Trends in Neuroscience and Education 17 (2019) 100122

Contents lists available at ScienceDirect

Trends in Neuroscience and Education

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

Research paper

Distributed practice in mathematics: Recommendable especially for students on a medium performance level?

Katharina Barzagar Nazari*, Mirjam Ebersbach**

Department of Psychology, University of Kassel, Hollaendische Str. 36-38, 34127 Kassel, Germany

A R T I C L E I N F O

Keywords:
Desirable difficulties
Distributed practice
Spacing effect
Mathematics
School

A B S T R A C T

In this study, the effect of distributed practice on the mathematical performance of 7th graders was investigated ($N = = 81$). After a stochastics lesson, one group of students worked three sets of exercises massed on one day, while the other group of students worked the same exercises distributed over three days. Bayesian analyses of the performance two weeks after the last practice revealed no evidence for an effect of practice condition. However, in a test after six weeks, strong evidence for a positive effect of distributed practice was revealed. Exploratory analyses indicated that especially students in the medium performance range benefitted from distributed practice. The results are discussed regarding the question under which circumstances distributed practice proves a useful strategy for mathematical learning.

1. Introduction

One important aim of learning in school and other real-world learning contexts is to acquire knowledge and skills that can be retrieved not only after a short delay but also in the long run. However, many strategies usually applied by learners, such as repeatedly reading the learning content, promote rather short-term retention [7]. Thus, the knowledge might be retrieved in the next exam but then quickly starts to decay. Several mechanisms have been proposed that aim at promoting long-term retention in particular, by aggravating the learning process for the learner. These mechanisms are labelled *desirable difficulties* [6] and are related to a range of successful learning strategies, such as, for instance, *spacing* or *distributed practice* – the strategy which is central for the present study. Distributed practice means that the total time for learning is distributed across at least two temporarily separated learning sessions – instead of massed in only one learning session [12]. That is, the only difference between distributed and massed practice is that with distributed practice, the total learning duration is not spent learning in one session straight but with one or more lags in between the sessions. However, the learning material and the total learning duration (or the number of repetitions) is the same for both distributed and massed practice. Several meta-analyses demonstrated a significant advantage for distributed compared to massed practice, with medium to large effect sizes [12, 21, 32]. Moreover, it could be shown that given a fixed retention interval, the size of the effect of distributed practice varies with the lag between the practice sessions. This phenomenon is called lag effect and empirical results led to the conclusion that longer retention intervals profit from longer temporal distances between the distributed practice sessions, although this relationship is not considered to be linear [12, 13, 39].

Several theoretical accounts try to explain the positive effect of distributed practice (for overviews see [38, 62]). One account assumes *deficient processing*. Here, it is supposed that when a person learns an item by successive repetitions (as in massed practice), the person (intentionally or not) processes the item more poorly than when the repetitions are separated by time lags (as in distributed practice). Thus, deficient processing takes place in massed practice, while distributed practice enhances processing and thereby improves retention performance [15, 17, 27]. Another potential explanation for the superior performance after distributed practice compared to massed practice is *study phase retrieval*. The central assumption here is that the retrieval of a learned item or chunk of information supports long-term retention, especially when the retrieval process itself is hard, yet successful. When the lag between learning situations increases, as in distributed practice, retrieval is harder than with massed practice, but it maximally strengthens the memory trace if retrieval is successful. In massed

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).

* Corresponding author.
** Co-corresponding author.
E-mail addresses: k.barzagar@posteo.de (K.B. Nazari), mirjam.ebersbach@uni-kassel.de (M. Ebersbach).

https://doi.org/10.1016/j.tine.2019.100122
Received 31 March 2019; Received in revised form 16 September 2019; Accepted 18 October 2019
Available online 24 October 2019

practice, on the other hand, retrieval throughout the learning period should be fairly easy, because little time passed since the last retrieval attempt. Unfortunately, thereby the memory trace is hardly strengthened, resulting in poorer long-term performance [5, 38, 61]. Finally, *encoding variability* theories are based on the assumption that a learned item is stored together with retrieval cues. These retrieval cues can be of various kinds and consist, for example, of the emotional state during learning, physical information of the learning environment, or associations between the item and prior knowledge. It is assumed that in massed practice, where repetitions follow immediately one after another, the item is encoded repeatedly in roughly the same way, resulting in little growth of relevant retrieval cues. With distributed practice, in contrast, more time passes between the repetitions and chance is thus higher that the item is encoded in a different way, accompanied by different retrieval cues. The greater variability of retrieval cues after distributed practice should then boost performance compared to massed practice [17, 24, 62]. By now, the different approaches proposed to explain the positive effect of distributed versus massed practice are assumed to be not exclusive. Instead, several mechanisms might work simultaneously. Most attempts to formulate such hybrid theories combine different variations of study phase retrieval and encoding variability approaches [19, 64, 65, 69].

1.1. Distributed practice of mathematical procedures

The large majority of studies on desirable difficulties in general, and on the distributed practice effect in particular, included adults as participants who were tested in the laboratory, involving rather incoherent learning contents that referred to rote memory of declarative knowledge, such as word lists or pictures [11, 12, 20, 34]. Although the interest in examining the effect of distributed practice in applied learning settings is increasing (e.g., [35]), it is still a fairly open question of whether distributed practice yields positive effects if curriculum-based learning contents are involved, aiming at promoting the acquisition of procedural knowledge, such as in mathematics. The present study addresses this question by examining whether distributed practice of mathematical procedures in the classroom leads to better performance than massed practice in secondary school students. Our main interest lies in the effect of distributed practice with repetitions. Therefore, the following short review concentrates mostly on studies that investigated the effect of distributed practice with repetitions of previously learned material. Another approach would be to distribute lessons or other forms of input sessions. This could be labeled distributed learning and constitutes a slightly different research area (e.g. [9]).

Concerning distributed practice in mathematics, there are two studies with college students who practiced permutation problems either in a distributed fashion with a 7-days-delay or in a massed fashion on one day. The distributed practice group clearly outperformed the massed practice group after one week [51] as well as after four weeks [50]. Yazdani and Zebrowski [68] used a combination of distributed and interleaved practice for the geometry homework of high school students over the course of six weeks and showed that working on exercises of previous and current topics (distributed + interleaved practice) improved performance compared to working on exercises only of current topics (massed practice), up until six weeks after the learning period. A similar approach was already pursued by Hirsch, Kapoor and Laing [28] and Saxon [53] who used a combination of distributed and interleaved practice over several months for mathematics assignments in college [28] and courses of the 9th grade in school [53] and found that this “mixed review homework strategy” led to better performance of students in particular in the lower and medium performance ranges (for a review on mixed review in mathematics see also [49]).

To our knowledge, there are only three studies investigating the isolated effect of distributed practice in mathematics in the school context. In a study by Schutte, Duhon, Solomon, Poncy, Moore and Story [54], 3rd graders practiced basic addition problems daily in four 1-min practice sessions (a) consecutively (i.e., massed), (b) distributed with two back-to-back sessions in the morning and two back-to-back sessions in the afternoon, and (c) all four sessions separated and distributed across one day. The three practice procedures were repeated across 19 days in total. Practice was explicitly timed and the number of correct digits per minute at the test, reflecting basic math fact fluency, served as dependent variable. Third graders in both distributed practice conditions gained more basic math fact fluency across the 19 days than 3rd graders in the massed practice condition. However, one might argue that all three groups practiced in a distributed manner as they practiced addition across multiple days.

Chen, Castro-Alonso, Paas and Sweller [14] compared the performance of 4th and 5th graders who practiced fraction addition (Grade 4), calculation with negative numbers, and solving fractional equations (both Grade 5) either massed on one day or distributed across three days. The authors found that distributing practice over three days significantly improved performance compared to massed practice in both grades. However, these results have to be interpreted with caution because the students of the massed practice condition were tested immediately after finishing their practice exercises while the students of the distributed practice condition were tested one day after they had finished their practice exercises. Thus, the retention interval was not held constant across both conditions, which might have caused an adverse effect on the attention and motivation of the massed practicing students.

