At the heart of learning: navigating towards educational neuroscience in health professions education

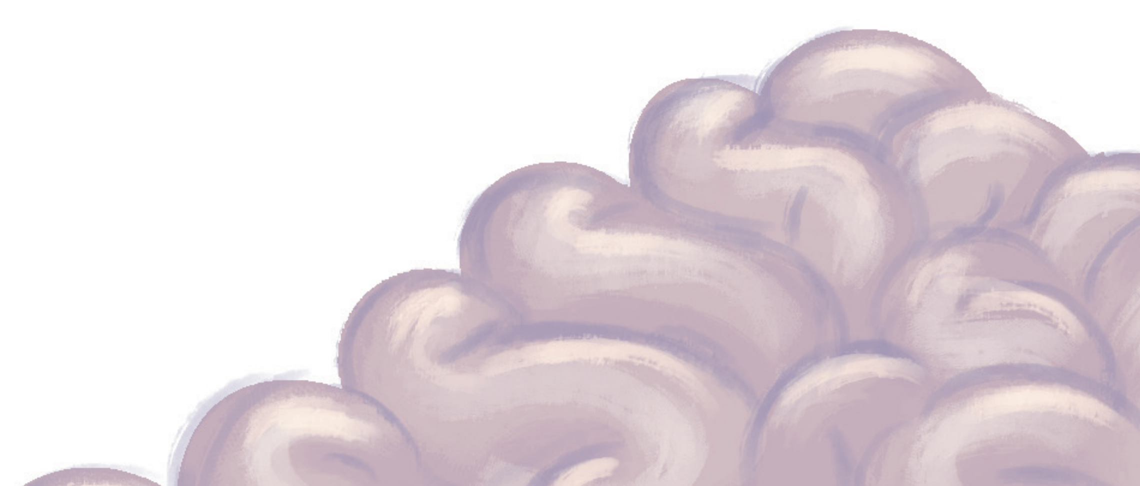
Versteeg, M.

Chapter 3

Making a lecture stick: the effect of spaced instruction on knowledge retention in medical education

Marnix C. J. Timmer · Paul Steendijk · Sandra M. Arend · Marjolein Versteeg

Published in: Medical Science Educator (2020), 30, 1211-1219.



Abstract

Introduction

Poor knowledge retention is a persistent problem among medical students. This challenging issue may be addressed by optimising frequently used instructional designs, such as lectures. Guided by neuroscientific literature we designed a spaced learning lecture in which the educator repeats the to-be-learned information using short temporal intervals. We investigated if this modified instructional design could enhance students' retention.

Materials and methods

Second-year medical students (N = 149) were randomly allocated to either the spaced lecture or the traditional lecture. The spaced lecture consisted of three 15-minute instructional periods, separated by 5-minute intervals. A short summary of the preceding information was provided after each interval. The traditional lecture encompassed the exact same information including the summary in the massed format, thus without the intervals. All students performed a baseline knowledge test two weeks prior to the lectures and students' knowledge retention was assessed eight days after the lectures.

Results

The average score on the retention test (α =.74) was not significantly different between the spaced lecture group (33.8±13.6%) and the traditional lecture group (31.8±12.9%) after controlling for students' baseline-test performance ($F(1,104) = 0.566$, p = .458). Students' narrative comments showed that the spaced lecture format was well received, and subjectively benefitted their attention span and cognitive engagement.

Discussion and Conclusion

We were unable to show increased knowledge retention after the spaced lecture compared to the traditional lecture. Based on these findings, we provide recommendations for further research. Ultimately, we aim for optimised spaced learning designs to facilitate learning in the medical curriculum and to help educate health professionals with a solid knowledge base.

Introduction

Medical students have a hard time recalling knowledge they acquire during medical training (Weggemans et al., 2017; Schneid et al., 2018; Simanton et al., 2012; Custers & ten Cate, 2011). Since successful clinical reasoning is built upon a solid foundation of knowledge, medical education is facing a serious problem (Herwaarden et al., 2009). This issue of forgetfulness may partly result from currently used teaching practices Namely, lectures in which a large volume of information is covered over a short uninterrupted time-span, so called massed learning, are still commonly used as a teaching modality by medical educators, especially in basic sciences education. This approach has shown to be rather ineffective when aiming for long-term knowledge retention (Rawson & Kintsch, 2005). One may consider adjusting such practices by using an alternative strategy, i.e. repeating information in several learning sessions distributed over time. This spaced learning approach is based on results from a century of psychological research (for a meta-analysis see Cepeda et al., 2006 or Carpenter et al., 2012) and could be a valuable addition to instructional designs in medical education.

