



Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Empirical article

Implementing Distributed Practice in Statistics Courses: Benefits for Retention and Transfer

Mirjam Ebersbach* and Katharina Barzagar Nazari

University of Kassel, Germany

The present study investigated the effect of distributed versus crammed practice before a course deadline on the retention and transfer of knowledge, and whether learner characteristics moderate the effect. In Experiment 1, only 41% ($N = 38$) of the initially enrolled students worked the voluntary but recommended practice tasks. Moreover, markedly fewer students did so in the distributed condition (12%) than the crammed practice condition (29%). In Experiment 2, working the practice tasks was mandatory and more students completed them ($N = 105$, i.e., 81%). Students who distributed practice clearly outperformed students who crammed practice on tests of knowledge retention and transfer five weeks after the practice deadline. No moderating effects of learner characteristics emerged. The study shows that distributed practice following knowledge acquisition is a powerful learning tool for fostering long-term retention and transfer with adults in authentic educational contexts.

General Audience Summary

Laboratory studies showed that distributing learning or practice time across multiple sessions (compared to practicing in only one session in a crammed or massed fashion) enhances memory performance. We investigated the effect of distributed practice in a field experiment at the university. After having acquired statistics skills in lectures, students were encouraged to practice these skills at home with a fixed number of practice tasks either distributed on three different days (with a gap of two and five days in between, respectively), or crammed on one day. In the first study, practice at home was recommended to the students but voluntary. As a result, only few students completed these practice tasks and less did so in the distributed condition than in the crammed condition—even though all students had been reminded to do so via email. In the second study, practice was mandatory for successfully completing the course, and most students completed the tasks. Similar as in laboratory studies, students in the distributed practice condition showed a better memory performance, tested after five weeks, than students in the crammed practice condition. The positive effect emerged not only for previously practiced skills but also in new tasks. The results suggest that distributing the practice of statistics skills can be recommended to university students and teachers—at least when memory performance is tested after a longer delay.

Keywords: Distributed practice, Statistics, Spacing, Transfer, University course, Long-term retention

Retention can be enhanced by distributing the time spent learning a certain piece of content across multiple sessions instead of cramming the learning of this content into only one session for the same total amount of time (i.e., spacing versus massing learning). The distribution of time spent learning can

be beneficial when it occurs during the acquisition of knowledge through repeated study or the practice of newly acquired knowledge or skills. The present study is concerned with the latter, and thus the term *distributed practice* is used in the following even though the reviewed literature includes studies on

Author Note

* Correspondence regarding this article should be addressed to Mirjam Ebersbach, University of Kassel, Department of Psychology, Hollaendische Str. 36-38, D-34127 Kassel, Germany.

E-mail address: mirjam.ebersbach@uni-kassel.de (M. Ebersbach).

distributed study as well. The benefit of distributed practice to retention appears to be a robust effect when investigated in the laboratory with rather simple, discrete content or skills, such as vocabulary, word lists, or pictures (e.g., Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1988, 1996; Donovan & Radosevich, 1999; Hintzman & Rogers, 1973; Kang, 2016). However, many open questions remain about the extent to which the benefits of distributed practice generalize more broadly. For example, only a few studies have investigated the benefit of distributed practice with more complex, interconnected content or skills, such as science concepts (Butler, Marsh, Slavinsky, & Baraniuk, 2014; Vlach & Sandhofer, 2012), mathematical procedures (Barzagar Nazari & Ebersbach, 2019a; Hopkins, Lyle, Hieb, & Ralston, 2016; Rohrer & Taylor, 2006, 2007), and comprehending a written text in a foreign language (Namaziandost, Rahimi Esfahani, & Hashemifardnia, 2018).

Given that most previous studies have been conducted in the laboratory or similar highly controlled contexts, one open question is about whether the benefits of distributed practice are reliably obtained with curriculum-relevant material in authentic educational contexts, which are less controllable in terms of potential confounding factors, such as students' learning activities outside of the manipulated learning conditions (e.g., Svihla, Wester, & Linn, 2018). Some studies suggest that the effects obtained in the laboratory do generalize to authentic educational contexts. For instance, one study found that distributing science lessons in elementary school enhanced the retention and transfer of science knowledge on a final test after a one-week delay (Gluckman, Vlach, & Sandhofer, 2014), and another study found a similar effect among university students (Kapler, Weston, & Wiseheart, 2015). Hopkins et al. (2016) assessed the effect of distributed practice in an "Introductory Calculus for Engineers" course. They used quizzes with questions addressing the course learning objectives and presented them in either a massed or distributed fashion. Distributing the quizzes enhanced short- and long-term performance on subsequent exams addressing these target objectives, even though some results only approached significance. However, an important consideration is that the procedure used in this study was quite complex due to the realistic educational context, including students' self-regulated execution of practice tasks and other quizzes ensuring that students had achieved a certain knowledge level. In addition, students were repeatedly tested on the same material via three exams, which might have produced a testing effect (Carpenter, Pashler, & Vul, 2006).

In contrast with the few studies that have demonstrated a benefit of distributed practice in authentic educational contexts with curriculum-relevant material, other studies have found mixed results. In one study, when third graders' practice of mathematical procedures was distributed, their performance was enhanced on a test after one week but not on another test after six weeks (Barzagar Nazari & Ebersbach, 2019a). Furthermore, university students practicing the solving of permutation tasks in either a massed or distributed manner in the laboratory did benefit in one study from the distribution in a test after five weeks but not after one week (Rohrer & Taylor, 2006). In addition, another study

using the same material found no benefit of distributed practice after one week and five weeks (Ebersbach & Barzagar Nazari, 2020). Given the paucity of research in authentic educational contexts and the mixed findings, further research is needed to better understand the effects of distributed practice in authentic educational contexts with curriculum-relevant content, and the factors that influence whether or not a benefit is obtained.

Most studies on distributed practice to date have assessed the retention of content, and thus another open question is whether distributed practice promotes transfer of learning (e.g., applying knowledge to new tasks, forming abstractions and generalizations, or making new inferences). Some evidence of a positive effect of distributed practice on transfer comes from studies with children (e.g., Gluckman et al., 2014). In particular, studies that use delayed tests often find a benefit of distributed practice on tasks that assess generalization performance, such as concerning novel nouns in toddlers (Vlach, Ankowski, & Sandhofer, 2012), science concepts in 5- to 7-year-olds (Vlach & Sandhofer, 2012), or category induction in 3-year-olds (Vlach, Sandhofer, & Kornell, 2008; see also Vlach et al., 2015; for a review concerning the effect of distributed practice in children's generalization performance, see (Vlach et al., 2014)). In adults, studies investigating the effect of distributed practice on transfer are scarce (Smith & Scarf, 2017). One exception is a study that found that distributed practice enhances the transfer of previously practiced surgery skills from models to real rats (Moulton et al., 2006). In another study (Kapler et al., 2015), undergraduate students attended a simulated university lecture referring to natural science and were asked to review the content one or eight days after this lecture. On a final test after five weeks, students who reviewed the material eight days after the lecture outperformed students who reviewed the material one day later, and this effect emerged for both factual and transfer knowledge. However, a pure massed condition in terms of reviewing the material without a delay, immediately after the lecture, was not implemented in this study.