Another direct comparison between the effects of massed and distributed practice of mathematical procedures in school was conducted by Barzagar Nazari and Ebersbach [3]. They taught 3rd- and 7th graders a mathematical procedure previously unknown to the students (i.e., semi-formal multiplication and probability calculation, respectively) that was then practiced in the classroom either massed for 45 min on one day or distributed across three consecutive days for 15 min each. The performance in similar but not identical tasks was assessed in two tests taking place one week and six weeks after the last practice session. Bayesian analyses suggested very strong evidence for a positive effect of distributed practice compared to massed practice one week after the last practice session in Grade 3, but the effect diminished in the second test, conducted six weeks after the last practice session. It was assumed that, while the 3rd graders did not practice semi-formal multiplication in class until the last test was conducted, they still most likely practiced related skills within other topics that were covered (e.g., they could have practiced mental multiplication, which is needed for semi-formal multiplication), which could then have led to a decreasing performance difference between the two practice conditions in the second test. In Grade 7, in contrast, there was strong evidence for a positive effect of distributed practice compared to massed practice on both the intermediate and long-term test performance. However, contrary to the hypothesis that the positive effect of distributed practice would become more pronounced in the long run [38], it diminished (as in Grade 3) or remained stable (as in Grade 7). One special feature of the described study was that the students of both grades did not receive feedback on their practice performance nor were they provided with the correct solutions. Though it is encouraging that there was evidence for a positive effect of distributed practice even without feedback or correct solutions, the question remains of whether the effect also emerges when feedback is provided during the practice sessions. Thus, students in the current study received feedback on their practice performance by being presented with the correct solution once they had completed a practice task.

1.2. The influence of individual characteristics on the effectivity of desirable difficulties

Another relevant – but so far rather neglected – aspect concerning the applicability of desirable difficulties in real learning contexts is whether the effects emerge for all learners in a similar manner or whether they depend on learners' *motivational and cognitive characteristics* (see also [22]). This question is particularly important if one strives to give recommendations to teachers concerning the use of desirable difficulties in school. Such aptitude-treatment-interactions [16, 56] have been established in many instructional areas and are supported by empirical findings (e.g., [26]). However, with regard to desirable difficulties, such interactions were theoretically assumed, too (e.g., [19, 43]), but rarely tested. With regard to distributed practice, indications of moderating variables are sparse. While the effect of distributed practice has also been demonstrated for a broad *age range* including children (e.g., [55]), the role of individual differences with regard to motivational or cognitive variables is less clear. A few studies examined the effect of *prior knowledge* – some yielding a larger benefit by distributed and interleaved practice for students with low prior knowledge [28, 53], while others found no such interaction [47]. Studies on the effect of working memory capacity seem to indicate that there is no interaction with the effect of distributed practice [18, 55].

Barzagar Nazari and Ebersbach [3] addressed the aspect of individual differences concerning the effect of distributed practice more broadly by additionally assessing various cognitive variables (i.e., initial practice performance, working memory capacity, sustained attention, metacognitive monitoring) and motivational variables (i.e., mathematical self-efficacy, effort motivation, performance-avoidance goals, work avoidance) of 3rd and 7th graders in their sample. However, none of these variables yielded a moderating effect. Given the sparse empirical basis concerning moderating variables on the effect of distributed practice, we addressed this aspect in an exploratory manner in the present study, too: One of the challenges of distributed practice compared to massed practice is that learners have to repeatedly engage in a topic that they may already have forgotten about, because the last learning or practice session was several hours or days ago. In the current study, academic self-efficacy, that is, the belief of a student in his or her capability to master a task or topic in the frame of formal education [71], could be challenged. For students lacking (mathematical) self-efficacy, this may result in a complete surrender to the topic because they cannot keep up while repeatedly being confronted with their own failings. A similar reasoning can be applied to individual effort motivation, which can be defined as the willingness to invest internal ressources to solve a task (c.f., [67]), with students displaying high effort motivation being more likely to keep working even when they face difficulties in distributed practice sessions. On the other hand, students displaying low effort motivation might be tempted to stop to engage with the topic in distributed practice, because they repeatedly have to retrieve information that has been forgotten again, eventually reducing the positive effect of this learning strategy. Concerning concentration ability, which in the current context is defined as the ability to focus attention on a given task (c.f., [67]), however, it might be especially those with poor concentration who benefit most from distributed practice, because it requires to concentrate for a shorter time compared to massed practice. Hence, *mathematical self-efficacy, effort motivation* and *self-rated concentration difficulty* were considered to be potential moderators for the effect of distributed practice in this study. Due to a programming error, however, only self-efficacy and concentration difficulty could be evaluated (see Design).

1.3. Research questions and hypotheses

With this study, we aimed at extending the empirical grounds regarding the usability of distributed practice for mathematical learning and regarding the question of whether all learners benefit from this learning strategy to a similar extend or not. Therefore, we examined the effect of distributed mathematical practice with feedback among secondary school students and investigated three potential moderating motivational variables. Students practicing mathematical procedures in a distributed fashion were expected to outperform students practicing the same procedures in a massed fashion in an intermediate test, conducted two weeks after the practice. In a long-term test, conducted six weeks after practice, the positive effect of distributed practice was expected to be even larger, because the effect of distributed practice is known to be especially pronounced in the long run [38]. In addition, we analyzed moderating effects of motivational variables of the learners that theoretically have the potential to explain individual differences regarding the effect of distributed practice in an exploratory manner. Additionally, mainly based on the results of Hirsch et al. [28] and Saxon [53], we decided to investigate the relationship between the initial practice performance (which can be conceived as baseline performance or content specific ability) and the effect of distributed practice. This question was pursued in an exploratory analysis as well.

2. Method

2.1. Participants

The initial sample included 142 7th graders of four schools, located in and around a middle sized German city in neighborhoods with a medium socio-economic status.[1] All students attended higher level courses aiming at the German higher education entrance qualification "Abitur". They participated voluntarily and could terminate their participation at all times. Prior to the experimental manipulation, the parents signed a consent form allowing their children to participate in the study. Only data of students who attended at least one of two lecture sessions and worked all three of the practice sheets as well as both tests were analyzed. This criterion and an irregularity of the testing time (see Practice and test sessions and Appendix B) led to the reduction of the initial sample so that the final sample included 81 7th graders (46 female, 35 male, M_{age} = 13 years 3 months, age range: 12–14 years).

2.2. Design

Practice condition served as independent variable and was manipulated between subjects: Within each class, one group of students practiced massed on one day and the other group practiced distributed on three different days. For the distributed practicing students, an expanding interval schedule was applied (see Material and Procedure; [40]). Before the students were assigned to one of the two practice conditions, they were ranked by their mathematics grade of their last school certificate, and then, within each grade group, they were randomly assigned to one of the two conditions, ensuring that the overall math performance was roughly equal in both conditions before manipulation. Their final performance in the practiced content (i.e., probability calculation) served as dependent variable and was tested two and six weeks after the last practice set was finished. To examine if individual learner characteristics moderated the effect of distributed practice, we additionally included a questionnaire assessing three motivational traits. However, for the items on effort motivation accidentally the wrong answer scale was programmed and because of this, the items could not be evaluated. In Table 1, the variable scales of the two remaining measures are displayed along with translated sample items and information on their reliability based on our data (Cronbach's alpha). Cronbach's alpha measures the internal consistency of a questionnaire scale and is used to assess how related the different questions of the scale are. High Cronbach's alpha values (>0.7) are taken as an indicator that the respective questionnaire items are closely related and measure a common construct. Finally, after finishing an exercise the students were asked how hard they considered the previous exercise.

