Author: Spruit, E.N.
Title: Increasing the efficiency of laparoscopic surgical training: assessing the effectiveness of training interventions
Issue Date: 2018-01-24
THE EFFECTS OF SPACING, NAPS AND FATIGUE ON THE ACQUISITION AND RETENTION OF LAPAROSCOPIC SKILLS

Edward N. Spruit MSc.
Dr. Guido P. H. Band
Dr. Kristiaan B. van der Heijden
Prof. Jaap F. Hamming

Based on the publication:
Abstract

**Background:** Earlier research has shown that laparoscopic skills are trained more efficiently on a spaced schedule compared to a massed schedule. The aim of the study was to estimate to what extent the spacing interval, naps and fatigue influenced the effectiveness of spacing laparoscopy training.

**Methods:** Four groups of trainees (aged 17-41 years; 72% female; N\textsubscript{massed} = 40; N\textsubscript{break} = 35; N\textsubscript{break-nap} = 37; N\textsubscript{spaced} = 37) without prior experience were trained in three laparoscopic tasks using a physical box trainer with different scheduling interventions. The first (massed) group received three 100-minute training sessions consecutively on a single day. The second (break) group received the sessions interrupted with two 45-minute breaks. The third (break-nap) group had the same schedule as the second group, but had two 35-minute powernap intervals during the breaks. The fourth (spaced) group had the three sessions on three consecutive days. A retention session was organized approximately three months after training.

**Results:** The results showed an overall pattern of superior performance at the end of training and at retention for the spaced group, followed by the break-nap, break and massed group, respectively. The spaced and break-nap group significantly outperformed the break and massed group, with effect sizes ranging from .20 to .37.

**Conclusions:** Spacing laparoscopic training over three consecutive days or weeks is superior to massed training, even if the massed training contains breaks. Breaks with sleep opportunity (i.e. lying, inactive, muted sensory input) enhance performance over training with regular breaks and traditional massed training. For optimal skill acquisition and retention of laparoscopic skills, larger spacing intervals (at least up to a week) are recommended.
Introduction

Acquiring laparoscopic motor skills requires an extensive amount of practice. Earlier research suggested that practice time allocated for surgical training can be used most efficiently when scheduled across multiple smaller time intervals (Moulton et al., 2006; Gallagher, Jordan-Black & O'Sullivan, 2012), preferably with several non-training days in between training sessions. In an earlier study (Spruit, Band & Hamming, 2014, chapter 7), we established that laparoscopic skills are better acquired and retained when learned on a spaced schedule as compared to a massed schedule. In the study, two groups of participants learned four basic and one advanced task (intra-corporeal suturing) on a physical box trainer. The first group received training on a massed 1-day schedule, while the second group received the same amount of training divided over three sessions spaced across three consecutive weeks. Performance at the end of training, at a two week retention session and a one year retention session was superior for the spaced group. These spacing effects were most pronounced for the advanced task.

Spacing training provides multiple benefits. Having multiple shorter training sessions reduces the likelihood of trainees becoming overly fatigued (Kahol et al., 2008) or bored, which may be the case when practicing laparoscopic tasks for an extended amount of time. Also, trainers and trainees get a more accurate reflection of their actual skill level during spaced training (Bjork, 1999). Since the training uses multiple sessions, trainees will discover they do not perform as smoothly at the start of the second training session as they did at the end of the first training session. This gives them a more accurate appraisal of their own skill level. The fresh start at the beginning of each new training session requires trainees to invest more effort to get their performance back up to the level they left at the end of the previous session. This allows for more elaborate and frequent activations of the neural pathways in the brain associated with learning the specific motor skill.