Another important question concerning distributed practice is whether learner characteristics potentially moderate its effect. This question has been largely neglected in previous research, particularly in studies taking place in authentic educational settings. Before issuing recommendations for teachers to integrate distributed practice into their teaching, one needs to know whether all learners benefit from it or only subgroups of learners. One study suggested that learners with a medium performance level might profit more from distributed practice than learners on a low- or high-performance level (Barzagar Nazari & Ebersbach, 2019b; see also Hirsch, Kapoor, & Laing, 1982). Another potential moderator might be performance-avoidance goals (e.g., Elliott & Dweck, 1988; Nicholls, 1984): Learners with marked performance-avoidance goals might be particularly challenged by the difficulty inherent in distributed practice, that is, to recall the once acquired skills in a later session. They might therefore be less motivated to meet this challenge compared to learners with weaker performance-avoidance goals. The same applies to learners scoring high in work avoidance (Nicholls, Cobb, Wood, Yackel, & Patashnick, 1990). A marked mathematical self-efficacy and effort motivation of the learners, on

the other hand, may help to overcome these additional obstacles inherent in distributed practice. Furthermore, learners with difficulties concentrating might benefit more from the distribution, compared to massed practice, because they only have concentrate for a shorter time due to multiple practice sessions that are shorter in duration (cf. Craik & Lockhart, 1972; Hintzman, 1974).

The Present Study

The present study addresses the previously introduced open questions by investigating the effects of distributed practice on the retention and transfer of knowledge in an authentic educational setting (i.e., following university lectures) with curriculum-relevant content (i.e., statistics) and taking learner characteristics as potential moderators into account. In contrast to many previous studies, the content was acquired first by all students in the same manner and only the subsequent practice phase was scheduled in a massed or distributed fashion. In addition, the present paradigm departs from the typical distributed practice procedure in that practice did not follow immediately after the knowledge acquisition; instead, practice was distributed up until a deadline for completion or massed immediately before that same deadline. In order to avoid confusion with prior work on this topic, we will refer to the massed practice condition as *crammed practice* instead (see Method section). The results were expected to reveal whether practice following knowledge acquisition should optimally be realized in a distributed or crammed manner when learners and teachers aim for long-term retention.

The distribution of practice can be realized in different ways. Usually, the temporal intervals between the single practice sessions are constant (see Cepeda et al., 2006). However, constant intervals might be too long, and thus impair successful retrieval, enhancing instead the retrieval of erroneous information. Furthermore, too short intervals can make retrieval too easy and thus produce no benefit of retrieval practice (Storm, Bjork, & Storm, 2010). Therefore, the effect of expanding intervals (i.e., prolonged periods between at least three practice sessions) and contracting intervals (i.e., shortened periods) was investigated. The optimal schedule of distributed practice depends on the retention interval: For shorter retention intervals up to seven days, a contracting interval yielded the largest effect, whereas for longer retention intervals, constant and expanding intervals were more beneficial (Küpper-Tetzl, Kapler, & Wiseheart, 2014; for a similar finding with children, see Vlach et al., 2010). Other studies suggested a larger benefit of expanding intervals for short retention intervals, too (e.g., Karpicke & Roediger, 2007). Toppino, Phelan, and Gerbier (2018) reported an advantage of an expanding interval for learners in a low training condition compared to learners in a high training condition. Storm et al. (2010) specified that an expanding interval might be beneficial when the learning content is susceptible to forgetting. In the present study with a retention interval of several weeks, we implemented an expanding interval in the distributed practice condition in line with Küpper-Tetzl et al. (2014).

We hypothesized that distributing practice in the frame of statistics lectures would foster students' retrieval of the practiced

contents as well as their transfer of knowledge, both examined in a long-term retention test. In addition, we assumed that the effect of the practice condition would be larger for learners on a medium performance level, with low performance- and work-avoidance goals, high mathematical self-efficacy and effort motivation, and more difficulties with concentrating.

Experiment 1

Method

Participants. Participants were recruited in the context of a statistics course (i.e., Statistics I) of the local Faculty of Social Sciences that was attended by about 250 students. Most students in this course were studying sociology or political science. To enroll for the study, students were provided at the beginning of the course with an online link, where they could enter a valid e-mail address and complete a short questionnaire (see Material and Procedure). In addition, their last grade in mathematics was collected via self-report as an indicator of their general prior math skills. They were then assigned to the two practice conditions by means of a randomized block design, taking into account their last math grade. Thus, on average, math grades were roughly equated across the two conditions. All students were informed that completing the practice tasks would support their performance in the final exam. Participation in the practice tasks was voluntarily and could be terminated at any time. In addition, attendance in the lectures was not mandatory at the university where the study took place. Initially, 94 students enrolled for the study, but only 62 took part actively in the experiment by providing any data during the practice or test phases. Of these 62 students, 35 studied sociology, 21 political science, and six other subjects. Thirty-three were assigned to the crammed practice condition (18 women, 15 men; $M_{age} = 23.2$ years, $SD = 3.2$ years) and 29 to the distributed practice condition (16 women, 11 men, 2 not specified; $M_{age} = 24.4$ years, $SD = 4.3$ years).

Design. The experiment followed a one-factorial design with *practice condition* (crammed versus distributed) manipulated between subjects. The dependent variable was the performance in a final test including tasks that assessed the retrieval of the previously acquired and practiced knowledge.

Material and procedure. For a schematic depiction of the procedure, see Figure 1. After enrolling for the study, students were provided in the pre-phase with a link directing them to an electronic questionnaire assessing their performance-avoidance goals and their work avoidance, each by eight items of the German SELLMO-ST (Spinath, Stiensmeier-Pelster, Schöne, & Dickhäuser, 2012), their mathematical self-efficacy by seven items of the German Academic Self-Efficacy Scale for School Children (Jerusalem & Satow, 1999), adapted to statistics, their effort motivation by eight items of the German LIST (Wild & Schiefele, 1994), and their self-rated difficulties to concentrate by six items of the German LIST (Wild & Schiefele, 1994). Different from the original LIST scale, we used an answer scale ranging from 1 (Does not apply at all) to 6 (Does apply completely) (see Boerner, Seeber, Keller, & Beinborn, 2005).

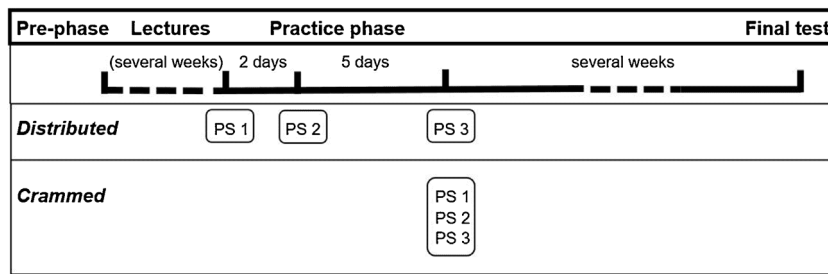


Figure 1. Schematic depiction of the procedure in the distributed and crammed condition in both experiments. The delay between the last practice set and the test was two and eight weeks in Experiment 1 and five weeks in Experiment 2. PS = Practice set.