[1] This study was carried out in accordance with the recommendation of the ethics committee of the Faculty of Human Sciences of the University of Kassel with written informed consent from all legal guardians of the subjects in accordance with the Declaration of Helsinki.

Table 1
Instruments used to assess potential moderators.

Motivational traits	Employed instrument	Reliability
Mathematical self-efficacy	7 items of a German *Academic Self-Efficacy Scale for School Children* [33], adapted to mathematics (Sample item: “In math, I can solve even the difficult problems if I try hard.”)	α = 0.87
Concentration difficulty	6 items of the German *Learning Strategies in College – LIST* [8, 67] (Sample item: “When I'm learning, I'm easily distracted.”)	α = 0.91

This question can be used to examine whether students of the distributed practice condition in fact perceived the exercises as more difficult than students of the massed practice condition.

2.3. Material and procedure

To investigate the effect of distributed practice in an applied setting, we chose a topic from the regular curriculum of Grade 7 (i.e., basic probability calculation) that had not been introduced before, and prepared the lessons and exercises in close collaboration with didactic experts and teachers. The students learned how to calculate simple probabilities and to draw one- and two-stage tree diagrams. Only classes who had not already covered probability calculation in the current school year participated in the study. The lesson scripts, all practice and test sets as well as the corresponding scoring schemes are provided in German language online (https://osf.io/d542q/?view_only=7896f90d809140d08b777fba6d564454).

Prior testing and introductory lessons. In the first session (about 45 min), all students individually answered a questionnaire on tablets concerning two motivational traits (see Table 1). The questionnaire was programmed with the survey tool LimeSurvey [45]. Thereafter, students' prior knowledge on probability calculation was assessed via pen-and-paper to ensure that the students had in fact no significant experiences concerning this topic before the study started. The session was directly followed by the first introductory lesson on probability calculation (45 min). A second, longer introductory lesson on probability calculation (90 min) was held one or two days after the first introductory lesson, depending on the class schedule. Both introductory lessons were held by student assistants experienced in teaching children and adolescents, who were supervised by the authors. The introductory lessons were designed as normal school lessons and included examples and short experiments (e.g., coin tossing). They were very similar to the lessons used for the abovementioned study by Barzagar Nazari and Ebersbach [3].

Practice and test sessions. Five to six days after the second introductory lesson, the practice period started. The exercises for the practice sheets and for the final performance tests were based on the topics covered in the introductory lessons and worked on the same tablets that were used for the questionnaire. This had the advantage that after each exercise the correct solution could be shown on screen individually and independently from the pace of the other students. One practice set consisted of three exercises each that involved the calculation of simple probabilities and the labeling as well as interpretation of a tree diagram. The exercises of one practice set could easily be worked in less than 15 min. An example of a practice set can be found in Appendix A. The total practice duration of 45 min resembles the optimal daily homework duration in secondary school (e.g., [23]). Students of the massed condition worked all three sets consecutively on the first day of practice. Students of the distributed condition started practicing the same day, but worked only the first practice set, after two days the second practice set, and another five days later the third practice set (i.e., expanding interval; s. Fig. 1) .[2] The practice sets consisted of conceptually similar but not identical exercises, that is, students could not memorize the exact solutions but only the solution process. The students were not allowed to use their material from the introductory lessons while working the exercises, but after finishing an exercise, they were shown the correct solution before going on to the next exercise (i.e., feedback). Right before the correct solution was shown, the students were asked how hard they considered the exercise on the previous page.[3] It was not possible to click back to previous pages.

Two and six weeks after the last practice set has been executed, the students were tested with tasks similar to the ones of the practice sets. Different from the practice sessions, no correct solution was provided after test tasks. In addition, the students were not explicitly told in advance that they would be tested to prevent them from preparing for the tests. Instead, both tests were announced as further exercises.

There were several deviations concerning the timing of the second test assessing long-term performance, originally scheduled six weeks after the last practice set, which led to the exclusion of one course. More information on the rationale behind this decision can be found in Appendix B.

The students did not receive any individual feedback on their performance in practice sets, but the same sample solution was provided regardless of the answer given by the student. No feedback at all was provided on the performance in test sessions. The student assistants supervising the practice and test sessions were instructed to respond only with strategic feedback to substantial questions concerning probability calculation (e.g., “Can you remember how we solved this kind of problem during the lesson?” or “Maybe you remember that there was a trick to solve this kind of problem?”). In addition, we asked the regular math teachers not to elaborate the topic before the last test had been realized. At the last test session, all teachers confirmed that they did not work on probability calculation during the study period.

All practice exercises and test tasks were programmed using the research tool “formr” [1] and worked in class on tablets in a self-paced manner. Students who finished first were asked to wait quietly until the rest of the class had finished too. On times when the students did not have to work any exercises, they were provided brain-teasers as filler tasks.

Coding. Correct solutions included only unequivocal numerical values or single words. In total, three raters (i.e., two per class) rated the answers in the practice and test sets using a predefined scoring scheme. Afterwards, the ratings were compared and potential differences were resolved among the raters. In all cases, the differences were due either to mistyped values or scores that were unintentionally not rated by one of the raters.

3. Results

The data sets and R scripts for data preparation and analyses are provided online [files will be added after acceptance of the manuscript].

The prior knowledge test, conducted before the lessons, confirmed that the participating sample had a rudimentary knowledge of probability calculation, at most, with only seven of 81 students gaining scores of 50% or more. It should be noted that no student was removed

[2] In one school, eight students of the distributed practice condition had lags of two and three days instead of two and five days, due to a national holiday.

[3] Question: “What would you say, how hard was the exercise you just worked?” The answers were given on a five-point scale: 1 “Very easy”, 2 “Easy”, 3 “Medium”, 4 “Hard”, 5 “Very hard”.

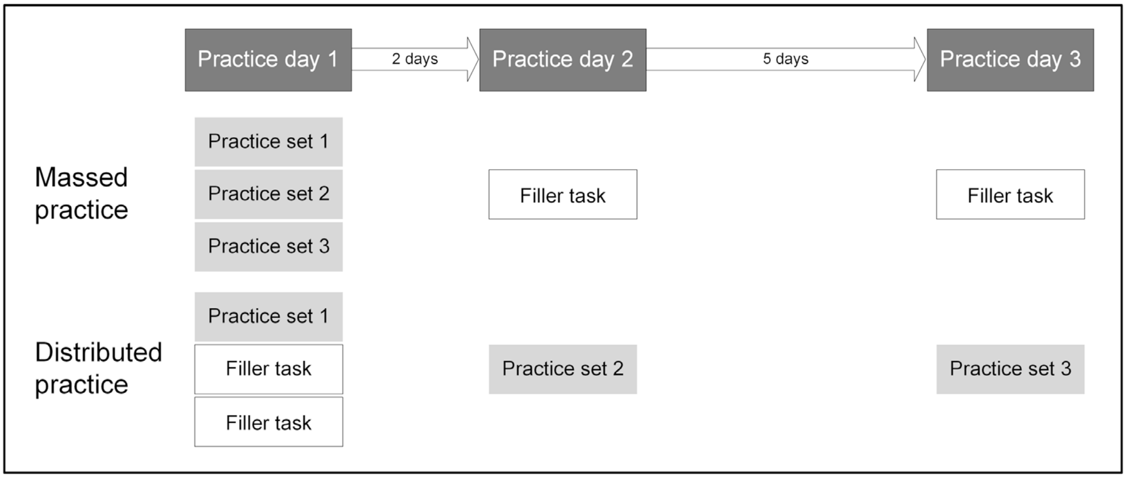


Fig. 1. Illustration of the two practice schedules. Each practice set and filler task took no more than 15 min.

Table 2
Mean performance scores of the students in practice and test sets.