Theorists have suggested that consolidation of memory is adaptive (Wang, Zhou & Shah, 2014), and that learning a skill is favored on a spaced training schedule because frequent short practice sessions on a training task give the brain the indication that it will encounter the same task more often in the future. This triggers more enduring memory encoding in order to accommodate for repetitive encounters with the task. Spacing provides time intervals between training sessions during which memory consolidation (Stickgold, 2005) can be strengthened. This in turn supports longer retention. Consolidation occurs in the brain when a person is disengaged from the trained activity and is enhanced during sleep (Brashears-Krug, Shadmehr & Bizzi, 1996), when perceptual input to the brain is reduced to a large extent. In designing training, one ought to be cognizant of the fact that most learning takes place during off-line periods, not during practice. During training, practice provides the learning input that will be consolidated at a later period of time post-
training. If too many similar training tasks are practiced right after another, the learning input created by the first training experience will be replaced and thus impaired by later training experiences (Stickgold, 2005). A more adequate strategy is to provide trainees rest intervals so the associated brain regions can have the opportunity to process the learning input before proceeding with more training. Research has shown consolidation is enhanced by overnight sleep (Walker, Brakefield, Morgan, Hobson & Stickgold, 2002) as well as power naps (Korman et al., 2007).

From our earlier results (Spruit, Band & Hamming, 2014, chapter 7), we were unable to distinguish to what extent each of these factors facilitates acquisition and retention of laparoscopic motor skills acquired on a physical box trainer. In the current study we use a more elaborate design in order to dissect the individual effects that might contribute to spacing training.

The goal of this study was two-fold: (1) to compare the effectiveness of an intervention with a smaller spacing interval to a bigger one (one day versus one week) and (2) to determine the influence of factors such as naps, mental fatigue and simple breaks on the skill acquisition and retention of laparoscopy training. With the current study we aimed to attain valid estimates of the extent to which each of these factors influences the effectiveness of spacing training.

First, we hypothesize that groups that have a larger time interval between training sessions will have superior performance at the end of training and at retention and experience lower levels of fatigue. Second, we hypothesize that sleep opportunities in between training sessions will lead to superior performance at the end of training and at retention and result in lower levels of fatigue.

**Methods**

**Participants**

149 university students (108 female) without any prior experience in laparoscopy training were enrolled in the study. Age ranged from 17-41y (mean = 21) and 128 participants were right-handed. All subjects filled out informed consent forms and were granted a training certificate as a reward for participating in the study.

**Apparatus**

Participants trained three laparoscopic tasks on a physical box trainer. All tasks have previously established construct validity (Kolkman, van de Put, Wolterbeek, Trimbo & Jansen, 2008). The tasks train perceptual and motor skills such as depth perception, adapting to the fulcrum effect and instrument handling, all key skills for mastering
laparoscopic surgery. In the first task, participants had to coordinate a pipe cleaner through a set of four rings. This task is utilized to improve a trainee’s bi-manual dexterity. In the second task, participants had to pick up small beads from a bucket and drop them on pins on a pegboard making the shape of a simple figure. This task requires caution and very careful handling of the pins with the instruments. In the final task, participants learn intra-corporeal suturing. The suturing sequence required correct insertion of the needle and three knots. The first knot required two throws, while the second and third knot required one throw. Participants were instructed to start with the needle in their right instrument for the first and third knot and to start with the needle in their left instrument for the second knot. This was done to ensure flexibility of the skill for both hands (see online video appendix listed in the reference list for all the laparoscopic training tasks’). During training, participants also learned suturing on an open model in order to prepare them before practicing in the laparoscopic box trainer.

Video footage of the performance of participants on the box-trainers was converted to .mpg files for each task at each moment of measurement using a video splitter, grabster (Terratec Grabster AV 400 MX) and VLC Media Player for Windows. Participants filled out self-report questionnaires covering demographics (sex, age, etc.), prior sport, music and gaming experience (0 = no experience, 1 = I used to play, 2 = yearly, 3 = monthly, 4 = weekly, 5 = daily), personality (Gosling, Rentfrow & Swann, 2003) and mental fatigue (RSME, Rating Scale Mental Effort, Zijlstra, 1993).