In addition, students' last grade in mathematics in school was collected via self-report.

In the following weeks, all students attended a regular statistics course, comprising seven lectures referring to the determination and interpretation of simple descriptive statistics. On the day of the last lecture addressing this topic, students in the distributed practice condition received an email with a link and the prompt to complete the first practice set, including four practice tasks online on their own within 48 h. This period resembles typical homework periods (see also Hopkins et al., 2016) and was allotted to assure that these students had sufficient time to work the practice tasks. The practice tasks were provided by means of the online survey tool "formr" (Arslan, Tata, & Walther, 2018). Two days later, students in the distributed practice condition received another e-mail with a link and prompt to complete the second practice set including another four practice tasks online. Five days later (i.e., seven days after the first email), a link to the third practice set including the last four practice tasks was provided. The distributed practice followed an expanding interval between the practice sessions (cf. Küpper-Tetzel et al., 2014).

Students in the crammed condition received only one email seven days after the last lecture with a link and the prompt to complete all twelve practice tasks at once online within 48 h. This temporal gap between the last lecture and the request to solve the practice tasks was implemented in the crammed practice group to ensure that (a) the interval between the last practice set and the final test was the same in both conditions, and that (b) the final test could take place at the same time in both conditions, ruling out differences in students' statistics knowledge meanwhile acquired in the subsequent lectures. Previous research has shown that the retention interval (i.e., the interval between the practice phase and the test phase) is a critical factor for the effect of distributed practice (e.g., Cepeda et al., 2009; Küpper-Tetzel & Erdfelder, 2012). Thus, students in the distributed condition received the e-mail with the prompt to complete the last practice set on the same day as students in the crammed condition (with the prompt to complete all practice sets at once). Given that this procedure differs from how many other studies operationalized massed practice (i.e., in one session immediately after the knowledge acquisition, without delay), the term "crammed practice" is used.

All practice tasks, including several subtasks, were structurally parallelized and involved simple descriptive statistics, in line with the contents of the lectures (i.e., determining and interpreting frequencies, medians, quartiles, information about

normal distributions, confidence intervals; for an example, see Appendix A). Each practice task addressed the same compilation of contents in the subtasks; the practice tasks only differed regarding the contexts and the numerical values provided. Students received no individual feedback, but after completing a practice task, a sample solution was presented. Two weeks and eight weeks after the last practice set, respectively, tests took place online that had to be completed without additional help on an electronic device. The tests were announced and presented as additional practice sessions to refrain students from additional rehearsal. Participants were again provided with the link via email. Even though we could not control whether the students used additional help for the tests due to the online testing, it can be assumed that such behavior would occur similarly often in both conditions. The tests aimed at assessing short- and long-term retention, addressing previously practiced knowledge, that is, the four test tasks had the same structure as the practice tasks but were embedded in other contexts and contained different numbers (for an example, see Appendix B). The percentage of correct responses in the final tests served as dependent variable (i.e., $[100 \times \text{achieved score}] / \text{total attainable score}$). All students were informed that the study included multiple sessions (but not the exact number) that may differ regarding their timing between students.

Data preparation. It was planned to include only those students in the analyses for testing the hypotheses who had been working all practice tasks and all test tasks. Therefore, they should have had at least tried to complete the tasks. If students did not know an answer, they could leave the response fields empty and click to see the next task. Only those students who abandoned the practice or test session by not clicking further were excluded. Students' performance in the practice tasks and the final tests was scored by two independent raters based on a predefined, detailed coding scheme. It listed the correct solutions for each test task as well as possible analogue expressions (e.g., $1/5$ or 0.2), and the corresponding scores. Thereafter, the two raters compared their scores and discussed the open differences. All differences could be resolved by the raters in that they could agree on the same scoring without any doubts. To ensure the reliability of this final rating, a third rater scored about 25% of the answers independently as well. The final ratings of the first two raters were largely in line with the ratings of the third (control) rater. Therefore, the final common score of the two first raters was used.

Results and Discussion

Of the 62 actively participating students (out of 94 initially enrolled students), only 26 worked all tasks in the three practice sets and all tasks in the two tests (i.e., 41% of the actively participating students; thereof 12% or $n = 8$ in the distributed practice condition and 29% or $n = 18$ in the crammed practice condition). This level of participation is a clear indicator that students do not voluntarily use practice in general and distributed practice in particular to improve their memory for the knowledge acquired during their statistics course. Given that the final sample sizes were too small, the initially planned analyses testing the effect of distributed practice could not be computed.

Instead, we examined the participation of the students in both practice conditions in more detail. Table 1 shows the number of students who completed all tasks of each single practice set and of each of the two tests. A substantially smaller percentage of students in the distributed condition completed all tasks in each single practice session and in the tests compared to the students in the crammed condition. A Bayesian contingency test (see Table 1) revealed moderate evidence (i.e., $BF_{10} = 4-8$) for this difference in each of the three practice sessions and the first test, and anecdotal evidence (i.e., $BF_{10} = 2$; Lee & Wagenmakers, 2013) in the second test. Thus, already in the first practice set, a larger proportion of students in the crammed condition worked the practice tasks compared to students in the distributed condition.

In addition, we analyzed whether the two practice conditions differed concerning students' continuance in the study. In the distributed practice condition, of the 20 students who completed practice set 1, only 15 (i.e. 52% of the initial sub-sample) completed practice set 2 too, and of those, only 13 students (i.e., 45%) also completed practice set 3. In the crammed practice condition, these proportions were identical to those in Table 1 (i.e., $n = 31, 29,$ and 29). A Friedman test revealed a clear drop across the practice sessions in the distributed condition, $\chi^2(2) = 11.14, p = 0.004, n = 29$, while no such drop emerged in the crammed condition, $\chi^2(2) = 4.00, p = 0.14, n = 33$.

Thus, when students were not directly or indirectly forced to practice the knowledge previously acquired in a lecture, they hardly did so—and they did it even less frequently and less continuously when they had to practice in a distributed manner compared to a crammed manner. This finding puts into question whether an unsupervised implementation of distributed practice into students' self-regulated learning could be successful (see also Barzagar Nazari & Ebersbach, 2018). Experiment 2 was conducted to follow up on Experiment 1 and investigated the effect of distributed practice on the retention and transfer of knowledge, but completing the practice tasks was mandatory to avoid massive drop-out.