Practice condition	Practice sets 1	2	3	Test sets 2 Weeks	6 Weeks
Massed (*n* = = 36)	5.53 (2.71)	7.19 (2.69)	6.89 (2.75)	6.22 (3.03)	5.90 (3.03)
Distributed (*n* = = 45)	5.58 (2.58)	7.22 (2.12)	6.99 (2.37)	6.56 (2.74)	6.87 (2.81)

Note. Mean scores, standard deviation in parentheses. Maximum score of each set: 9.5. The tasks per set were of the same type but not identical.

based on the performance in the pre-test. However, we used the performance in the first practice set to control for performance differences that were independent from the experimental manipulation (see model specification). The performance of students in the massed and distributed condition in the practice and test sets are displayed in Table 2.

Bayesian regression models were used to test the hypotheses. Among others, a reason to choose Bayesian statistics over the frequentist approach was the relatively small sample size. While the binary results (significant or not significant) of the frequentist approach can make reliable conclusions complicated or even impossible when small sample sizes are involved, Bayesian statistics allows to assign each parameter a range of values together with the respective probability for the effect, even with small sample sizes [37]. Each of the models included the practice condition (distributed vs. massed practice) to test the respective hypothesis and the score in the very first practice set as control variable. The control variable was included to ensure that the effect of distributed practice was not due to performance differences that existed prior to the experimental manipulation. It was decided to use the initial practice performance as control variable instead of the prior knowledge score, because it is the closest approximation for performance differences that existed *after* the introductory lessons were given but *before* the practice manipulation was introduced. Because the main question is how practice following an input session should be scheduled optimally, this is the critical moment to control for performance differences. The models were estimated in R (R [46]) using the brms package[4] [10] and checked for autocorrelation and proper chain conversion. These checks indicated that the sampling process resulted in samples that showed no autocorrelation and were evenly distributed around the mean of the posterior values. Because there are only few existing results on the effect of distributed practice on mathematical performance in school, no priors were specified. That is, for the two independent variables an improper flat distribution over the reals served as prior distribution [10].

To test the first hypothesis that distributing mathematical practice improves performance two weeks after the last exercises more than massing the same amount of practice, a Bayesian linear regression model with the performance in the first test as dependent variable was computed. Practice condition and initial practice performance served as independent variables. In contrast to the hypothesis, there was only little evidence for a positive effect of distributed practice: The mean of the posterior distribution for the effect of distributed practice was 0.31 (95% credible interval = −0.79 to 1.40). That is, the expected performance difference between students of the distributed practice condition and those of the massed practice condition is expected to be between −0.79 and 1.40, with the most likely value being around 0.31 points (with regard to a test score ranging from 0 to 9.5 points). The evidence ratio of 2.41 confirms that it is only about twice as likely that distributed practice has a positive effect than that it has no effect or a negative effect on the performance two weeks after the last practice set, compared to massed practice. In other words, based on the presented data it is only twice as likely that any of the values above 0 is true – compared to any value below zero (mathematically, the probability area for values above 0 are set in relation to the probability area for values below zero, resulting in a ratio of approximately 2.4). Referring to Lee and Wagenmakers [41], this is only “anecdotal evidence” for a positive effect of distributed practice. The mean for the effect of the performance in the first practice set was 0.58 (95% credible interval = 0.37 to 0.78), that is, a higher score in the first practice set was related to a higher score in the first test.

The second hypothesis was that the positive effect of distributed practice would be more pronounced in the long run, that is, the difference between distributed and massed practicing students should be larger in the second test, conducted six weeks after the last practice session, compared to the first test. To test this hypothesis, the performance change between the second and first test was calculated. According to the hypothesis, the change score should be higher (or less negative, in terms of a performance decrease) in the distributed practice group than in the massed practice group. Indeed, the mean of the

[4] Further R-packages we used for data preparation and analysis were (in alphabetical order): BayesFactor [44], gridExtra [2], partykit [31], psych [48], rstan [57] and tidyverse [66].

posterior distribution for the effect of distributed practice was 0.63 (95% credible interval = −0.33 to 1.57) and the evidence ratio in favor of a positive effect of distributed practice on the change score was 10, which can be considered as strong evidence for a positive effect of distributed compared to massed practice in the long run [41]. The performance in the first practice set, however, did not seem to influence the performance change between first and second test: The posterior distribution for the initial practice performance was −0.01 (95% credible interval = −0.19 to 0.16).

Because the first hypothesis, assuming a positive effect of distributed practice in a first test after two weeks, could not be supported by the data, the performance difference between distributed and massed practicing students in the second test was additionally compared directly. For the second test, the mean of the posterior distribution of distributed practice was 0.93 (95% credible interval = −0.21 to 2.07) and the evidence ratio in favor of a positive effect of distributed practice compared to massed practice on the long-term test performance was 18, which is considered strong evidence [41]. In practical terms, this means that the performance advantage for the 7th graders of the distributed practice condition compared to the 7th graders of the massed practice condition is most likely about one point (with a maximum test score of 9.5 points). Similar to the first test performance, the posterior distribution for the effect of the initial practice performance was 0.56 (95% credible interval = 0.35 to 0.78) confirming that students in both practice conditions who scored higher in the first practice set on average also achieved a higher long-term test performance.

We additionally checked exploratory whether the performance benefit of distributed over massed practice in the second test was due to the fact that students in the distributed condition had – if they had performed correctly already in the first test – an additional learning opportunity. Therefore, an additional Bayesian regression analysis was conducted with practice condition, performance in the first test, and the interaction of both variables as predictors of students' performance in the second test. However, there was no evidence for an interaction of learning condition and first test performance on the second test performance (most likely effect of −0.04, 95% credible interval = −0.35 to 0.27). This result suggests that the advantage of the distributed practicing students in the second test cannot be explained by the fact that they had an additional learning opportunity in the first test that markedly improved their performance.

Finally, Bayesian t-tests were used to investigate the relationship between the practice condition and the perceived difficulty regarding the different practice and test sets. However, these tests did not indicate that the practice and test sets were perceived as more or less difficult by the students of one of the two conditions.

3.1. Exploratory analyses

Conditional inference tree models, as proposed by Hothorn, Hornik and Zeileis [29, 30] were used to analyze the potentially moderating effects of mathematical self-efficacy and concentration difficulty on the effect of distributed practice in an exploratory manner. These models are based on recursive binary partitioning and can be used to explore linear and non-linear relationships between a dependent variable and multiple independent variables, which can be measured on different scales. Furthermore, they automatically detect interactions between independent variables and allow for an easy interpretation of complex regression problems by means of the visualization of the fitted decision trees [70].

Basically, the model first checks globally whether the null hypothesis (i.e., that the dependent variable is independent from all tested independent variables) can be rejected or not. If this null hypothesis cannot be rejected, the independent variable with the strongest association to the dependent variable is selected. Based on this selected independent variable, the observations are split into two groups with the criterion of minimizing the *p*-value referring to the difference between the two groups. The result of this process are two different subsamples – distinguished based on the selected independent variable – that show maximally different distributions of the dependent variable. Thereafter, it is tested again within each subsample whether the independent variable is unaffected by the remaining independent variables. If this is not the case (i.e., if the null hypothesis can be rejected again), the independent variable that shows the strongest association to the dependent variable in the currently considered subsample is selected, and the subsample is split based on this variable in a way that minimizes the *p*-value referring to the difference between these two groups. This process is repeated for each resulting level of subsample until the null hypothesis of independence cannot be rejected any longer. Note that each split of the sample does not result in equal-sized subsamples but in subsamples that maximally differ with regard to the dependent variable.

The conditional inference tree models were calculated for each test time separately. Each of the models estimated for the presented study included the practice condition and the potential moderators (i.e., mathematical self-efficacy and self-rated concentration difficulty) as independent variables, and the respective test performance as dependent variable. In none of the two models a significant interaction of the learner characteristics and the effect of distributed practice on the test performances was revealed.