The spacing effect is a robust phenomenon that forms the basis of spaced learning methods. During spaced learning, knowledge or skills that have to be acquired are repeated in several learning sessions that are distributed over time. Spaced learning is usually contrasted with massed learning where information is packed together in a single learning session and only repeated consecutively, if repeated at all. The beneficial effects of spaced learning on retention have been shown for a variety of learning tasks concerning factual knowledge, e.g. Cepeda et al., (2008), conceptual knowledge, e.g. Gluckman et al., (2014), Rohrer & Taylor, (2007), and procedural knowledge, e.g. Simmons, (2011).

Over the last ten years, research has proven spaced learning to be successful in various medical disciplines, including surgery, urology, radiology and general clinical reasoning (Gyorki et al., 2013; Kerfoot et al., 2007a; Smeds et al., 2016; Nkenke et al., 2012; Boettcher et al., 2018; Moulton et al., 2006; Patocka et al., 2015). Spaced learning has mostly been investigated in online learning environments and simulation settings. In surgical skill training, for example, a recent systematic review showed that spacing practice sessions resulted in increased retention of skills compared to massed training (Cecilio-Fernandes et al., 2018). Similar results have been found for spacing instructional designs. For instance, a study has shown that the dispersion of 4hrs of direct instruction over 4 weeks, i.e. 1hr/week, significantly enhanced knowledge retention after one month (Raman et al., 2010). Interestingly, neuroscientific research on mechanisms of memory implies that the spacing effect may already occur using much shorter intervals in the timescale of minutes to hours (for a review, see Smolen et al., 2016). This notion gave rise to our idea of implementing spaces within traditional massed 45-60 minute lectures to promote long-term knowledge storage among medical students.

Researchers in higher education have already reported the successful application of spaces with short intervals. Kelley and Whatson (2013) compared a 4-month biology course with a single 60-minute spaced learning session. Students following the spaced learning session repeatedly received an intensive 20-minute presentation (three times in total), intervened by 10-minute breaks. During the breaks they were asked to perform physical distractor activities, e.g. to clay, take a walk, or play basketball. These physical tasks were specifically selected to prevent any interference with the memory formation process regarding the learning material. Students' final test results were compared between groups per hour of education, and outcomes highly favoured the spaced learning cohort. This study showed that implementation of spaced learning in an instructional setting can establish long-term memory rapidly. Recent initiatives were inspired by these findings and illustrated the benefits of spaced instruction in different educational contexts (O'Hare et al., 2017; Garzia et al., 2016).

Educational initiatives promoting the use of short intervals to enhance long-term memory formation are inspired by neuroscientific evidence regarding the mechanisms of memory. An important phase in the process of long-term memory formation is the stabilization of a memory trace after the initial acquisition, referred to as consolidation (Kandel et al., 2014). Research has shown that consolidation of a memory on the molecular level, referred to as long-term potentiation (LTP), is elicited particularly by spaced trials and to a lesser extent by massed trials (Cao et al., 2014; Mauelshagen et al., 1998). A hypothesis has been formulated stating that adequately spaced stimuli overcome a refractory period that is needed for reinforcement of LTP (Smolen et al., 2016). In more detail, this refractory period may provide neurons time that is needed to synthesise molecular factors and/or facilitate feedback loops that underlie the initiation of LTP. In line with this reasoning, massed stimuli would not produce sufficient levels of molecular factors needed to support LTP even if the stimuli are repeated. Another hypothesis states that separate rounds of stimuli induce 'priming': meaning that LTP can be formed by a first stimulus, and is strengthened by a properly timed second stimulus (Smolen et al., 2016). Both hypotheses are not mutually exclusive and congruent with consolidation theory. Importantly, memory consolidation involves various molecular processes that each have their own temporal dynamics. Further research on long-term memory encoding in the brain has suggested that some specific molecular processes underlying LTP occur on the timescale of minutes and may contribute to the superiority of spaced learning (Genoux et al., 20020; Farah et al., 2009; Pagani et al., 2009; Ajay et al., 2004; Naqib et al., 2012; Xue et al., 2011; Menzel et al., 2011; Fields, 2005).