Experiment 2

Method

Participants. The sample that was recruited for Experiment 2 was largely identical to Experiment 1: The students were now attending the Statistics II course, including advanced statistics,

which took place in the subsequent term. To enroll for the study, students were again provided at the beginning of the new course with a link, where they could enter a valid e-mail-address and complete a short questionnaire (see Material and Procedure of Experiment 1). In contrast to Experiment 1, the assignment to the two practice conditions was based on students' Statistics I grade instead of their math grade, and only one test took place five weeks after the last practice session to avoid a testing effect (e.g., Carpenter et al., 2006). Initially, 129 students enrolled for the study; 69 were assigned to the crammed condition and 60 to the distributed condition. The final sample, including only those students who had worked all of the practice tasks and retention tests, consisted of $N = 105$ students (i.e., 81% of all enrolled students; 85 studied sociology, 19 political sciences, 1 did not indicate study). Of the students included in the final sample, 64 were in the crammed practice condition (35 women, 29 men; $M_{age} = 23.2$ years, $SD = 2.6$ years) and 41 were in the distributed practice condition (30 women, 8 men, 3 not specified; $M_{age} = 24.1$ years, $SD = 4.2$ years). The unequal distribution of students to the two conditions is again a result of self-selection processes in terms of less students in the distributed condition having worked all of the practice tasks and retention tests.

Design. Like Experiment 1, Experiment 2 used a one-factorial design with *practice condition* (crammed versus distributed) manipulated between subjects. The dependent variable was the performance score (percentage of correct responses) in the final test assessing the retention of the previously acquired and practiced knowledge as well as students' ability to generalize this knowledge in terms of transfer.

Material and procedure. The instruments assessing the learner characteristics, the procedure as well as the data preparation were the same as in Experiment 1 (see Figure 1), with the following exceptions: (a) the learning content involved the determination and interpretation of inferential statistics (i.e., linear regressions), (b) practice was mandatory as part of the study (for an example of a practice task, see Appendix C), (c) a sole final test took place five weeks after the last practice session, in which retention and (d) transfer performance were assessed. Retention was assessed by four test tasks including structurally similar tasks as the practice tasks that were embedded in other contexts and included different numbers (for an example, see Appendix D). The test tasks assessing transfer were designed as nine questions, assessing the more general understanding of the subject, to be answered with true or false (e.g., "A disadvantage of the beta coefficient is that it can only be interpreted in relation of the corresponding scale" or "To identify in a linear regression model the independent variable with the largest effect on the dependent variable, the significant beta coefficients have to be compared"). The scoring of the solutions in the test tasks was executed as in Experiment 1, with a predefined scoring scheme including correct responses and potential alternative expressions and the corresponding scores. Two raters scored the responses independently and thereafter solved discrepancies. The final score of the two raters corresponded well with the score of a third rater who independently rated about 25% of

Table 1

Number of active participants ($N = 62$) who completed all tasks of each single practice set and of each test.

Condition	Practice set 1 n (%)	Practice set 2 n (%)	Practice set 3 n (%)	Test 1 n (%)	Test 2 n (%)
Crammed	31 (94)	29 (88)	29 (88)	26 (79)	18 (55)
Distributed	20 (69)	18 (62)	17 (59)	15 (52)	9 (31)
BF ₁₀	6 [0.38, 3.35]	4 [0.22, 2.69]	8 [0.37, 2.82]	3 [0.13, 2.32]	2 [-0.05, 1.98]

Note. Crammed condition: $n = 33$, Distributed condition: $n = 29$; BF₁₀: Bayes factor indicating differences between both learning conditions concerning the number of active participants with 95% credible interval. BF₁₀ = 1 to 3: anecdotal evidence for H1; BF₁₀ = 3 to 10: moderate evidence for H1 (Lee & Wagenmakers, 2013).

Table 2

Descriptive statistics of the performance in the three practice sets (percent correct), separate for each practice condition.

Condition	Practice set 1 $M(SD)$	Practice set 2 $M(SD)$	Practice set 3 $M(SD)$
Crammed	32 (15.7)	36 (17.3)	40 (19.1)
Distributed	30 (13.2)	40 (21.1)	41 (17.6)

Note. Crammed condition: $n = 64$, Distributed condition: $n = 41$.

the answers (i.e., interrater reliability of greater than 86% for knowledge retention and transfer).

Results and Discussion

Like in Experiment 1, the data again suggested that more students in the crammed condition worked all practice tasks and the test tasks referring to retention (i.e., 93%) than students in the distributed condition (i.e., 68%). A Bayesian test of this difference clearly indicated a moderate negative effect (−8 percentage points) of the distributed practice condition on the working of the practice tasks with a BF₁₀ > 100 (i.e., extreme evidence; Lee & Wagenmakers, 2013), 95% credible interval: −3 to −14 percentage points.

Another preliminary analysis was conducted to check whether the performance in the practice sets differed between the two conditions. Such a difference could have occurred because students in the crammed condition solved the practice tasks one week after the last lecture, whereas students in the distributed condition started to solve the practice tasks one day after the last lecture (see Method of Experiment 1 for the rationale). Moreover, stronger learners might have decided to complete all practice tasks, particularly in the distributed condition, in terms of self-selection processes. Table 2 shows the performance in the three practice sets per practice condition. Bayesian t -tests for independent samples revealed no evidence for a difference between the practice conditions, all BF₁₀ = 0.

To test our hypotheses that distributed practice results in a better final test performance than crammed practice, two Bayesian linear regressions were computed with practice condition as independent variable and the test scores (percentage correct) referring to retention and transfer, respectively, as dependent variables. With regard to retention (i.e., test tasks that addressed previously practiced knowledge), the analysis revealed moderate evidence for a positive effect of distributed compared to crammed practice of about 4 percentage points on the test perfor-

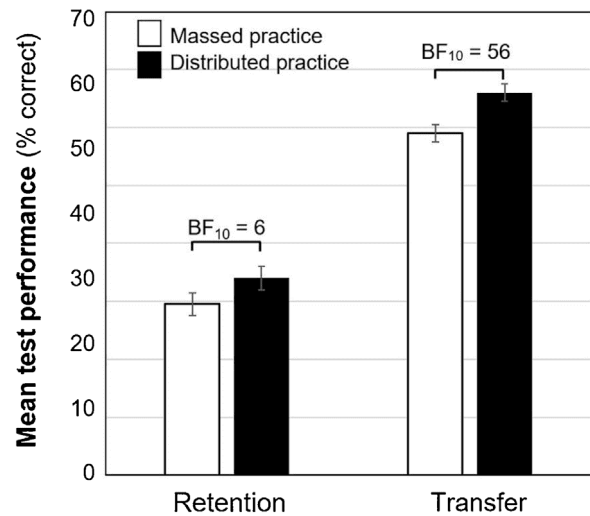


Figure 2. Mean test performance (percent correct) in the crammed and distributed practice condition, separately for retrieval and transfer of knowledge.

mance (BF₁₀ = 6; 95% credible interval: −3 percentage points to 12 percentage points, $d = 0.22$; see Figure 2). Please note that in the first practice set there was an accidental mistake in the wording of one of the practice tasks that could have misled the students into focusing on the “wrong” mistake (the task was to state a mistake in the statement of a fictional student). The mistake in the task was not recognized until after the experiment had started and appeared in both conditions. We nevertheless decided to check the result of the final analyses addressing retrieval by re-running the same model, but only considering the performance in the test tasks that were not related to the task with the mistake. This re-analysis confirmed our results with an even larger Bayes factor (10 instead of 6). Hence, we decided to report the results of the analysis with the complete test performance, including the task with the wording mistake in the first practice set. Concerning transfer performance, there was very strong evidence in favor of a positive effect of distributed compared to crammed practice of about 7 percentage points (BF₁₀ = 56; 95% credible interval: 1 percentage point to 14 percentage points, $d = 0.43$; see Figure 2).