Two additional conditional inference tree models were calculated (again, one for each test time) to investigate whether the effect of distributed practice depends on the initial performance level, as results by Hirsch et al. [28] and Saxon [53] suggested. In these models, practice condition and performance in the first practice set served as independent variables and the respective test performance as dependent variable. These analyses revealed that in the second test, conducted six weeks after the last practice session, distributed practicing students performed better than massed practicing students in particular if their performance in the first practice set was in a medium score range (i.e., between 3 and 7 points out of 9.5 points, s. Fig. 2). In the low and high score ranges, the test performance was rather independent from the practice condition (i.e., low performers / massed: $M = = 3.5$, $SD = 3.2$; low performers / distributed: $M = = 2.9$, $SD = 2.0$; high performers / massed: $M = = 8$, $SD = 1.8$; high performers / distributed: $M = = 8.3$, $SD = 1.5$). This result indicates that the effect of distributed practice, which was supported by the Bayesian regression model above, may be mainly due to the students in a medium performance range. This interaction did not appear for the first test.

4. Discussion

Despite the strong empirical indicators that suggest a positive effect of distributed practice on the long-term memory of verbal content (e.g., [12]), little is known about the effect of distributed practice on the learning of mathematics. Though few studies found positive effects of distributed practice using mathematical content (e.g., [3, 50, 51, 54]), the question if and how distributed practice affects performance in more complex domains like procedural mathematical learning still lacks sufficient empirical support.

In the current study, the effect of distributed practice with feedback on mathematical learning was examined in 7th grade. Contrary to the central hypotheses and the results by Barzagar Nazari and Ebersbach [3], there was no evidence for a positive effect of distributed practice compared to massed practice on the intermediate performance of the students, tested after two weeks. However, regarding the long-term performance tested six weeks after the last practice session, strong evidence for a positive effect of distributed practice was revealed. The latter finding is similar to the one reported by Barzagar Nazari and Ebersbach [3]. Though it is theoretically assumed that the effect of distributed practice emerges especially in the long term, it is not clear why there was no empirical evidence for a positive effect two weeks after the last practice set in the present study at all. Most of the previous

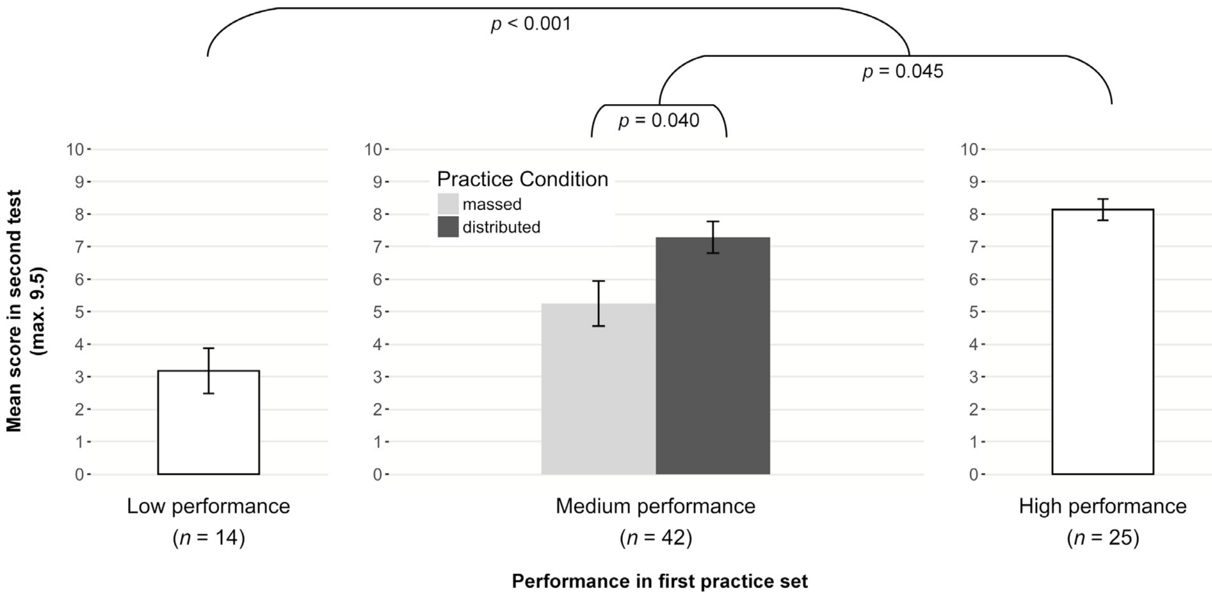


Fig. 2. Interaction between students' performance in the first practice phase and learning condition concerning their long-term test performance (i.e., six weeks after the last practice session). The sample sizes within the three performance range groups were as follows: low performers: $n = = 7$ for massed practice, $n = = 7$ for distributed practice; medium performers: $n = = 16$ for massed practice, $n = = 26$ for distributed practice; high performers: $n = = 13$ for massed practice, $n = = 12$ for distributed practice.

studies only investigated the effect of distributed with similar retention intervals of up to one or two weeks and still found positive effects of distributed practice. Perhaps, feedback in the present study boosted the performance in both conditions so that the performance was still similar after two weeks, overshadowing the effect of practice condition that became apparent only in the long term. No feedback was provided in an earlier study by Barzagar Nazari and Ebersbach [3], where an effect of distributed practice became apparent also in an intermediate test among 7th graders. Future studies should try to implement distributed practice of complex material with feedback with a greater variance of retention intervals to further investigate its effect.

Initial exploratory analyses did not indicate any moderating effects of mathematical self-efficacy or concentration difficulty on the effectivity of distributed practice. However, further exploratory analyses suggested that the positive effect of distributed practice in the second test may be largely ascribed to an effect for students in the medium performance range (i.e., between 3 and 7 points out of 9.5 points in the first practice set). Among these students, those who practiced distributed performed significantly better in the long-term performance test after six weeks than those who practiced in a massed fashion. For students in the very low or very high performance range, no such pattern was revealed.

These results, though exploratory, are interesting in that they suggest that regarding mathematical learning there might be subgroups of learners whose performance is rather unaffected by the distributed or massed practice strategy, possibly because they are either not able to grasp the underlying concept at all (i.e., the low performance group) or because they grasp it that quickly that the practice strategy does not matter, too (i.e., the high-performance group). In each case, the performance is relatively stable and independent from practice, and hence independent from the practice strategy. It is only the students in the medium performance range who might benefit from distributed practice. Interestingly, the aforementioned studies examining the effect of mixed review of mathematical homework by Hirsch et al. [28] and Saxon [53] point into a similar direction since the authors found that low to medium performing students profited more from distributed (and interleaved) practice than high performing students. More studies on distributed practice of mathematical content should be conducted to further investigate the relationship between students' initial ability or baseline performance and the effect of the practice strategy.

Regarding mathematical self-efficacy and concentration difficulty, there are different reasons that may explain the missing interactions with the effect of distributed practice. First, it could be that there is a relationship between the discussed characteristics and the effect of distributed practice, but the effects are too small to detect them with the given sample size. Second, it is possible that in the investigated context there are actually no effects of mathematical self-efficacy and concentration difficulty on the effectivity of distributed practice because this strategy works independent from these two traits. This would not mean, however, that the interaction cannot be expected under any other circumstances. On the contrary, it is quite possible that when practice relies more on self-regulated learning, motivational and cognitive traits become increasingly important. For example, they could additionally influence the actual use of certain strategies [4, 60]. Another potential factor could be the general difficulty of the tasks. For example, it was assumed that especially for those who struggle with the practice exercises, mathematical self-efficacy and concentration difficulty could affect the effect of distributed practice. The descriptive practice performance indicated that the exercises in this study were not particularly hard. It is possible that with more demanding exercises, these relationships could be revealed.