Based on previous educational experiments in higher education and the evidence derived from the neuroscientific framework, we aimed to examine the effect of short spaces on knowledge retention in medical students. Therefore, we compared a spaced lecture design with a traditional massed lecture and measured students' knowledge retention. We believe that the potential benefits of incorporating short spaces during teaching might help medical educators to make their lectures more effective.

Methods

Participants and setting

Second-year medical students enrolled in a course on disease mechanisms at the Leiden University Medical Center (LUMC) were invited to voluntarily participate in the study. More than 80% of contact hours in this course consists of lectures. The intervention was conducted in a lecture on the Dutch national vaccination program. In previous academic years information about the national vaccination program was covered by a self-study assignment. This topic was selected for this spaced learning study specifically, because students had received no prior formal education on this topic. The lectures were delivered as live presentations in a lecture hall, supported by a digital slideshow (Microsoft Powerpoint). This is common practice for lecturing at the LUMC.

Ethical considerations

Study participation was on a voluntary basis as the lectures were not mandatory. Students were notified that the supplied information was part of their exam material. Those who decided not to attend any session could still access the exam material using the existing self-study assignment. Students autonomously decided if their test results could be used for research purposes by signing the informed consent form prior to the baseline-test, and again prior to the retention-test. They were informed that data would be anonymised and that they could withdraw their consent at any given time. Moreover, they were ensured that the test results would not affect their course grades. Students did not receive any additional credit for their participation. The study protocol was reviewed and approved by the Educational Research Review Board of the LUMC: OEC/ERRB/20180612/2.

Study design

This was an experimental study for which two lectures were designed: an experimental lecture based on spaced learning principles, i.e. spaced lecture, and a control lecture using the traditional, massed approach, i.e. traditional lecture. Participants were randomly allocated to one of the lecture sessions. The spaced and traditional lecture were held consecutively on one day to facilitate that both lectures could be given by the same lecturer (SMA). The lectures were video recorded, to explore and reveal any substantial differences if suspected by the test results. The lecturer was a highly experienced teacher and is considered an expert in the field of infectious diseases. To assess students' knowledge at baseline, they were tested approximately two weeks prior to the intervention, i.e. baseline-test. The retention-test was taken 8 days after the intervention. For a detailed scheme of the study procedure, see Figure 1.

Intervention

The lectures were designed for this experiment specifically, serving as a substitute for the self-study assignment that was used during preceding academic years. The lecture material comprised characteristics of the diseases covered by the Dutch national vaccination program (Topic A), type and moment of vaccinations (Topic B) and a regional and international comparison of participation and program components (Topic C). Both lectures used the same supporting slides in identical order.

Spaced lecture

The total presentation-time of 45 minutes was divided over three instructional periods of approximately 15 minutes, separated by breaks, i.e. intervening gaps of 5 minutes, resulting in a 60-minute lecture. Each break was followed by a short rehearsal of the previously discussed information presented by the lecturer using 2-3 summary slides (Figure 1). We explicitly chose for summary slides as a passive rehearsal strategy, to be able for the lecturer to continue direct instruction after each break. In this way, we could study the effect of spacing on knowledge retention specifically, instead of inducing additional effects caused by active retrieval. The rationale for the 5-minute spaces was based on our interpretation of neuroscientific literature, similar educational implementation studies, and practical feasibility (Smolen et al., 2016; Kelley & Whatson, 2013; Menzel et al., 2001; Fields, 2005). Topics A, B and C were allocated to instructional periods one, two and three respectively. During the 5-minute breaks, students performed distractor activities (in our case three different origami tasks) that were not in any way related to information provided in the lecture, as this is considered to prevent possible cognitive interference with the memory formation process, conform the design of Kelley and Whatson (2013).

Traditional lecture

The traditional lecture followed the conventional setup for lectures in medical education at the LUMC, which is a 45-minute presentation without breaks. To control for potential confounders and ascertain same time on task, the traditional lecture also contained the summary slides, i.e. repetition, and the same amount of time dedicated to distractor activities (Figure 1). The latter were scheduled before and after the 'massed' instructional session.