To account for the possibility that weaker students in the distributed practice condition could have decided to drop the practice tasks across the experiment, and that this self-selection might have contributed to the benefit of distributed practice, another Bayesian linear regression analysis was run. It included all students who completed at least the first practice task (not

necessarily all practice tasks) and the final test. This sample consisted of 65 students of the 69 students initially assigned to the crammed condition and 48 of the 60 students initially assigned to the distributed condition. This analysis confirmed the benefit of distributed over crammed practice (i.e., 4 percentage points in the test on previously practiced knowledge, $BF_{10} = 5$; 95% credible interval: -4 percentage points to 11 percentage points; and 8 percentage points in the test on transfer knowledge, $BF_{10} > 100$; 95% credible interval: 2 percentage points to 14 percentage points).

Given the relatively small sample size, learner characteristics were not included as further predictors in the regression models because the power would have been too small. Instead, their role concerning the effect of distributed practice was exploratory analyzed by means of conditional inference tree models (CIT; Hothorn, Hornik, & Zeileis, 2006; Hothorn & Zeileis, 2015). These models are based on recursive binary partitioning and explore potential relationships between the independent variable(s) and the dependent variable. First, the null hypothesis is tested that the dependent variable is unaffected by the independent variable(s). If the null hypothesis can be rejected, the independent variable with the strongest relationship to the dependent variable is selected and the sample is then divided into two maximally differing groups regarding its effect on the dependent variable. The analysis continues until no further independent variable can explain variance of the dependent variable. Thus, instead of testing linear relationships between independent and independent variables, the sample is post-hoc clustered into (at least) two groups that differ maximally. Regarding the retention performance as dependent variable, none of the learner characteristics (including students' performance in the first practice set) yielded a significant effect in these analyses. Concerning transfer performance as dependent variable, the analyses revealed that students differed regarding their performance in the first practice set. For students with a rather poor performance in the first practice set (i.e., max 3.5 points out of 9, $n = 71$), no further variables were relevant for their transfer performance—not even the practice condition. However, for students with a better performance in the first practice set (i.e., more than 3.5 points, $n = 34$), there was an effect of practice condition ($p = 0.026$), with cramming practicing students within this group yielding a poorer transfer performance ($M = 5.6$, $SD = 1.3$, $n = 22$) than distributed practicing students within this group ($M = 7.1$, $SD = 0.9$, $n = 12$). These results suggest that the positive effect of distributed practice on transfer performance might become evident particularly in students with a better initial practice performance. Besides this variable, no other learner characteristics yielded a discriminative effect on transfer.

General Discussion

The aim of the present study was to examine whether statistics knowledge, acquired in a university course, should be worked crammed or distributed to be retained, and whether distributed practice yields positive long-term effects for the retention and transfer of knowledge. In Experiment 1, students were asked to work practice tasks in either a crammed or distributed manner

to improve their memory for the previously acquired contents of the lecture that were relevant for their exam. Only about 41% of the students originally enrolled in the study worked all practice tasks and the final tests (the latter were not announced as such). The effect of distributed practice therefore could not be analyzed. However, exploratory analyses of the participation behavior showed differences between the practice conditions already the first practice set: It was completed by a smaller proportion of students in the distributed condition compared to students in the crammed condition. Because this difference emerged that early, it cannot be ascribed to the distribution of practice. In fact, the schedules of the conditions differed: Students in the distributed condition received the first practice set one day after the last lecture, whereas students in the crammed condition received it (and the other two practice sets) seven days after the last lecture. The initial difference concerning the completion of practice set 1 suggests that students might be more motivated to work practice tasks with a certain delay after the lecture than shortly after the lecture (even if it is more difficult to reactivate the previously acquired knowledge in terms of desirable difficulties; Bjork, 1994).

However, we also found a marked drop-out across practice sessions of Experiment 1 in the distributed condition: Clearly more students worked only the first practice sets than the subsequent practice sets. No such drop was revealed in the crammed practice condition. This pattern underlines the assumption that distributed practice is an obstacle for learners and is therefore hardly used in the context of self-regulated learning (Barzagar Nazari & Ebersbach, 2018; McCabe, 2011; Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013; Wissman, Rawson, & Pyc, 2012; cf. Michael, 1991, who proposes a “procrastination scallop” indicating a lower exam and course success the later students start to complete the required tasks because less time is left). On the other hand, massed or crammed practice is often promoted in real learning contexts by the fact that a university course is usually finalized by an exam at the end. However, multiple tests across the period of the whole course can increase the realization of repeated, distributed practice phases of students in the context of their self-regulated learning (Mawhinney, Bostow, Laws, Blunenfeld, & Hopkins, 1971; see also Barenberg, Roeder, & Dutke, 2018). Our results of Experiment 1 furthermore suggest that learners might need additional instructional support to benefit from distributed practice. Learners should, for example, explicitly be informed about the efficacy of this strategy. It might also be useful to provide them with a corresponding temporal structure and externally enhance their motivation to practice at all (Mullet, Butler, Verdin, von Borries, & Marsh, 2014) and in a distributed manner, in particular.

This was realized in Experiment 2 by making the practice tasks mandatory for students' study certificate. This instruction massively boosted their participation in the practice and test tasks, so that the effect of distributed practice could be analyzed. In a final test, five weeks after the last practice session, students in the distributed practice condition clearly outperformed students in the crammed practice condition concerning (a) the knowledge that they had previously practiced and (b) the generalization of knowledge in terms of transfer knowledge. This is, to

our knowledge, the first study showing that distributing practice enhances adults' generalization of coherent learning contents, referring to statistics, acquired in an authentic educational setting. The results further suggest that distributed practice is an effective tool to improving learners' memory performance in statistics when they are aware that the subject matter is relevant for their exam and when they are provided with a corresponding practice structure (i.e., prompts to work the practice tasks in a distributed manner; see also Hopkins et al., 2016). The latter can easily be realized by assigning homework not all at once but in parts that are temporarily distributed. This approach is slightly more work for the lecturer but helps students improving their memory performance in the long run and fosters their comprehension, enabling them to transfer their knowledge to new tasks.