Another issue raised in the current study is whether distributed practice is in fact a difficulty for learners, as assumed for instance by Bjork [6]. Descriptively, there seemed to be no differences concerning the practice performance of massed and distributed practicing students in the present study (see Table 2). Additionally, exploratory analyses did not indicate that the perceived difficulty of the practice and test sets differed between the conditions. However, indicators that distributed practice is regarded as effortful strategy by learners is reflected by the fact that the majority of students does not think that it is an effective study strategy [42] and therefore tends to resort to massed practice and devote their time for learning on the two days prior to the test [25, 59]. Thus, students show hardly any insights into the actual benefit of distributed practice and feel more confident with massed practice [7, 36]. Studies like the present one may serve to undeceive learners and teachers about the effectivity of distributed practice for learning in school.

One potential limitation of the current study concerns the manipulation of the test delays within subjects. Retrieval attempts might evoke a testing effect that is usually expressed by enhanced learning and longer retention periods ([22]; for a meta-analysis, [52]). Thus, the first test after two weeks might have affected the memory performance in the second test after six weeks. Such an effect would be uncritical if it would apply to both learning conditions in the same manner. Only if the learning condition and testing would interact, results might be blurred. A testing effect cannot fully be ruled out due to the within-subjects design.

Another limitation refers to the exploratory finding that distributed practice worked in particular for students in a medium performance range. In fact, with regard to mathematical knowledge, there is usually a broad variance within the same age group or class. It could be assumed that the intervals between practice sessions were not optimal very low and very high performing students to uncover an effect of distributed practice also in these performance groups. Studies involving verbal learning material that also implemented distributed practice with an expanding interval suggest that an expanding interval yields better effects than a uniform or contracting interval when forgetting is highly probable due to the task or material [58] or due to learners' prior skills [63]: An expanding interval allows learners to consolidate their knowledge by initially shorter intervals that promote successful retrieval. Accordingly, it could be inferred that low performing learners in our study would have profited from shorter (expanding) intervals when practicing mathematical skills, while high performing learners would have profited from longer (expanding) intervals. It would be worthwhile to investigate this interaction further with mathematical material. One might also think of implementing distributed practice in an adaptive way, accounting for the prior knowledge of students. In doing so, one could discriminate students' prior knowledge from their more general mathematical performance level and determine for which particular group of students distributed practice works best. Another account could be to test the effect of distributed practice on novel mathematical contents students have absolutely no prior knowledge about.

Furthermore, there were some deviations from the scheduled procedure for a subsample of students who were tested late. However, it should be noted that these deviations mostly concerned the second test, originally scheduled six weeks after the last practice. The fact that 16 of the distributed practicing students were tested one week later should not have influenced the results markedly.

Additionally, as in the previous study by Barzagar Nazari and Ebersbach [3], only one specific topic was covered in the current study and generalizations concerning other topics based on this study should be drawn with caution. The topic stochastics was picked for different reasons. First, the topic is part of the curriculum for Grade 7, but regularly covered at the end of the school year or not at all. That is, with this topic the chance was high to find enough courses who had not already covered the topic at the beginning of the study period. Second, stochastics as a mathematical topic is relatively independent from other domains such as analysis or geometry. That is, the courses were roughly comparable in this specific topic as their performance was hardly influenced by the topics covered in the regular mathematics class before the start of the study. Third, the goal was to expand the results of the previous study on the same topic, so roughly the same material was used. Nevertheless, research on distributed practice with mathematical content would benefit from more studies using a broader range of material.

5. Conclusions

The current study contributes to answer the questions of why and under which circumstances distributed practice proves a useful learning strategy in realistic learning contexts, even beyond learning of rather simple verbal content. We argue that while the effect of distributed practice emerges for roughly all learners if simple verbal content or mathematical routines are practiced [11, 50, 51], the effect of distributed practice on learning of more complex content, like the acquisition and application of more advanced mathematical procedures as in the present study, might especially be effective for students in the medium performance range (see also [28, 53]). These assumptions, however, are based on exploratory analyses and need further empirical confirmation. In sum, distributed practice remains a promising learning strategy and more studies on a broad range of content and learners could help to deepen our understanding of when and why it works.

6. Ethical statement

We confirm that any aspect of the work covered in this manuscript that has involved either experimental animals or human patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

We would like to thank Wiebke Eden, Carina Haines, Ulrike Haßfurther, Fabian Kinzl, Amrei Schmieg and Christina Vogelsang for their support in conducting the experiments and Jan Bender, Klara Schelbert and Maike Siegle for also teaching the participating courses.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tine.2019.100122.

Appendix A. Exercise examples

Note: The complete material is provided in German language online (https://osf.io/d542q/?view_only=7896f90d809140d08b777fba6d564454).

One practice set consisted of three exercises similar to the following:

Exercise 1

Class 8a draws by lot which student has to start with the poem presentation. The teacher writes all the names of the 27 students on little notes. Afterwards he draws one blindly. What are the chances for Lisa to go first?

Answer: ________________

Exercise 2

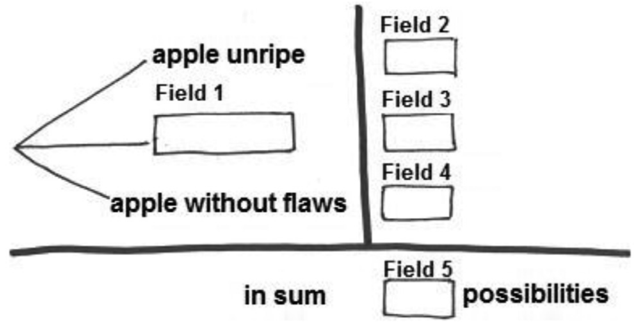
Drawing one card out of a deck with 52 cards, how high is the probability to…

a) …draw one of the four jacks? Answer:________________
b) …draw one of the two red queens? Answer:________________

c) …draw one of the 13 cards of hearts? Answer:________________

Exercise 3

There are 80 apples in one basket. 12 of them are unripe, 8 contain a worm. Label the missing pieces of the tree diagram and determine the probability of grabbing an apple with flaws (containing a worm or unripe).



Field 1:________________
Field 2: ________________
Field 3:________________
Field 4: ________________
Field 5:________________

Probability of grabbing an apple with flaws (unripe or containing a worm): ________________

Appendix B. Exclusion procedure

There were several deviations from the scheduled procedure that led to the exclusion of one course. In one school the teacher forgot the last session of the distributed practicing students and a new appointment had to be arranged. This led to a one week longer second test interval for 16 students. That is, instead of six weeks, they were tested seven weeks after the last practice session. In addition, in another school there was a complete course with 27 students whose long-term performance was tested seven and nine days too early (i.e., massed and distributed condition, respectively) due to the summer holidays.

The first deviation of being tested one week too late concerned only students of the distributed practice condition. Keeping these students in the sample poses the risk to *underestimate* the expected effect of distributed practice as they might have forgotten more of the practiced contents than the regularly tested students. However, for the sake of a larger sample and more balanced group sizes we decided to leave the respective group in the analyzed sample. In contrast, additionally including the students tested too early would have caused a broad range in long-term test lags (between 33 days after the last practice set for the early tested distributed group and 49 days for the late tested distributed group). Because we wanted to reduce the variance in test lags and were especially interested in long-term performance, we decided to remove the early tested students and to maintain the group of distributed practicing students who were tested one week late.

By removing the early tested students, the lag between the last practice and the long-term performance test was between the scheduled six weeks and seven weeks for the 16 distributed practice condition students being tested especially late. Additionally, five students of the massed practice condition were tested two days too late due to other school activities. They were included, too.