Baseline-test and Retention-test

The baseline-test and retention-test consisted of short open-ended questions. The test questions were designed by the researchers and evaluated by two independent test-experts. The lecturer was not involved in designing these tests and was not

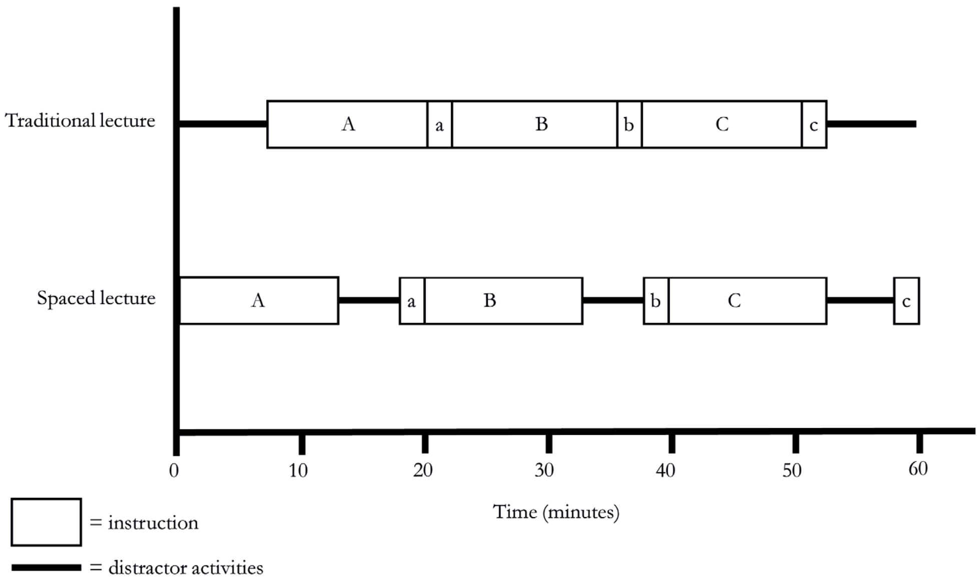


Figure 1 | Study design. The capital letters (A, B and C) represent the regular instructional phase on the specific topics. The small letters (a, b and c) represent short small summaries of the previous instruction block. In the experimental group, the regular instructional phase and small summaries were intervened by a 5-minute gap, where students were asked to perform an origami task. The traditional lecture was preceded and followed by an origami task, but the instructional phase lasted 45 consecutive minutes. Both lecture sessions ended after 60 minutes.

allowed to view the questions during the experiment. The tests were validated by performing a pilot-test with independent associates (three PhD candidates with a medical degree and two undergraduate medical students), which reassured a score of 100% could be obtained by deriving the correct answers from the lecture slides.

Participants were tested for recall of factual information that was covered by the lecture. The test consisted of short open-ended questions to reassure that recall was assessed rather than recognition. Students could obtain one point for each correct answer. If students were asked to mention two or more aspects or items in their answers to a specific question, e.g. "Which two human papilloma virus (HPV) serotypes are primarily targeted by the HPV-vaccination?", they were only awarded the point if their answer was completely correct. No penalty was given for incorrect answers. Using a pre-made answers key, students' answers were scored by one of the investigators, who was blinded for the lecture-condition. All answers considered eligible for discussion were discussed by two researchers until consensus was reached.

Baseline-test

The baseline-test was used to assess baseline knowledge regarding the topics covered in the lecture. The test consisted of 10 short open-ended questions. The test was performed in a lecture hall two weeks prior to the intervention to minimise priming effects.

Retention-test

Eight days after the intervention, students were invited to perform a retention-test. Students were requested not to study the lecture material between the lectures and the retention-test. The test included 30 short open-ended questions: the 10 questions of the baseline-test plus 20 new questions. Several additional questions were included to reveal students who had violated the study-protocol, e.g. "Did you study the lecture material between lecture and the retention-test? If yes, in what way?"

Data collection

All attendants at the baseline-test were linked to an anonymous study code. This study code was used to couple retention-test and baseline-test data. If students admitted they had studied or did not follow their allocated lecture, i.e. violated the study protocol, they were marked and excluded from further analyses.

Outcome measures

The primary outcome measure was students' performance on the full retention-test. Final scores were expressed as the percentage of the maximal score. Secondary outcome measures were: i) performance on the 20 new retention-test questions and on the 10 baseline-test questions included in the retention-test, ii) narrative comments of students and the lecturer on the used lecture-formats. Test and item analyses were conducted for assessment of internal consistency and item-characteristics (Supplementary E).