**Training Programs**

Training was divided into three sessions. The duration of each session was 100 minutes, divided into 50 minutes of practice on the training tasks, 15 minutes of instructions and 35 minutes of measurement (laparoscopic tasks, questionnaires). All three laparoscopic tasks were practiced and measured during each of the three sessions, but the duration of practice for the pipe cleaner and beads task was reduced for the later sessions, while practice time on intra-corporeal suturing was increased. Total practice time for the pipe cleaner, beads and intra-corporeal suturing task was 27.5, 32.5 and 60 minutes, respectively. During the first session, measurement took place before practice to establish a baseline. At the baseline measurement for intra-corporeal suturing, only the insertion of the needle in the model is performed, since the knot tying part of the task is too difficult for a novice without any prior training. For the second and third session, measurement took place after practice at the end of the session. RSME questionnaires were filled out right before measuring the laparoscopic tasks on each session. A retention session was scheduled approximately three months after training. All three laparoscopic tasks were measured during retention without any prior practice.
Participants received standardized instructions by the trainer and instructional videos (see online appendix in reference list*) and no feedback in order to minimize confounding effects on the learning curve of the trainees. If trainees asked for feedback they were reminded to pay close attention next time the instructional video would be shown, since all the required information to learn is present in the video. 

There were four groups, each with a different time schedule. The spaced group received the training sessions on three consecutive days. The massed group had all sessions consecutively on a single day without any breaks. The break and break-nap group received the sessions on one day with two 45 minute breaks in between, with the break-nap group having a 35 minute powernap opportunity during both breaks. In the break-nap group, participants had a powernap opportunity on inflatable mattresses in a dark room while wearing earplugs and a sleeping mask to minimize sensory input. After each powernap, a self-report sleep questionnaire was filled out. All participants in the break and break-nap group wore activity trackers (Flex, Fitbit) on their wrist to measure their activity levels during training and during the breaks. The instrument uses three-dimensional motion sensing technology, records data in one-minute epochs, and has shown to be an acceptable instrument for sleep/wake monitoring in normative populations (Montgomery-Downs, Salvatore & Bond, 2012). Participants were assigned randomly to each group. All groups trained their laparoscopic skills for an equal amount of time.

Performance outcome measures

The video files of the participants were assessed by the first author for completion times of the task. If participants were unable to complete the task within a set amount of measurement time (maximum of ten minutes during training and fifteen minutes during retention), a score of 601 or 901 (seconds) was assigned in order to avoid selective drop-out from our sample based on poor performance. This score would automatically be assigned as the highest rank in the non-parametric tests. Thus, there were three outcome measures (pipe-cleaner, beads, suturing task) at each moment of measurement (at baseline, end of session 2, at the end of training, and retention). A total score was computed as a fourth measure using z-scores from the laparoscopic tasks (we chose this method in order to make sure all three tasks contribute equally to the total score, since the laparoscopic tasks have differing mean completion times).

Statistical Analysis

Data were checked for normality and statistical tests were chosen accordingly (One-way ANOVA, independent samples and paired samples t-tests in the case of normal data and Kruskal-Wallis, Mann-Whitney and Wilcoxon Signed-rank tests in the case of non-normal data) using the statistical software SPSS 23.0. A significance level of .05 was
used. We tested whether groups were comparable at baseline in terms of age, sex, hand preference, academic year, food, caffeine, alcohol intake, sleep prior to training, mental fatigue, musical, gaming, sports activity, and personality factors. For our main analysis, we tested all four laparoscopic scores (pipe cleaner, beads, suturing and total scores) on all four moments of measurement (baseline, at the end of session two, at the end of training and at the retention session) with 16 Kruskal-Wallis tests.

Results

149 participants (N\textsubscript{massed} = 40; N\textsubscript{break} = 35; N\textsubscript{break-nap} = 37; N\textsubscript{spaced} = 37) completed the training and 134 participants (N\textsubscript{massed} = 36; N\textsubscript{break} = 27; N\textsubscript{break-nap} = 35; N\textsubscript{spaced} = 36) returned to the lab for the follow-up retention session.