References

[1] Arslan, R.C., & Tata, C.S. (2017). formr.org survey software (Version 0.16.13).
[2] Auguie, B. (2017). gridExtra: miscellaneous functions for "grid" graphics (R package version 2.3). Retrieved from https://CRAN.R-project.org/package=gridExtra.
[3] K. Barzagar Nazari, M. Ebersbach, Distributing mathematical practice of third and seventh graders: applicability of the spacing effect in the classroom, J. Appl. Cognit. Psychol. 33 (2019) 288–298, https://doi.org/10.1002/acp.3485.
[4] K. Barzagar Nazari, M. Ebersbach, Distributed practice: rarely realized in self-regulated mathematical learning, Front. Psychol. 9 (2018) 2170, https://doi.org/10.3389/fpsyg.2018.02170.
[5] R.A. Bjork, Retrieval as a memory modifier: an interpretation of negative recency and related phenomena, in: R.L. Solso (Ed.), Information Processing and Cognition: The Loyola Symposium, Lawrence Erlbaum, 1975, pp. 123–144.
[6] R.A. Bjork, Memory and metamemory considerations in the training of human beings, in: J. Metcalfe, A.P. Shimamura (Eds.), Metacognition: Knowing about Knowing, The MIT Press, Cambridge, 1994, pp. 185–205.
[7] R.A. Bjork, J. Dunlosky, N. Kornell, Self-regulated learning: beliefs, techniques, and illusions, Annu. Rev. Psychol. 64 (1) (2013) 417–444, https://doi.org/10.1146/annurev-psych-113011-143823.
[8] S. Boerner, G. Seeber, H. Keller, P. Beinborn, Lernstrategien und Lernerfolg im Studium: zur Validierung des LIST bei berufstätigen Studierenden, Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie 37 (1) (2005) 17–26, https://doi.org/10.1026/0049-8637.37.1.17.
[9] L. Budé, T. Imbos, M.W. van de Wiel, M.P. Berger, The effect of distributed practice on students' conceptual understanding of statistics, Higher Educ. 62 (1) (2010) 69–79, https://doi.org/10.1007/s10734-010-9366-y.
[10] P.-C. Bürkner, brms: an r package for Bayesian multilevel models using Stan, J. Stat. Softw. 80 (1) (2017) 1–28, https://doi.org/10.18637/jss.v080.i01.
[11] S.K. Carpenter, N.J. Cepeda, D. Rohrer, S.H.K. Kang, H. Pashler, Using spacing to enhance diverse forms of learning: review of recent research and implications for instruction, Educ. Psychol. Rev. 24 (3) (2012) 369–378, https://doi.org/10.1007/s10648-012-9205-z.
[12] N.J. Cepeda, H. Pashler, E. Vul, J.T. Wixted, D. Rohrer, Distributed practice in verbal recall tasks: a review and quantitative synthesis, Psychol. Bull. 132 (3) (2006) 354–380, https://doi.org/10.1037/0033-2909.132.3.354.
[13] N.J. Cepeda, E. Vul, D. Rohrer, J.T. Wixted, H. Pashler, Spacing effects in learning. A temporal ridgeline of optimal retention, Psychol. Sci. 19 (11) (2008) 1095–1102, https://doi.org/10.1111/j.1467-9280.2008.02209.x.
[14] O. Chen, J.C. Castro-Alonso, F. Paas, J. Sweller, Extending cognitive load theory to incorporate working memory resource depletion: evidence from the spacing effect, Educ. Psychol. Rev. (2017) 1–19, https://doi.org/10.1007/s10648-017-9426-2.
[15] F.I.M. Craik, R.S. Lockhart, Levels of processing: a framework for memory research, J. Verbal. Learn. Verbal. Behav. 11 (6) (1972) 671–684, https://doi.org/10.1016/S0022-5371(72)80001-X.
[16] L.J. Cronbach, R.E. Snow, Aptitudes and Instructional Methods: A Handbook for Research on Interactions, Irvington, Oxford, England, 1977.
[17] R.G. Crowder, Principles of Learning and Memory, John Wiley & Sons, New York, 1976.
[18] P.F. Delaney, N.R. Godbole, L.R. Holden, Y. Chang, Working memory capacity and

the spacing effect in cued recall, Memory 26 (6) (2018) 784–797, https://doi.org/10.1080/09658211.2017.1408841.

[19] P.F. Delaney, P.P.J.L. Verkoeijen, A. Spirgel, Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature, in: B.H. Ross (Ed.), The Psychology of Learning and Motivation, 53 Academic Press, London, 2010, pp. 63–147.

[20] F.N. Dempster, The spacing effect: a case study in the failure to apply the results of psychological research, Am. Psychol. 43 (8) (1988) 627–634, https://doi.org/10.1037/0003-066X.43.8.627.

[21] J.J. Donovan, D.J. Radosevich, A meta-analytic review of the distribution of practice effect: now you see it, now you don't, J. Appl. Psychol. 84 (5) (1999) 795–805, https://doi.org/10.1037/0021-9010.84.5.795.

[22] J. Dunlosky, K.A. Rawson, E.J. Marsh, M.J. Nathan, D.T. Willingham, Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology, Psychol. Sci. Publ. Interest 14 (1) (2013) 4–58, https://doi.org/10.1177/1529100612453266.

[23] R. Fernández-Alonso, J. Suárez-Álvarez, J. Muñiz, Adolescents' homework performance in mathematics and science: personal factors and teaching practices, J. Educ. Psychol. 107 (2015) 1–11, https://doi.org/10.1037/edu0000032.

[24] A.M. Glenberg, Component-levels theory of the effects of spacing of repetitions on recall and recognition, Mem. Cognit. 7 (2) (1979) 95–112, https://doi.org/10.3758/BF03197590.

[25] M.K. Hartwig, J. Dunlosky, Study strategies of college students: are self-testing and scheduling related to achievement, Psychon. Bull. Rev. 19 (1) (2011) 126–134, https://doi.org/10.3758/s13423-011-0181-y.

[26] J. Hattie, Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement, Routledge, London, 2008.

[27] D.L. Hintzman, Theoretical implications of the spacing effect, in: R.L. Solso (Ed.), Theories in Cognitive psychology: The Loyola Symposium, Lawrence Erlbaum, Oxford, England, 1974.

[28] C.R. Hirsch, S.F. Kapoor, R.A. Laing, Alternative models for mathematics assignments, Int. J. Math. Educ. Sci. Technol. 13 (3) (1982) 243–252, https://doi.org/10.1080/0020739820130301.

[29] T. Hothorn, K. Hornik, A. Zeileis, Unbiased recursive partitioning: a conditional inference framework, J. Comput. Graph. Stat. 15 (3) (2006) 651–674 10.1198/106186006x133933.

[30] T. Hothorn, K. Hornik, A. Zeileis, ctree: conditional inference trees, Comprehens. R Arch. Netw. (2015) Retrieved from https://rdrr.io/rforge/partykit/f/inst/doc/ctree.pdf.

[31] T. Hothorn, A. Zeileis, partykit: a modular toolkit for recursive partytioning in R, J. Mach. Learn. Res. 16 (2015) 3905–3909.

[32] C. Janiszewski, H. Noel, A.G. Sawyer, A meta-analysis of the spacing effect in verbal learning: implications for research on advertising repetition and consumer memory, J. Consum. Res. 30 (1) (2003) 138–149, https://doi.org/10.1086/374692.

[33] M. Jerusalem, L. Satow, Schulbezogene Selbstwirksamkeitserwartung, Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der Psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen, Freie Universität Berlin, Berlin, 1999 Retrieved from http://www.psyc.de/skalendoku.pdf.

[34] S.H.K. Kang, Spaced repetition promotes efficient and effective learning: policy implications for instruction, Policy Insights Behav. Brain Sci. 3 (2016) 12–19, https://doi.org/10.1177/2372732215624708.

[35] I.V. Kapler, T. Weston, M. Wiseheart, Spacing in a simulated undergraduate classroom: long-term benefits for factual and higher-level learning, Learn. Instr. 36 (2015) 38–45, https://doi.org/10.1016/j.learninstruc.2014.11.001.