Narrative comments from the students and from the lecturer were gathered for qualitative assessment. To this end, students were actively encountered by the researchers immediately following the lectures and the retention-test, and they were encouraged to express their thoughts on the lecture format in an informal way. At that time, students were unaware whether they were part of the intervention or control group. Narrative comments for both sessions were noted and stored digitally afterwards. The lecturer was interviewed after both lectures were concluded.

Statistical analyses

For the power-analysis, the researchers agreed that a mean difference of at least one standard deviation should be detected, as this was regarded relevant for practice. Consequently, a minimum of 42 students (21 in each group) was needed to achieve 90% power at two-sided 5% significance.

The average total scores on the retention-test and the subscores for the 20 new questions and for the 10 repeated baseline-test questions were compared between study groups using ANCOVA tests, adjusting for students' baseline knowledge scores. Reliability and item characteristics (difficulty and distinctiveness) of the retention-test were evaluated by Cronbach's alpha, p-values and Rir-values (Supplementary E) (Berkel & Bax, 2006). Only data of students who completed both the baseline- and the retention-test were included for analysis. Those that did not follow their allocated lecture or restudied lecture material between the lecture and retention-test violated the study-protocol and consequently were marked and excluded from the analyses. Sensitivity analyses were carried out to reveal any major influence of these students on the outcome measures. If any of the test-questions should be removed for any cause, another sensitivity analysis for its effect on score differences would be carried out.

Results

A total of 344 students were enrolled in the course, of whom 171 students completed the baseline-test at the start of the course (Figure 2). Halfway through the course, 149 of these students attended the lectures. Half of the participating students followed the spaced lecture (*n* = 74) and the other half followed the traditional lecture (*n* = 75). One week after the lectures, 116 students completed the retention-test. 9 participants violated the study protocol, resulting in data of 107 students to be included for final analyses. On the retention-test, there was a higher participation rate among students who attended the spaced lecture (86.5%) compared to the traditional lecture (68.0%) attendees ($\chi^2(1, 149) = 7.228$, $p < .007$). However, further statistical analyses did not need to be adjusted since groups showed similar variance for all outcome measures.

Demographics

The mean age of participants was 19.3 ± 0.9 years (Table 1). Of these participants 90 (77.6%) were women, which resembles the overall gender distribution in the LUMC medical school. Both groups had similar average scores on the baseline-test (Spaced: 10.8% ± 8.8%, Traditional: 10.4% ± 8.8%).

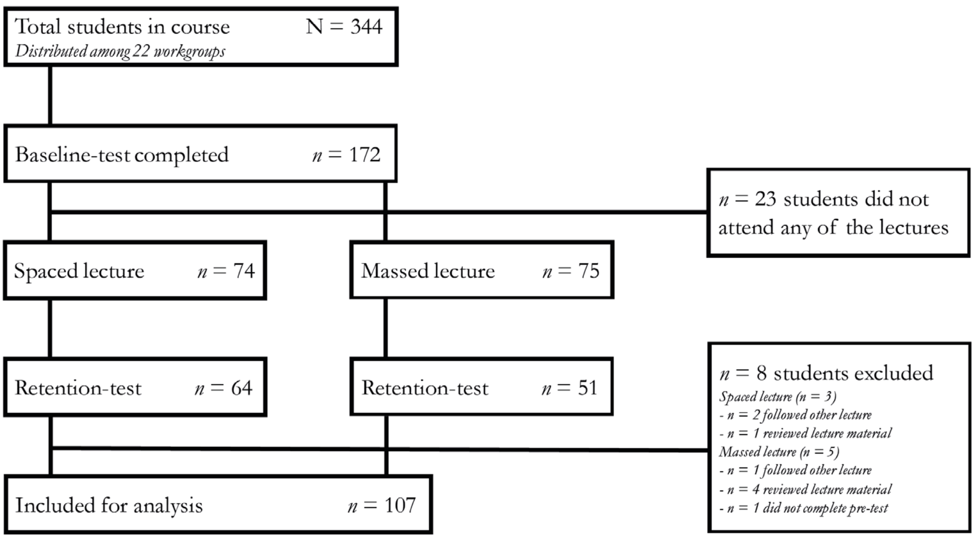


Figure 2 | Participant flow diagram.

Table 1 | Group characteristics.