Participant characteristics

Chi-square tests showed no significant differences between all the groups in terms of sex, hand preference and food intake. Caffeine intake was lower in the break-nap group (35.1%) compared to the other groups (massed: 62.5%; break: 51.4%; spaced: 70.2%; p = .022). Alcohol intake (on the night prior to training) differed between groups (massed: 62.5%; break: 37.1%; break-nap: 35.1%; spaced: 13.5%; p < .001). In order to check for any effects of caffeine or alcohol intake on performance on the laparoscopic tasks, eight Mann-Whitney tests were performed (two within each of the four groups). None of the tests were significant, indicating no influence of caffeine or alcohol intake in the sample. Mann-Whitney tests revealed no significant differences between groups in age, musical gaming and sports activity. However, the break-nap group had more participants in the initial years of their study (Median(Mdn)\textsubscript{break-nap} = 1.0y; Mdn\textsubscript{massed} = 3.0y; Mdn\textsubscript{break} = 3.0y; Mdn\textsubscript{spaced} = 2.0y). Non-parametric correlations showed no significant relations between years of study and performance on any of the laparoscopic tasks.

One-way ANOVAs showed no significant differences between groups in any of the big five personality factors and hours of sleep (the night prior to training). For quality of sleep (the night prior to training) the results were mostly comparable, with the exception of a slightly lower sleep quality in the break group compared to the break-nap group (p = 0.03, Mean\textsubscript{break} = 3.38, SD\textsubscript{break} = 0.92; Mean\textsubscript{break-nap} = 3.92, SD\textsubscript{break-nap} = 0.76), but without differences for the other two groups (Mean\textsubscript{massed} = 3.78, SD\textsubscript{massed} = 0.66; Mean\textsubscript{spaced} = 3.57, SD\textsubscript{spaced} = 0.83). Also, baseline mental fatigue was significantly higher in the spaced group compared to the break-nap group (p = 0.001, Mean\textsubscript{spaced} = 22.53, SD\textsubscript{spaced} = 12.25; Mean\textsubscript{break-nap} = 12.42, SD\textsubscript{break-nap} = 8.40), but without differences for the other two groups (Mean\textsubscript{massed} = 18.94, SD\textsubscript{massed} = 12.10; Mean\textsubscript{break} = 21.45, SD\textsubscript{break} = 12.18). Non-parametric correlations
showed no significant relations between sleep quality (the night prior to training) or baseline mental fatigue and performance on any of the laparoscopic tasks.

Results of the activity trackers showed a large difference in the estimated minutes spent asleep between the break and the break-nap group for both of the two breaks (Break 1: \( p < 0.001; \) \( \text{Mdn}_{\text{break}} = 0; \) \( \text{Mdn}_{\text{break-nap}} = 10; \) Break 2: \( p < 0.001; \) \( \text{Mdn}_{\text{break}} = 0; \) \( \text{Mdn}_{\text{break-nap}} = 12; \) Self-reported minutes asleep in the break-nap group ranged from 0 to 30 (\( \text{Mean}_{\text{nap1}} = 5.95; \) \( \text{Std.Dev.}_{\text{nap1}} = 6.62; \) \( \text{Mean}_{\text{nap2}} = 6.73; \) \( \text{Std.Dev.}_{\text{nap2}} = 8.22 \)). The activity trackers’ measure of minutes spent asleep correlated significantly with the associated self-report measure (\( r = .353, p = .032 \)). No significant correlations were found between self-reported sleep the night prior to training and self-reported (\( r = -.178, p = .291 \)) and activity trackers’ measure (\( r = -.148, p = .218 \)) of minutes spent asleep during the breaks, although the coefficients were in the direction one would expect.

