tr>
<tr>
<td>Gallagher et al. 32</td>
<td>Laparoscopic box trainer task</td>
<td>Randomized controlled</td>
<td>Study 1: 24 novices &lt;br&gt; Study 2: 16 novices</td>
<td>Study 1: Massed condition completed the training of all 6 MIST VR tasks 3 times in 12 hour (n = 8). &lt;br&gt; Control group did not receive any training on MIST VR tasks (n = 8). &lt;br&gt; Study 2: Interval condition completed the 6 MIST VR tasks once per day on 3 consecutive days (n = 8). &lt;br&gt; Study 2: 1 and 2 wk after the last training</td>
<td>Study 1: The cutting task performance was judged as correct or incorrect. &lt;br&gt; Study 2: The same as the first study.</td>
<td>Study 1: The practice condition group was significantly better than massed and control groups. &lt;br&gt; Study 2: Practice group performed better than no practice group in T3.</td>
</tr>
<tr>
<td>Study 2: Practice group:</td>
<td>Mitchell et al. 33</td>
<td>End-side vascular anastomosis</td>
<td>Randomized</td>
<td>24 surgical interns</td>
<td>Weekly group: 1 training 4 mo per week ( (n = 12) ). Monthly group 1 training per month ( (n = 12) ).</td>
<td>Validated procedural checklist scores and global rating scores. Final product analysis and overall performance.</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
<td>------------</td>
<td>---------------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>Nonpractice: no additional practice ( (n = 8) ).</td>
<td>Stefanidis et al. 34</td>
<td>Laparoscopic suturing</td>
<td>Randomized controlled</td>
<td>18 medical students</td>
<td>Control group: No additional training ( (n = 9) ). Ongoing training group: practiced until proficiency after 1 and 3 mo retention tests ( (n = 9) ).</td>
<td>The students performed 3 repetitions of laparoscopic suturing at 2 wk, 1, 3, and 6 mo.</td>
</tr>
<tr>
<td>Moulton et al. 35</td>
<td>Microvascular anastomosis</td>
<td>Stratified randomized</td>
<td>38 surgical residents</td>
<td>Massed group: practices 4 sessions in 1 day ( (n = 19) ). Distributed group: practices the 4 sessions once a week ( (n = 19) ).</td>
<td>Expert-based evaluations of performance</td>
<td>The distributed group performed significantly better than the massed group in most of the outcomes on the retention test. Group B performs better than group A in overall scores and time. There was no difference between groups A and C.</td>
</tr>
<tr>
<td>Mackay et al. 36</td>
<td>Laparoscopic transfer place</td>
<td>Randomized controlled</td>
<td>41 undergraduate or postgraduate students</td>
<td>Group A: 20 min training without interval ( (n = 14) ). Group B: 5 blocks of 4 min with 2.5 min interval ( (n = 14) ). Group C: 5 blocks of 3 min with 2.5 min interval ( (n = 13) ).</td>
<td>5 min overall scores, time, errors, and path length economy</td>
<td></td>
</tr>
</tbody>
</table>
The difference between very simple and highly complex motor skills may lie in the amount of declarative and procedural knowledge that is necessary for the task. More complex tasks such as end-to-side vascular anastomosis may require more declarative knowledge than more simple tasks such as suturing, and therefore would profit more from spacing. This may also explain our somewhat unexpected finding that even suturing profited from spacing. But when we look more closely at the suturing task, we can see that for this so-called simple motor skill, declarative knowledge was required: knowledge of the anatomical structures that were sutured, knowing the preferred position of the knot, how many knots were necessary, etc. This may imply that dividing the components of a skill into declarative and procedural knowledge may be a way to optimize skill training, as it allows us to use the best teaching strategies based on the necessary knowledge.

This study has a few limitations. First, the final number of articles obtained in the systematic review is low. Despite our comprehensive search in different databases, we only retrieved 11 articles about the spacing effect with respect to surgical skills acquisition. In our search, we only included randomized trials, forcing us to exclude studies that did address the spacing effect in other fashions. The participants included in the reviewed articles ranged from medical students to residents. Interestingly, however, in all the studies, participants in the spaced group scored higher on the retention test than those in the massed group. Because of the small number of articles and differences in methodology and measurements, it was not possible to conduct a meta-analysis. Despite these limitations, the results of our systematic review suggest that spacing the training sessions may improve long-term retention of surgical skills.

For the training of surgical residents, simulation training is nowadays common practice. Very often, massed training sessions are being used for logistic reasons. Based on our findings, we believe that such a training strategy may be less effective than spreading multiple sessions over time. The optimal gap between the study sessions, however, has yet to be established.

CONCLUSION

Our systematic review showed evidence that the spacing effect improves surgical skills retention. When setting up a skills training for novices, spacing the training sessions should be employed such that simulation lab time is used as effectively as possible.

CONTRIBUTORS

All 4 authors contributed significantly to the design of the review. D.C.F. conducted the background research to develop and search key words and MESH terms. Data were extracted by the D.C.F. D.C.F. and R.A. independently reviewed all title and abstracts. The included articles were interpreted by all 4 authors. D.C.F. wrote the first draft of the manuscript. All authors provided critical revision of the manuscript for important intellectual content. All authors approved the final manuscript for submission and agreed to be accountable for all aspects of the work.

FUNDING

This work was supported by CAPES – Brazilian Federal Agency for Support and Evaluation of Graduate Education—Grant 9568-13-1 awarded to Dario Cecilio-Fernandes.

ACKNOWLEDGMENTS

The authors thank Mrs. Tineke Bouwkamp-Timmer for her feedback on the final version of the article and editorial help.

REFERENCES


kind of practice makes perfect?: a randomized, controlled trial *Ann Surg.* 2006;244(3):400-409.