Variable	Spaced lecture (n = 61)	Traditional lecture (n = 46)
Gender, female	68.9%	89.1%
Age, M (SD)	19.3 (0.9)	19.3 (0.9)
Pretest performance, M (SD)[1]	10.8% (8.8)	10.4% (10.1)

[1]*Maximum score was 90%, after exclusion of question 9.*

Retention

The spaced lecture group obtained a higher average score on the retention-test compared with the traditional lecture group (Spaced: 33.8% ± 13.6%, Traditional: 31.9% ± 12.9%), which, however, was not statistically significant ($F(1,104) = 0.566$, $p = .454$), also see Table 1. Separating the performance outcomes on the repeated original baseline-test items from the novel items revealed no significant differences between groups.

On one question of the retention-test the spaced lecture group outperformed the traditional lecture group (Spaced: 59.0% correct, Traditional: 2.2% correct). This difference likely resulted from extra information that was unintentionally provided by the lecturer during the spaced lecture. Consequently, we excluded this question in our analyses which resulted in a maximum score of 9 points on the baseline-test and 29 points on the retention-test, respectively. An alpha of 0.74 was obtained for the retention-test indicating acceptable internal consistency (see the Supplementary E for a full summary on test psychometrics).

Sensitivity analyses

Sensitivity analyses were conducted to investigate possible influences of the excluded data on the main results. Firstly, including participants that violated the study protocol in our analyses did not result in any significant differences versus the current reported results. Secondly, the mean difference on total retention-test scores between groups was higher when the aberrant item of the baseline-test and retention-test was not excluded, but remained not significant ($F(1,104) = 2.199$, $p = .141$). However, inclusion of this data significantly influenced the difference in retention-test scores regarding the subset of original baseline-test items ($F(1, 104) = 3.956$, $p = .049$).

Narrative comments

Generally, students who followed the spaced lecture responded positively towards the spaced lecture format. After the retention-test one student noted: "I really hope that we [spaced lecture group] did better on the test, as I would really like to do this more often." Several participants of the spaced lecture mentioned a positive effect of the intervening gaps on their attention and productivity. For example, a student said: "During those breaks I was totally distracted from the lecture material, resulting in a feeling of total reset. I really liked this as it enhanced my attention during the whole session." Others were more doubtful: "I am not sure if the breaks improved my attention because I had a hard time to reboot after each break, therefore missing most of the small summaries." The origami task was emphasised as an enjoyable distractor activity. "Normally, everyone grabs for his or her smartphone during occasions like this [the breaks]. Now, everyone started the break trying to complete the origami task, and only switched to their phones when they gave up or failed." Another student marked the negative side effect of this particular task: "...at sudden moments I was too busy on thinking of what the next origami model would be than on the actual lecture content".

Lastly, there were some comments on the intensity of the spaced lecture: "I found it rather intense, I would hate to think of doing this four times in a row, but I could imagine it being preferable for a revision lecture."

Students who followed the traditional lecture had some positive comments on the structure of the lecture but they generally agreed that they did not notice any difference with a normal lecture. On the question: "did you experience any differences in the lecture apart from starting and ending with the origami task?",

Table 2 | Test results for spaced lecture and traditional lecture cohorts on the retention-test.

	Spaced lecture *M (SD)*	Traditional lecture *M (SD)*	Difference (95% CI)	F	*p*	Effect size[1] (η^2)
Retention-test score (%)	33.8 (13.6)	31.8 (12.9)	2.0 (-3.1;7.2)	0.566	.454	.005
Retention-test score on 20 new questions (%)	36.6 (14.9)	33.6 (13.2)	3.0 (-2.5;8.5)	1.108	.295	.011
Retention-test score on 9 baseline-test questions (%)	27.7 (14.7)	27.8 (15.7)	-0.1 (-6.0;5.8)	0.007	.934	.000

[1]*Partial η^2 effect size as predicted by the ANCOVA model*

one student answered: "hmm... well, to be honest, no."

The lecturer enjoyed the spaced lecture format as she could recover during the breaks and experienced less fatigue afterwards. She noted: "I really had the idea that students' attention on the last part [of the lecture] was higher than normal."

Discussion

This study investigated the effect of spaced learning during a lecture on medical students' knowledge retention. We hypothesised that incorporating short spaces in the lecture would increase its effectiveness. We used an experimental design, comparing a spaced lecture with a traditional massed lecture. Our results showed that the effect of both lecture formats was not significantly different. Notably, the positive narrative comments indicated that the spaced lecture format was generally well-received by students.