### Table 1

Mean ranks (from multiple Kruskal-Wallis Tests) for all four training groups at baseline, the end of session II, at the end of training (\( N_{\text{massed}} = 40; \) \( N_{\text{break}} = 35; \) \( N_{\text{break-nap}} = 37; \) \( N_{\text{spaced}} = 37 \)) and at retention (\( N_{\text{massed}} = 36; \) \( N_{\text{break}} = 27; \) \( N_{\text{break-nap}} = 35; \) \( N_{\text{spaced}} = 36 \)).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Massed</th>
<th>Break</th>
<th>Break-nap</th>
<th>Spaced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipe cleaner _baseline</td>
<td>76.95</td>
<td>71.37</td>
<td>69.47</td>
<td>81.85</td>
</tr>
<tr>
<td>Pipe cleaner _end of session II</td>
<td>78.95</td>
<td>79.03</td>
<td>62.36</td>
<td>77.66</td>
</tr>
<tr>
<td>Pipe cleaner _end of training</td>
<td>84.71</td>
<td>81.37</td>
<td>65.58</td>
<td>67.89</td>
</tr>
<tr>
<td>Pipe cleaner _retention</td>
<td>82.99</td>
<td>79.56</td>
<td>55.70</td>
<td>54.44</td>
</tr>
<tr>
<td>Beads _baseline</td>
<td>82.61</td>
<td>70.21</td>
<td>73.78</td>
<td>70.49</td>
</tr>
<tr>
<td>Beads _end of session II</td>
<td>82.85</td>
<td>72.36</td>
<td>75.82</td>
<td>64.07</td>
</tr>
<tr>
<td>Beads _end of training</td>
<td>84.22</td>
<td>72.89</td>
<td>69.09</td>
<td>71.19</td>
</tr>
<tr>
<td>Beads _retention</td>
<td>74.97</td>
<td>74.63</td>
<td>63.71</td>
<td>56.42</td>
</tr>
<tr>
<td>Suturing _baseline</td>
<td>71.41</td>
<td>76.79</td>
<td>76.42</td>
<td>73.81</td>
</tr>
<tr>
<td>Suturing _end of session II</td>
<td>82.98</td>
<td>79.39</td>
<td>72.03</td>
<td>61.46</td>
</tr>
<tr>
<td>Suturing _end of training</td>
<td>86.78</td>
<td>72.94</td>
<td>72.15</td>
<td>63.32</td>
</tr>
<tr>
<td>Suturing _retention</td>
<td>81.10</td>
<td>73.00</td>
<td>57.37</td>
<td>59.63</td>
</tr>
<tr>
<td>Total z-scores _baseline</td>
<td>77.63</td>
<td>72.37</td>
<td>73.29</td>
<td>72.30</td>
</tr>
<tr>
<td>Total z-scores _end of session II</td>
<td>83.25</td>
<td>76.58</td>
<td>69.00</td>
<td>66.70</td>
</tr>
<tr>
<td>Total z-scores _end of training</td>
<td>88.51</td>
<td>77.06</td>
<td>64.50</td>
<td>65.05</td>
</tr>
<tr>
<td>Total z-scores _retention</td>
<td>83.42</td>
<td>79.74</td>
<td>54.71</td>
<td>52.64</td>
</tr>
</tbody>
</table>
Main analysis: Laparoscopic tasks

For our main analysis, the mean ranks of the Kruskal-Wallis tests are shown in Table 1. If differences between groups were observed, an independent samples Mann-Whitney test was performed to compare individual groups with each other. Figures 1-4 show the results of these tests with asterisks to flag whenever the outcome of these Mann-Whitney tests proved significant. Figures 1-3 show median scores with 25 to 75 % inter-quartile ratio’s on each task at each session and serve as an indication of the different learning curves between the training groups. Note that the baseline of the suturing task is lower since only a smaller segment of the task was performed at baseline. This measure was only used for baseline between group analysis, not within-person progress.

At baseline, none of the groups showed significantly different levels of performance on any of laparoscopic tasks. At the end of session two, we observed a significant difference on the pipe cleaner task between the break and the break-nap group (p = 0.044), while the spaced group performed better at the beads task (p = 0.024), the suturing task (p = 0.012) and the total score (p = 0.044) as compared to the massed group. The spaced group also performed significantly better than the break group (p = 0.047) on suturing at the end of the second session. At the end of training both the spaced and break-nap group outperformed the massed group on the pipe cleaner (p = 0.044; p = 0.034, respectively) and the total score (p = 0.007; p = 0.009, respectively).