Exploratory analyses in Experiment 2 suggested that the effect of distributed practice on transfer was stronger for students who had performed better already in the first practice tasks. This finding is in line with the assumption that poorer students might not profit from distributed practice as long as they do not sufficiently grasp the learning content (Barzagar Nazari & Ebersbach, 2019b; Hirsch et al., 1982) or just forgot the content between the practice sessions. However, it is still unclear why no such effect was revealed for retention in the present study.

Our findings are important in that the benefit of distributed practice has not always been observed with mathematical procedures (e.g., Barzagar Nazari & Ebersbach, 2019a; Ebersbach & Barzagar Nazari, 2020; Rohrer & Taylor, 2006). Potential moderators for the effect of distributed practice might be the complexity of the learning material (see the meta-analysis of Donovan & Radosevich, 1999) and the type of knowledge addressed in terms of factual versus procedural knowledge (Cepeda et al., 2006). Further research on such potential moderator effects is necessary.

A limitation of the present study is one difference between the practice conditions: Students in the crammed condition started the practice tasks one week after the last lecture, whereas students in the distributed practice condition started the practice tasks one day after the last lecture. This procedure was done to keep the interval between the completion of the last practice set and the final test constant between both conditions and to schedule the test at the same time in both groups. However, it also established some sort of temporal distribution (i.e., between the last lecture and the onset of the practice set) in the crammed condition, which might have boosted the performance in terms of an additional spacing effect in this condition. In addition, it might also have impaired the performance in the crammed condition because students might have forgotten more of the content from the lectures than students in the distributed condition.

Given the fact that more students in the distributed practice condition than in the crammed practice condition abandoned the study before the end in Experiment 1 and 2, it is also possible that systematic self-selection processes took place. More poorly performing students might have terminated the study prematurely in the distributed practice condition, which might have resulted in a better final test performance in this condition. However, we

checked for performance differences between both conditions in the practice sets in Experiment 2 in the frame of preliminary analyses, and there were none. Thus, systematic effects of the schedule or of self-selection might not account for the effect of distributed practice.

A second limitation is that partly the same sample of students participated in both experiments. Thus, their prior experience with crammed or distributed practice might have affected their acceptance and execution of the practice schedule in Experiment 2.

Finally, the sample was too small to detect further moderating effects of learner characteristics besides initial practice performance. However, one might also infer that potential moderating effects—if they exist—are rather small given that they did not become evident in the sample of the present study. As a result, one might recommend distributed practice to all learners.

To conclude, distributing practice subsequently to the acquisition of statistics knowledge in the context of university lectures fosters both retention and transfer. Distributed practice can be recommended to learners striving to boost their knowledge in the long run in order to establish robust prior knowledge that facilitates their further learning. This recommendation is highly relevant within the frame of university learning in which students must retain and transfer knowledge in both the short term (e.g., a course exam) and the long term (e.g., building prior knowledge as a basis for future courses). In particular, statistics courses quite often stretch across several terms, and retaining prior knowledge is critical to facilitating the acquisition of further knowledge (e.g., Brod, Lindenberger, Wagner, & Shing, 2016).

Author Contributions

Both authors contributed to equal parts to this manuscript (i.e., shared first authorship). ME provided the general idea of this research, supervised its realization, and wrote large parts of the manuscript. KBN specified the design of this study, collected and analyzed the data, and wrote the results section. Both authors revised the previous versions of this manuscript.

Funding

This research was funded by the LOEWE Excellence Initiative of the State of Hesse, Germany, as part of the project “Desirable Difficulties in Learning” (2015–2018).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We would like to thank Manuela Pötschke who allowed us to run the study in the frame of her lectures, Anna Helfers for proofreading the manuscript, and the students who took part despite the “desirable difficulties” in the practice phase. Furthermore, we owe special thanks to Andrew Butler for his helpful suggestions that contributed to finalizing this manuscript.

Appendix A

Example of a Practice Task (Experiment 1)

In the frame of a survey, students were asked how important it is for them to pass their exam with a good result (1 = *totally unimportant*, 2 = *unimportant*, 3 = *rather unimportant*, 4 = *rather important*, 5 = *important*, 6 = *totally important*). Please complete the first three lines of the corresponding frequency table:

k	n_k	p_k
1	3	
2	10	
3	6	
4	11	
5	13	
6	7	
Σ	50	1

Appendix B

Example of a Test Task (Experiment 1)

In the frame of a survey, students were asked how important it is for them that the canteen daily offers at least one vegetarian dish (1 = *totally unimportant*, 2 = *unimportant*, 3 = *rather unimportant*, 4 = *rather important*, 5 = *important*, 6 = *totally important*). Please complete the first three lines of the corresponding frequency table:

k	n_k	p_k
1	18	
2	3	
3	9	
4	12	
5	12	
6	6	
Σ	60	1

Appendix C

Example of a Practice Task (Experiment 2)

In 2002, the International Social Survey Programme (ISSP) conducted an international survey concerning "Family and changing sex roles". It assessed aspects like life satisfaction, satisfaction with job and with family. Based on these data, a data output concerning life satisfaction was generated, showing the relationship between life satisfaction and the subjective perception of different areas of life, such as work place and employment, family and household, and stress.

Source	SS	df	MS	Number of obs	=	23,493
Model	8819.39537	8	1102.42442	F(8, 23484)	=	2146.06
Residual	12063.664	23,484	.513697156	Prob > F	=	0.0000
				R-squared	=	0.4223
				Adj R-squared	=	0.4221
Total	20883.0594	23,492	.888943444	Root MSE	=	.71673

happy	Coef.	Std. Err.	t	P> t	Beta
job_satisfied	.1475367	.0043231	34.13	0.000	.1805946
family_satisfied	.4967155	.004779	103.94	0.000	.5513109
todo_home	-.0066546	.0041355	-1.61	0.108	-.00883
stress_rare_home	.0324853	.0043003	7.55	0.000	.039928
todo_work	-.0064863	.0042801	-1.52	0.130	-.0084001
stress_rare_job	.0018592	.0041573	0.45	0.655	.0023819
fam_responsible	-.0555479	.0110647	-5.02	0.000	-.0274612
conc_difficult	-.0756494	.0158227	-4.78	0.000	-.0252528
_cons	1.44582	.0282884	51.11	0.000	.

Dependent variables:

- “General life satisfaction” (happy)
 - Assessed by a 7-stage scale (0: totally unhappy, 1: very unhappy, 2: unhappy, 3: neither happy nor unhappy, 4: happy, 5: very happy, 6: totally happy)
 - “It is difficult to fulfil my responsibility concerning my family.” (fam_responsible)
 - “It is difficult for me to concentrate.” (conc_difficult)
 - each binary coded (0: seldom, 1: sometimes well)
- Independent variables:

average higher compared to persons who are ‘very satisfied’ with their family life?

Appendix D

Example of a Test Task (Experiment 2)

In 2008, the International Social Survey Programme (ISSP) conducted an international survey concerning “Religion and religiosity”. It assessed aspects like the acceptance of other religions or how one judges the tolerance of religious people. Based on these data, a data output concerning life satisfaction was generated that allows statements concerning the tolerance of religious people.