[36] N. Kornell, Optimising learning using flashcards: spacing is more effective than cramming, Appl. Cogn. Psychol. 23 (9) (2009) 1297–1317, https://doi.org/10.1002/acp.1537.

[37] J. Kruschke, Doing Bayesian data Analysis: A Tutorial With R, JAGS, and Stan, 2nd ed., Academic Press, London, 2015.

[38] C.E. Küpper-Tetzel, Understanding the distributed practice effect. Strong effects on weak theoretical grounds, Zeitschrift für Psychologie 222 (2) (2014) 71–81, https://doi.org/10.1027/2151-2604/a000168.

[39] C.E. Küpper-Tetzel, E. Erdfelder, O. Dickhäuser, The lag effect in secondary school classrooms: enhancing students' memory for vocabulary, Instr. Sci. 42 (3) (2014) 373–388, https://doi.org/10.1007/s11251-013-9285-2.

[40] C.E. Küpper-Tetzel, I.V. Kapler, M. Wiseheart, Contracting, equal, and expanding learning schedules: the optimal distribution of learning sessions depends on retention interval, Mem. Cognit. 42 (5) (2014) 729–741, https://doi.org/10.3758/s13421-014-0394-1.

[41] M.D. Lee, E.-J. Wagenmakers, Bayesian Cognitive Modeling: A Practical Course, Cambridge University Press, Cambridge, 2013.

[42] J. McCabe, Metacognitive awareness of learning strategies in undergraduates, Mem. Cognit. 39 (3) (2011) 462–476, https://doi.org/10.3758/s13421-010-0035-2.

[43] M.A. McDaniel, A.C. Butler, A contextual framework for understanding when difficulties are desirable, in: A.S. Benjamin (Ed.), Successful Remembering and Successful Forgetting: A festschrift in Honor of Robert A. Bjork, Psychology Press, New York, 2011, pp. 175–198.

[44] Morey, R.D., & Rouder, J.N. (2015). BayesFactor: computation of Bayes factors for common designs (R package version 0.9.12-2). Retrieved fromhttps://CRAN.R-project.org/package=BayesFactor.

[45] Project Team / Schmitz, C, LimeSurvey: An Open Source Survey Tool, LimeSurvey Project, Hamburg, 2012 Retrieved from http://www.limesurvey.org.

[46] R. Core Team, R: A language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, 2016.

[47] C.P. Rea, V. Modigliani, The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists, Hum. Learn. 4 (1) (1985) 11–18.

[48] W. Revelle, psych: Procedures for Personality and Psychological Research (R Package Version 1.6.12), Northwestern University, Evanston, Illinois, USA, 2016 Retrieved from https://CRAN.R-project.org/package=psych.

[49] D. Rohrer, The effects of spacing and mixing practice problems, J. Res. Math. Educ. 40 (1) (2009) 4–17.

[50] D. Rohrer, K. Taylor, The effects of overlearning and distributed practice on the retention of mathematics knowledge, Appl. Cogn. Psychol. 20 (9) (2006) 1209–1224, https://doi.org/10.1002/acp.1266.

[51] D. Rohrer, K. Taylor, The shuffling of mathematics problems improves learning, Instr. Sci. 35 (6) (2007) 481–498, https://doi.org/10.1007/s11251-007-9015-8.

[52] C.A. Rowland, The effect of testing versus restudy on retention: a meta-analytic review of the testing effect, Psychol. Bull. 140 (2014) 1432–1463, https://doi.org/10.1037/a0037559.

[53] J. Saxon, Incremental development: a breakthrough in mathematics, Phi Delta Kappan 63 (7) (1982) 482–484.

[54] G.M. Schutte, G.J. Duhon, B.G. Solomon, B.C. Poncy, K. Moore, B. Story, A comparative analysis of massed vs. distributed practice on basic math fact fluency growth rates, J. Sch. Psychol. 53 (2) (2015) 149–159, https://doi.org/10.1016/j.jsp.2014.12.003.

[55] R. Seabrook, G.D.A. Brown, J.E. Solity, Distributed and massed practice: from laboratory to classroom, Appl. Cogn. Psychol. 19 (1) (2005) 107–122, https://doi.org/10.1002/acp.1066.

[56] R.E. Snow, Aptitude-treatment interaction as a framework for research on individual differences in learning, in: P.L. Ackerman, R.J. Sternberg, R. Glaser (Eds.), A Series of Books in Psychology. Learning and Individual Differences: Advances in Theory and Research, W H Freeman/Times Books/ Henry Holt & Co, New York, 1989, pp. 13–59.

[57] Stan Development Team. (2018). RStan: the r interface to stan (R package version 2.17.3). Retrieved fromhttp://mc-stan.org/.

[58] B.C. Storm, R.A. Bjork, J.C. Storm, Optimizing retrieval as a learning event: when and why expanding retrieval practice enhances long-term retention, Mem. Cognit. 38 (2010) 244–253, https://doi.org/10.3758/MC.38.2.244.

[59] R. Taraban, W.S. Maki, K. Rynearson, Measuring study time distributions: implications for designing computer-based courses, Behav. Res. Method. Instrum. Comput. 31 (2) (1999) 263–269, https://doi.org/10.3758/BF03207718.

[60] M. Theobald, H. Bellhäuser, M. Imhof, Identifying individual differences using logfile analysis: distributed learning as mediator between conscientiousness and exam grades, Learn. Individ. Differ. 65 (2018) 112–122, https://doi.org/10.1016/j.lindif.2018.05.019.

[61] S.J. Thios, P.R. D'Agostino, Effects of repetition as a function of study-phase retrieval, J. Verbal Learn. Verbal Behav 15 (5) (1976) 529–536, https://doi.org/10.1016/0022-5371(76)90047-5.

[62] T.C. Toppino, E. Gerbier, About practice: repetition, spacing, and abstraction, in: B.H. Ross (Ed.), The Psychology of Learning and Motivation, 60 Academic Press, Waltham, MA, 2014, pp. 113–189.

[63] T.C. Toppino, H.-A. Phelan, E. Gerbier, Level of initial training moderates the effects of distributing practice over multiple days with expanding, contracting, and uniform schedules: evidence for study-phase retrieval, Mem. Cognit. 46 (2018) 969–978, https://doi.org/10.3758/s13421-018-0815-7.

[64] P.P.J.L. Verkoeijen, R.M.J.P. Rikers, H.G. Schmidt, Detrimental influence of contextual change on spacing effects in free recall, J. Exp. Psychol. 30 (4) (2004) 796–800.

[65] P.P.J.L. Verkoeijen, R.M.J.P. Rikers, H.G. Schmidt, Limitations to the spacing effect: demonstration of an inverted u-shaped relationship between interrepetition spacing and free recall, Exp. Psychol. 52 (2005) 257–263.

[66] Wickham, H. (2017). tidyverse: easily install and load "tidyverse" packages (R package version 1.1.1). Retrieved fromhttps://CRAN.R-project.org/package=tidyverse.

[67] K.P. Wild, U. Schiefele, Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens, Zeitschrift für Differentielle und Diagnostische Psychologie 15 (1994) 185–200.

[68] M.A. Yazdani, E. Zebrowski, Spaced reinforcement: an effective approach to enhance the achievement in plane geometry, Journal of Mathematical Sci. Math. Educ. 1 (2006) 37–43.

[69] D.R. Young, F.S. Bellezza, Encoding variability, memory organization, and the repetition effect, J. Exp. Psychol. 8 (6) (1982) 545–559.

[70] A. Zeileis, T. Hothorn, K. Hornik, Model-based recursive partitioning, J. Comput. Graph. Stat. 17 (2) (2008) 492–514 0.1198/10618600SX319331.

[71] B.J. Zimmerman, Self-efficacy and educational development, in: A. Bandura (Ed.), Self-efficacy in Changing Societies, Cambridge University Press, New York, 1995, pp. 202–231.