The spaced and break-nap group also showed a lower score on the suturing task compared to the massed group, but only borders close to significance for the break-nap group (p = 0.008; p = 0.056, respectively). At retention, both the spaced and break-nap condition outperformed the massed (p = 0.001; 0.001) and the break group (p = 0.005; p = 0.01, respectively) on the pipe cleaner task. On the beads task, the spaced group significantly outperformed both massed (p = 0.017) and break (p = 0.032) groups. On the suturing task, both the spaced and break-nap group outperform the massed group (p = 0.007; p = 0.005, respectively). Both the spaced and break-nap group have a significantly lower total score compared to the massed (p < 0.001; p = 0.001, respectively) and the break group (p = 0.002; p = 0.007, respectively).
Figure 1. Median completion times (in seconds) for the pipe cleaner task at baseline, at the end of the second session, at the end of training and at retention for all training groups (* = \( p < .05 \); ** = \( p < .01 \); *** = \( p < .001 \)).

Figure 2. Median completion times (in seconds) for the beads task at baseline, at the end of the second session, at the end of training and at retention for all training groups (* = \( p < .05 \); ** = \( p < .01 \); *** = \( p < .001 \)).
Figure 3. Median completion times (in seconds) for the intra-corporeal suturing task at baseline, at the end of the second session, at the end of training and at retention for all training groups (* = p < .05; ** = p < .01; *** = p < .001). Please note that only a segment of the suturing task was performed at baseline (see method section), hence the lower completion times.

Figure 4. Median total z-scores at baseline, at the end of the second session, at the end of training and at retention for all training groups (NS = non-significant; * = p < .05; ** = p < .01; *** = p < .001).
One week versus one day spacing comparison

In sum, the Mann-Whitney tests revealed substantial differences in completion times on the laparoscopic tasks between the spaced and massed group. Estimates of effect sizes ($r_{rb}$) for completion times at the end of the second session, at the end of training, and at retention are displayed in Table 2, along with the effect sizes found in a prior study that used a spacing interval of one week. Since the retention interval was different from the current study, only end of training comparisons have been made.

Sleep and laparoscopy measures

To differentiate between the effects of estimated sleep and mere rest (and muted sensory input), we assessed the relation between self-reported sleep at breaks, self-reported sleep the night prior to training, the results of the activity trackers at breaks, and performance on the laparoscopic tasks within the subsample of the break-nap group.

Non-parametric correlations showed coefficients ranging from $r = .040$ ($p = .813$) to $r = .465$ ($p = .004$) for self-reported sleep during the naps and performance on the subsequent laparoscopic tasks. These correlations were most prevalent for the end of training and retention measures, but also present at baseline. However, none of the laparoscopy measures were significantly correlated with the results of the activity trackers’ estimate of minutes spent asleep. No significant correlations were found between performance on the laparoscopic tasks and self-reported sleep the night prior to training.

Table 2. Effect sizes ($r_{rb}$) of the Mann-Whitney tests for the three training tasks. Effect sizes are only mentioned when tests were significant and when a comparison was viable (for the previous spacing study).

<table>
<thead>
<tr>
<th>Task</th>
<th>End of Session II</th>
<th>End of Training</th>
<th>Retention</th>
<th>spacing</th>
<th>massed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipe cleaner task</td>
<td>0.20</td>
<td>0.21</td>
<td>0.29</td>
<td>0.19</td>
<td>0.32</td>
</tr>
<tr>
<td>End of Training</td>
<td></td>
<td>0.65</td>
<td>0.23</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Retention</td>
<td></td>
<td></td>
<td>0.31</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>Suturing Task</td>
<td>0.20</td>
<td>0.26</td>
<td>0.31</td>
<td>0.29</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Means Rating Scale Mental Effort scores with paired samples t-tests assessing the progression in mental fatigue for all four training groups at baseline, the end of session II and at the end of training (no significant difference; ↑/↑ for significant increase/decrease, * = p < .05; ** = p < .01; *** = p < .001).