Source	SS	df	MS	Number of obs	=	47,123
Model	18199.1928	6	3033.19881	F(6, 47116)	=	3205.98
Residual	44576.7226	47,116	.946105836	Prob > F	=	0.0000
				R-squared	=	0.2899
				Adj R-squared	=	0.2898
Total	62775.9154	47,122	1.33219972	Root MSE	=	.97268

rel_intolerant	Coef.	Std. Err.	t	P> t	Beta
rel_conflict	.5120476	.0038233	133.93	0.000	.5247562
respect_rel	.0018374	.0046858	0.39	0.695	.0015699
accept_marriage	.1798959	.0130181	13.82	0.000	.0637884
accept_candidate	.0599141	.0134936	4.44	0.000	.0205904
extremists_meeting	-.1185455	.0124414	-9.53	0.000	-.0508715
extremists_publish	.0174408	.0124356	1.40	0.161	.0075258
_cons	1.170035	.0194875	60.04	0.000	.

Dependent variable:

- “satisfaction with main job” (job_satisfied)
 - “satisfaction with family life” (family_satisfied)
 - each assessed by a 7-stage scale (0: totally unsatisfied, 1: very unsatisfied, 2: unsatisfied, 3: neither satisfied nor unsatisfied, 4: satisfied, 5: very satisfied, 6: totally satisfied)
 - “At home, there are so many things to do.” (todo_home)
 - “My life at home is rarely stressful.” (stress_rare_home)
 - “In my job, there are so many things to do.” (todo_job)
 - “My job is rarely stressful.” (stress_rare_job)
 - each assessed by a 5-stage scale (0: totally disagree, 1: disagree, 2: neither agree nor disagree, 3: agree, 4: totally agree)
- (a) Consider the output concerning the general life satisfaction. Which variable shows with an error rate of 5% ($\alpha = 0.05$) the strongest, significant effect? Justify your answer.
- (b) One acquaintance of you claims that persons who rarely perceive stress in their job are in general more satisfied with their life. Take up position based on the provided model. Assume an error rate of 5% ($\alpha = 0.05$).
- (c) A classmate explains: “The p-value for the effect of the variable ‘My life at home is rarely stressful’ (stress_rare_home) is smaller than 0.05. That means, we can assume with a probability of 95% that the variable ‘My life at home is rarely stressful’ affects general life satisfaction also in the population. Is this interpretation correct? If not: Which mistake did your classmate make?”
- (d) To what extent is the expected general life satisfaction of persons, who are ‘totally satisfied’ with their family life, on

- “Religious people are too intolerant” (rel_intolerant)
 - Assessed by a 5-stage scale (0: totally disagree, 1: disagree, 2: neither disagree nor agree, 3: agree, 4: totally agree)
- Independent variables:
- “Religions evoke conflicts” (rel_conflicts)
 - “One should respect all religions” (respect_rel)
 - each assessed by a 5-stage scale (0: totally disagree, 1: disagree, 2: neither disagree nor agree, 3: agree, 4: totally agree)
 - “Acceptance of a person of another religion who marries a relative” (accept_marriage)
 - “Acceptance of a person of another religion who candidates for a political party” (accept_candidate)
 - “Religious extremists should be allowed to held public meetings.” (extremists_meeting)
 - “Religious extremists should be allowed to burn books.” (extremists_publish)
 - each variable binary coded (0: no acceptance, 1: acceptance)
- (a) Consider the output concerning religious tolerance. Which variable shows with an error rate of 5% ($\alpha = 0.05$) the strongest, significant effect? Justify your answer.
- (b) One acquaintance of you claims that persons who accept the marriage between one of their relatives and a person of another religion tend to believe that religious people are tolerant. Take up position based on the provided model. Assume an error rate of 5% ($\alpha = 0.05$).

- (c) A classmate explains: “The p-value for the effect of the variable ‘Acceptance of a person of another religion who candidates for a political party’ (accept_candidate) is smaller than 0.05. That means, one can assume with a probability of 95% that this variable has an impact on the opinion on intolerance of religious people in the population. Is this interpretation correct? If not: Which mistake did your classmate make?”
- (d) To what extent is the expected convincement concerning the intolerance of religious people who ‘totally agree’ with the statement ‘religions evoke conflicts’, on average higher compared to persons who do ‘not agree’ with this statement?

References

- Arslan, R. C., Tata, C. S., & Walther, M. P. (2018). *formr: A study framework allowing for automated feedback generation and complex longitudinal experience sampling studies using R. (version v0.17.21)*. <http://dx.doi.org/10.5281/zenodo.1345615>
- Barenberg, J., Roeder, U.-R., & Dutke, S. (2018). Students’ temporal distributing of learning activities in psychology courses: Factors of influence and effects on the metacognitive learning outcome. *Psychology Learning & Teaching*, 17(3), 257–271. <http://dx.doi.org/10.1177/1475725718769488>
- Barzagar Nazari, K., & Ebersbach, M. (2018). Distributed practice: Rarely realized in self-regulated mathematical learning. *Frontiers in Psychology*, 9, 2170. <http://dx.doi.org/10.3389/fpsyg.2018.02170>
- Barzagar Nazari, K., & Ebersbach, M. (2019a). Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom. *Journal of Applied Cognitive Psychology*, 33, 288–298. <http://dx.doi.org/10.1002/acp.3485>
- Barzagar Nazari, K., & Ebersbach, M. (2019b). Distributed practice in mathematics: Recommendable especially for students on a medium performance level? *Trends in Neuroscience and Education*, 17, Article 100122 <http://dx.doi.org/10.1016/j.tine.2019.100122>
- Bjork, R. A. (1994). *Memory and meta-memory considerations in the training of human beings*. In J. Metcalfe, & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: MIT-Press.
- Boerner, S., Seeber, G., Keller, H., & Beinborn, P. (2005). Lernstrategien und Lernerfolg im Studium: Zur Validierung des LIST bei berufstätigen Studierenden. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37(1), 17–26. <http://dx.doi.org/10.1026/0049-8637.37.1.17>
- Brod, G., Lindenberger, U., Wagner, A. D., & Shing, Y. L. (2016). Knowledge acquisition during exam preparation improves memory and modulates memory formation. *The Journal of Neuroscience*, 36, 8103–8111. <http://dx.doi.org/10.1523/jneurosci.0045-16.2016>
- Butler, A. C., Marsh, E. J., Slavinsky, J. P., & Baraniuk, R. G. (2014). Integrating cognitive science and technology improves learning in a STEM classroom. *Educational Psychology Review*, 26, 331–340. <http://dx.doi.org/10.1007/s10648-014-9256-4>
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, 24(3), 369–378. <http://dx.doi.org/10.1007/s10648-012-9205-z>
- Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, 13, 826–830. <http://dx.doi.org/10.3758/BF03194004>
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. <http://dx.doi.org/10.1037/0033-2909.132.3.354>
- Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, 56(4), 236–246. <http://dx.doi.org/10.1027/1618-3169.56.4.236>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [http://dx.doi.org/10.1016/S0022-5371\(72\)80001-X](http://dx.doi.org/10.1016/S0022-5371(72)80001-X)
- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork, & R. A. Bjork (Eds.), *Handbook of perception and cognition* (2nd ed., pp. 317–344). San Diego, CA, US: Academic Press. <http://dx.doi.org/10.1016/B978-012102570-0/50011-2>
- Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, 43(8), 627–634. <http://dx.doi.org/10.1037/0003-066X.43.8.627>
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don’t. *Journal of Applied Psychology*, 84(5), 795–805. <http://dx.doi.org/10.1037/0021-9010.84.5.795>
- Ebersbach, M., & Barzagar Nazari, K. (2020). No robust effect of distributed practice on the short- and long-term retention of mathematical procedures. (in press), *Frontiers in Psychology: Cognition*, <http://dx.doi.org/10.3389/fpsyg.2020.00811>
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54, 5–12. <http://dx.doi.org/10.1037/0022-3514.54.1.5>
- Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children’s science curriculum. *Applied Cognitive Psychology*, 28(2), 266–273. <http://dx.doi.org/10.1002/acp.2997>
- Hintzman, D. L. (1974). *Theoretical implications of the spacing effect*. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Oxford, England: Erlbaum.
- Hintzman, D. L., & Rogers, M. K. (1973). Spacing effects in picture memory. *Memory & Cognition*, 1, 430–434.
- Hirsch, C. R., Kapoor, S. F., & Laing, R. A. (1982). Alternative models for mathematics assignments. *International Journal of Mathematical Education in Science and Technology*, 13(3), 243–252. <http://dx.doi.org/10.1080/0020739820130301>
- Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. S. (2016). Spaced retrieval practice increases college students’ short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, 28(4), 853–873. <http://dx.doi.org/10.1007/s10648-015-9349-8>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15, 651–674. <http://dx.doi.org/10.1198/106186006X13393>
- Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partitioning in R. *Journal of Machine Learning Research*, 16, 3905–3909. <http://jmlr.org/papers/v16/hothorn15a.html>
- Jerusalem, M., & Satow, L. (1999). Schulbezogene Selbstwirksamkeitserwartung. In *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin. Retrieved from <http://www.psyc.de/skalendoku.pdf>

- Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 12–19. <http://dx.doi.org/10.1177/2372732215624708>
- Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, 36, 38–45. <http://dx.doi.org/10.1016/j.learninstruc.2014.11.001>
- Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 704–719. <http://dx.doi.org/10.1037/0278-7393.33.4.704>
- Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory*, 20(1), 37–47. <http://dx.doi.org/10.1080/09658211.2011.631550>
- Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition*, 42(5), 729–741. <http://dx.doi.org/10.3758/s13421-014-0394-1>
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.
- Mawhinney, D. E., Bostow, D. E., Laws, D. R., Blunenfeld, G. J., & Hopkins, B. L. (1971). A comparison of studying-behavior produced by daily, weekly, and three-week testing schedules. *Journal of Applied Behavior Analysis*, 4, 257–264. <http://dx.doi.org/10.1901/jaba.1971.4-257>
- McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, 39(3), 462–476. <http://dx.doi.org/10.3758/s13421-010-0035-2>
- Michael, J. (1991). A behavioral perspective on college teaching. *The Behavior Analyst*, 14, 229–239. <http://dx.doi.org/10.1007/bf03392578>
- Moulton, C.-A. E., Dubrowski, A., Macrae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: What kind of practice makes perfect? *Annals of Surgery*, 244(3), 400–409. <http://dx.doi.org/10.1097/01.sla.0000234808.85789.6a>
- Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*, 3, 222–229. <http://dx.doi.org/10.1016/j.jarmac.2014.05.001>
- Namaziandost, E., Rahimi Esfahani, F., & Hashemifardnia, A. (2018). The comparative effect of spacing instruction and massed instruction on intermediate EFL learners' reading comprehension. *SAGE Open*, 8(4), 21–28. <http://dx.doi.org/10.1177/2158244018811024>
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328–346. <http://dx.doi.org/10.1037/0033-295X.91.3.328>
- Nicholls, J., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing students' theories of success in mathematics: Individual and classroom differences. *Journal of Research in Mathematics Education*, 21, 109–122. <http://dx.doi.org/10.2307/749138>
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology*, 20(9), 1209–1224. <http://dx.doi.org/10.1002/acp.1266>
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6), 481–498. <http://dx.doi.org/10.1007/s11251-007-9015-8>
- Smith, C. D., & Scarf, D. (2017). Spacing repetitions over long timescales: A review and a reconsolidation explanation. *Frontiers in Psychology*, 8, 962. <http://dx.doi.org/10.3389/fpsyg.2017.00962>
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C., & Dickhäuser, O. (2012). *Skalen zur Erfassung der Lern- und Leistungsmotivation: SELMO* Retrieved from. Göttingen [u.a.]: Hogrefe. <https://ub-madoc.bib.uni-mannheim.de/38090/>
- Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, 38(2), 244–253. <http://dx.doi.org/10.3758/MC.38.2.244>
- Svihla, V., Wester, M. J., & Linn, M. C. (2018). Distributed practice in classroom inquiry science learning. *Learning: Research and Practice*, 4(2), 180–202. <http://dx.doi.org/10.1080/23735082.2017.1371321>
- Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: do people interleave or block exemplars during study? *Psychonomic Bulletin & Review*, 20(2), 356–363. <http://dx.doi.org/10.3758/s13423-012-0319-6>
- Toppino, T. C., Phelan, H.-A., & Gerbier, E. (2018). Level of initial training moderates the effects of distributing practice over multiple days with expanding, contracting, and uniform schedules: Evidence for study-phase retrieval. *Memory & Cognition*, 46(6), 969–978. <http://dx.doi.org/10.3758/s13421-018-0815-7>
- Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, 83(4), 1137–1144. <http://dx.doi.org/10.1111/j.1467-8624.2012.01781.x>
- Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1), 246–254. <http://dx.doi.org/10.1037/a0025260>
- Vlach, H. A., Sandhofer, C. M., & Bjork, R. A. (2014). Equal spacing and expanding schedules in children's categorization and generalization. *Journal of Experimental Child Psychology*, 123, 129–137. <http://dx.doi.org/10.1016/j.jecp.2014.01.004>
- Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, 109(1), 163–167. <http://dx.doi.org/10.1016/j.cognition.2008.07.013>
- Wild, K. P., & Schiefele, U. (1994). *Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens*. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 15, 185–200.
- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory*, 20(6), 568–579. <http://dx.doi.org/10.1080/09658211.2012.687052>

Further reading

H.A. Vlach. The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn *Child Development Perspectives* 2014; 8: 163–168. doi:10.1111/cdep.12079

Received 26 September 2019;
received in revised form 6 May 2020;
accepted 28 August 2020
Available online xxx