In our study, we incorporated short 5-minute gaps between instruction sessions in a lecture to enhance the memory formation process. However, beneficial effects on knowledge retention were not found, suggesting that 5-minute spaces might have been too short to stimulate the consolidation process. They may have been insufficient to overcome the refractory period needed for stabilization of the memory trace, for example. Apparently, 10-minute spaces seemed to be more effective as Kelley and Whatson (2013) were successful with their spaced learning strategy in the classroom where they incorporated short 10-minute gaps. However, one should be careful interpreting these results, since findings can be highly dependent on the study design . For instance, in our study we measured knowledge retention at 8 days, whereas Kelley and Whatson measured it at 5 days. The 8-day period was selected since the Ebbinghaus forgetting curve indicates that forgetting declines exponentially and most of the forgetting occurs in the first week after initial learning (Ebbinghaus, 1885; Murre & Dros, 2015). It might be that a shorter retention period, i.e. less than eight days, had raised the ability to reveal differences between the study-groups. This is in line with evidence indicating that short intervening gaps potentially promote advantages on shorter retention periods (Cepeda et al., 2006; Cepeda et al., 2008; Cepeda et al., 2009). Another notable difference is that we chose to incorporate small summaries of preceding information in our lectures, whereas Kelley & Whatson repeated their 15-minute instructional blocks three times. Our rationale for this design was that it was closer to the traditional teaching style and was expected to be an easy-to-incorporate tool for medical educators if it was found to be effective.

Despite some empirical evidence including our own study, researchers acknowledge that optimal spacing protocols for humans remain unknown and we may need to gain more fundamental knowledge of the mechanisms of memory formation in order to develop these protocols (Smolen et al., 2016). Our study contributes by informing the research community that 5-minute spaces in a lecture setting seem insufficient to promote knowledge retention, and that other approaches should be investigated to develop optimal spacing formats. In future studies on spaced learning during instruction, one may specifically investigate the influence of (1) the duration of spaces (2) the number of spaces (3) in relation to the duration of the retention gap. Furthermore, one should be specific about the characteristics of the setting in which the study was performed, to determine if findings can be generalizable across educational contexts. Finally, future research may combine spaced learning with other effective learning strategies such as retrieval practice and/or test-enhanced learning to further promote knowledge retention in medical education (Roediger & Butler, 2011; Karpicke & Roediger, 2008; Ayyub & Mahboob, 2017).

Strengths and limitations

Our experiment was embedded in an obligatory course of the medical curriculum, so some practical limitations should be noted. For sake of time and anonymity, we did not register attendees at the time of lecture. Consequently, we were unable to question students who were absent on the retention-test about their reasons for a no-show. The higher drop-out rate in the control group thus remains unexplained. Furthermore, some sort of testing effect is inherent to our pre- post-test design. We aimed to minimise the testing effect by incorporating a two-week gap between the baseline-test and the intervention, and by including new questions in our retention-test. Specific strengths of the study design should also be delineated. First, the protocol included reliable tests in which we did not observe any floor or ceiling effects. We showed an increase in overall test scores from 10% to over ~30%. However, it is hard to contextualise this result as previous literature on retention following a lecture is heterogenous, with a high variability of the moment of delayed testing, e.g. one week (Mahler et al., 2011), 4-months (Giles et al., 1982), 5-months (Razzel & Weinman, 1977), type of testing, and restudying opportunities, e.g. summarise, note-taking or self-questioning (King, 1992). Second, the analyses included narrative comments which indicated that our new spaced lecture format was well-received. The majority of students and the lecturer noted that this format increased their attention and engagement and improved their productivity. It would be interesting to investigate whether this experience of enhanced attention could be quantified, using any approach that previously assessed mind wandering and its effect on retention in a lecture context (Farley et al., 2013; Lindquist & McLean, 2011).

Conclusion

Our findings showed that a spaced lecture did not enhance knowledge retention in medical students compared with a traditional, massed lecture. However, positive narrative comments indicated that this new spaced lecture format was generally well-received by students and the lecturer. Additionally, a theoretical and practical elaboration on our findings resulted in recommendations for educators and researchers on how to implement and study future spaced learning projects.