<table>
<thead>
<tr>
<th>Measure:</th>
<th>Mean_massed</th>
<th>Mean_break</th>
<th>Mean_break-nap</th>
<th>Mean_spaced</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSME_baseline</td>
<td>18.94</td>
<td>21.45</td>
<td>12.42</td>
<td>22.53</td>
</tr>
<tr>
<td>RSME_end of session II</td>
<td>33.81 ↑***</td>
<td>40.67 ↑***</td>
<td>35.43 ↑***</td>
<td>25.15 *</td>
</tr>
<tr>
<td>RSME_end of training</td>
<td>50.61 ↑***</td>
<td>48.50 *</td>
<td>48.30 ↑***</td>
<td>27.55 *</td>
</tr>
</tbody>
</table>

Rating Scale Mental Effort Analysis

No significant correlations were found between scores on the RSME and estimated time spent asleep at breaks (self-report and activity tracker). At baseline and at the end of session two, no effects of mental fatigue on laparoscopic performance were found. Fatigue at the end of session three did show significant correlations with the pipe cleaner (r = .192, p = 0.019), suturing (r = .206, p = 0.012) and total score (r = .219, p = 0.008) at the end of training and the suturing task at retention (r = .173, p = 0.046).

The RSME scores are shown in Table 3., along with the results of multiple paired samples t-tests illustrating the progression of the scores for all four training groups. A clear reduction in mental fatigue can be observed for the spaced group compared to the three other groups.

Discussion

The findings in this study further nuanced the current theory regarding the spacing effect as it applies to laparoscopic surgical training. We found that spacing three training sessions across three consecutive days is advantageous compared to a massed 1-day schedule, even if that schedule accounts for substantial amounts of breaks in training sessions. Alternatively, we found that including powernaps in between the training sessions on a 1-day schedule, can enhance long-term retention, an option which may be more beneficial in terms of logistics when organizing training events, when time constraints are a bigger motive.

The Effects of Naps

One of the more striking findings was that within the break-nap group, participants who reported spending more time asleep during the naps, had worse performance on the laparoscopic tasks at the end of training and retention. This seems counter-intuitive since
the break-nap group as a whole shows a pattern of better performance compared to the break group, the most similar condition that does not allow sleep during the break. A possible explanation is that the lower performance was caused by sleep inertia, a time period of impaired performance and disorientation as a person transitions from sleep to wakefulness (Miccoli, Versace, Koterle, Cavallero, 2008). We accounted for this possibility in the schedule of our sessions, since all performance measurements were done directly prior to the naps, with the naps being followed by a brief period to fill out a questionnaire and resuming training on a basic task in the next session. Still, sleep inertia may have effected trainees while practicing during the following session or the inertia effects may not have fully wore off by the time of the next measurement (Burke, Scheer, Ronda, Czeisler & Wright, 2015). The occurrence of sleep inertia can differ between contexts. The recommendation on the duration for daytime naps is to keep them below 30 minutes, as longer naps are associated with lower productivity and increased sleep inertia (Dhand & Sohal, 2006). However, when employees work in night shifts or have prolonged working hours, longer naps (40-60 minutes) show more performance benefits (Mulrine, Signal, Berg & Gander, 2012). Additionally, effects on motor performance can be influenced by whether a trainee habitually takes naps or not (Milner, Fogel & Cote, 2006).

Another explanation is that participants who slept (longer) within the break-nap group did so because they were more sleep deprived and sleepy to begin with, and this led to a suboptimal state for learning throughout the day of training (Miccoli et al., 2008), while the naps were not sufficient to decrease sleep deprivation and sleepiness. Contrary to this explanation, there was no association between sleep quantity and sleep quality on the night before the training and performance on the laparoscopic tasks (both within the break-nap group and the overall sample).

An alternative explanation is that within the break-nap group, trainees with a lower aptitude for learning laparoscopic skills experienced more arousal due to their difficulties in learning the tasks and reported being more fatigued and having spent more time asleep during the breaks as a misattribution of their arousal levels (Cotton, 1981). In any case, the findings should be interpreted cautiously since no correlations were found between the performance on the laparoscopic tasks and estimated sleep derived from the activity monitoring devices. Regardless of whether sleep during naps deteriorates performance, the results do indicate that trainees benefit from a period of rest while they are inactive, lying down, and while their sensory input is muted as compared to a traditional break.

**The Effects of Fatigue and Consolidation**

Even though participants in the massed group reported experiencing significantly more fatigue compared to the spaced group, we observed weak correlations between fatigue and performance on the laparoscopic tasks in the sample, which suggests a small influence
of fatigue contributing to the spacing effect. It is also worth considering that fatigue may have a higher impact on learning efficiency when a higher level of fatigue is reached, but that this threshold was simply never reached using the current design of training.

Learning of motor skills largely occurs in the timeframe following training (Walker, Brakefield, Morgan, Hobson & Stickgold, 2002), when there is opportunity for consolidation of the training stimuli. Consolidation occurs during ‘off-line’ periods simply as time passes, but is enhanced with a short nap or during overnight sleep (Brashers-Krug, Shadmehr & Bizzi, 1996; Doyon et al., 2009; Korman, Raz, Flash & Karni, 2003), although this may depend on the nature of the task (Doyon et al., 2009). Typically, effects of overnight sleep consolidation are enhanced for more difficult tasks (Kuriyama, Stickgold & Walker, 2004) and when a break is planned early in the training schedule rather than later (Duke, Allen, Cash & Simmons, 2009). Other authors have suggested that sleep effects can be attributed to the type of design used (sleep deprivation control groups that tend to impair performance) and that mere periods of rest show similar improvements as sleep (Rieth, Cai, McDevitt & Mednick, 2010). Regardless of whether sleep enhances consolidation or whether sleep deprivation impairs performance (or both), one can conclude that sleeping sufficiently is beneficial after training. The results of the current study also suggest that night sleep is more beneficial in between training sessions than just after the entire training, since the spaced group (who had two nights of sleep in between training sessions) had the best performance at the end of training and at retention, followed by break-nap group, the break group and the massed group, respectively.

When comparing the two spaced groups (the one in the current study and the one from the previous study, Spruit, Band & Hamming, 2014, chapter 7), we observed higher effect sizes for the intervention with the bigger spacing interval (one week). Future research should investigate what the optimal spacing interval is, although this is likely influenced by the desired retention interval (McDaniel, 2012). Finally, trainers ought to be cautious in the design of their training, since consolidation can also be disrupted during the acquisition (active training) phase when multiple conflicting training conditions are interleaved during training (Banai, Ortiz, Oppenheimer & Wright, 2010; Spruit, Kleijweg, Band & Hamming, 2016, chapter 6).

**Limitations and Conclusion**

One of the limitations of the study is that we did not include any measures of self-efficacy in our design. In the introduction we noted that spacing can enhance more accurate appraisal of the skill level of a trainee (by both trainee and instructors) and with the current design we were unable to estimate the extent to which this factor influenced the effectiveness of spacing. Also, we used a different retention interval from the previous study (Spruit,
Band & Hamming, 2014, chapter 7) so we only have comparisons of effects sizes for the end of training and not retention. The current study only used completion times, so no generalizations to other performance outcome measures (accuracy, instrument path length, force, etc.) can be made from these findings. Also, we used self-reports and activity monitoring devices to measure sleep, but these only serve as an approximation of time spent asleep. Future studies could make use of polysomnography to provide a more valid estimate on the influence of sleep consolidation on skill acquisition and retention of laparoscopic motor skills.

To answer the initial research question, we conclude that consolidation has the biggest influence in the effectiveness of spacing. Mitigating mental fatigue by inclusion of breaks and providing small periods of sensory muting are beneficial, but play a smaller role. When applying powernaps in training, one ought to be cautious for sleep inertia and it is recommended to schedule short (below 30 minutes, Dhand & Sohal, 2006) with ample rest time after waking to allow for waning of potential sleep inertia. Another way of preventing a poor learning state and poor consolidation post-training is to instruct trainees to sleep sufficiently before and after each training session. Finally, we found that spacing over three weeks leads to better learning efficiency and retention than spacing over three days.

Acknowledgements

We would like to acknowledge Gertjan Hultzer, Rene Rodenburg, Racheal Cheung, Joram van Ketel, Jolanda Hus, Tirza Loman, Remco Duijn, Enola van Maarsseveen and Frederique Arntz for their assistance in the LUMC skillslab.