



## RESEARCH ARTICLE

WILEY

# Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom

Katharina Barzagar Nazari  | Mirjam Ebersbach

University of Kassel, Kassel, Germany

**Correspondence**

Katharina Barzagar Nazari, University of Kassel, Department of Psychology, Kassel D-34127, Germany.

Email: k.barzagar@uni-kassel.de

**Funding information**

LOEWE Excellence Initiative of the State of Hesse, Germany (project "Desirable Difficulties in Learning", 2015-2018)

**Summary**

We examined the effect of distributed practice on the mathematical performance of third and seventh graders ( $N = 213$ ) in school. Students first received an introduction to a mathematical topic, derived from their curriculum. Thereafter, they practiced in one of two conditions. In the massed condition, they worked on three practice sets in 1 day. In the distributed condition, they worked on one practice set per day for 3 consecutive days. Bayesian analyses of the performance in two follow-up tests 1 and 6 weeks after the last practice set revealed a positive effect of distributed practice as compared with massed practice in Grade 7. In Grade 3, a positive effect of distributed practice was supported by the data only in the test 1 week after the last practice set. The results suggest that distributed practice is a powerful learning tool for both elementary and secondary school students in the classroom.

**KEYWORDS**

desirable difficulties, distributed practice, mathematics, school (elementary school; secondary school), spacing effect

## 1 | INTRODUCTION

Practicing is fundamental for consolidating learned skills and knowledge. Whether learning a language, a motor skill, or a mathematical procedure, students must practice their developing skills to maintain and improve their performance. School subjects, such as mathematics, also rely on practice. Often, students acquire a new mathematical concept or procedure and practice it with similar tasks. That is, solutions are not learned by heart, but the application of mathematical procedures is practiced until a certain confidence level is reached. Maintaining acquired knowledge and skills is important in mathematics, because subsequent topics often build on previously learned mathematical content. Thus, forgetting may impede not only present but also future achievement. The aspect of long-term retention is especially important in school because blocked courses are common. For example, mathematical analysis is taught for a few weeks in a blocked course, then tested, followed by several weeks of geometry. The next

topic requiring analysis may not occur for weeks or even months. Thus, learned content is often not accessed for long periods before it is retrieved and applied again. Given the cognitive challenges in this type of learning environment, knowing which learning strategies enhance long-term retention and minimize forgetting is essential for preserving knowledge even beyond intermediately covered topics.

Several processes that have been found to boost encoding and retrieval were subsumed under the term *desirable difficulties* (Bjork, 1994, 2013). Learning strategies relying on these desirable difficulties are supposed to make the learning process harder for the learner but contribute to the consolidation of knowledge over longer periods of time. Within cognitive psychology research, these strategies have received much attention (e.g., Carpenter, Pashler, Wixted, & Vul, 2008; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Rohrer, Dedrick, & Stershic, 2015; Toppino & Gerbier, 2014).

One widely investigated and potent learning strategy related to desirable difficulties is the *spacing* of learning sessions, also known

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors Applied Cognitive Psychology Published by John Wiley & Sons Ltd

as *distributed practice* (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012). The terms *spacing* and *distributed practice* are often used interchangeably. The term *distributed practice* is used in the current paper because it focuses on the spacing of practice. Generally, *distributed practice* implies that a given practice duration is distributed across several learning sessions, whereas the same learning duration is massed within one learning session during *massed practice* (Bjork & Bjork, 2011). Long-term retention performance after distributed practice exceeds performance after massed practice (Küpper-Tetzel, 2014). The effect of distributed practice is widely studied and considered to be robust (Carpenter et al., 2012).

Different theoretical approaches have tried to explain the positive effect of distributed practice on long-term retention and presumably, different mechanisms contribute to this effect simultaneously (Küpper-Tetzel, 2014; Toppino & Gerbier, 2014). There are three main theoretical approaches: deficient processing theories (Challis, 1993; Craik & Lockhart, 1972; Crowder, 1976; Hintzman, 1974; Toppino & Gerbier, 2014), theories based on study phase retrieval mechanisms (Bjork, 1975; Braun & Rubin, 1998; Küpper-Tetzel, 2014; Thios & D'Agostino, 1976), and encoding variability theories (Glenberg, 1979; Küpper-Tetzel, 2014; Toppino & Gerbier, 2014; Tulving & Thomson, 1973). Although originally referring to verbal learning, all three approaches appear to be suited to explain effects of distributed practice in procedural mathematical learning as well.

Given the persuasive evidence regarding the positive effect of distributed practice for verbal memory recall, ample grounds exist to expect that distributed practice might also work for learning mathematical procedures—but this assumption has yet to be investigated.

### 1.1 | Distributed practice in the classroom

In the classroom, distributed learning already occurs in some regards (e.g., a topic is frequently and repetitively taught for several weeks), whereas in other regards, it is not as common. For example, homework is mostly used to practice only the information most recently learned. In our experience, teachers also rarely encourage their students to systematically distribute their homework across several days. Practice is usually massed within relatively short periods of time, and even when learning and practice happens to be distributed, this strategy is not implemented systematically. That is, assuming that distributed practice is superior to massed practice, several possible applications of distributed practice in the educational context remain untapped (Kang, 2016).

In addition, the learning conditions in school are much less controllable than in the laboratory, where most of the studies regarding distributed practice take place, because students in school, for example, practice in potentially noisy environments. Given the lack of ecologically valid studies in this line of research, the field needs studies that investigate whether the promising laboratory results on the effects of learning strategies related to desirable difficulties, or distributed practice in particular, also hold in applied learning settings. Only a few studies on distributed practice, however, have adopted learning contents based on actual curricula (e.g., Kang, 2016; Kapler, Weston, & Wiseheart, 2015), for example, by investigating the distributed practice effect on mathematical procedures in classrooms or college

courses (for an overview, see Kang, 2016). Moreover, studies investigating distributed practice effects have addressed only particular aspects, for example, by delivering information verbally and focusing on declarative knowledge only (Carpenter et al., 2012; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Verkoeijen, & Spigel, 2010; Son & Simon, 2012; Toppino & Gerbier, 2014). Though findings from these studies likely can be generalized to school subjects that require much declarative memory recall (e.g., foreign language learning), they fail to address the extent to which distributed practice shows similar positive effects in school subjects that rely more on implicit, procedural memory such as mathematics. Procedural mathematical learning differs from memorization and can be defined as “the ability to execute action sequences (i.e., procedures) to solve problems”—in contrast to conceptual knowledge that can be conceived as “knowledge of concepts” (Rittle-Johnson & Schneider, 2015, pp. 1119–1120).

Despite the growing interest in examining distributed practice effects on mathematical learning (Rohrer & Taylor, 2006, 2007), or distributed practice in authentic educational settings (Goossens et al., 2016; Kapler et al., 2015), both aims have rarely been combined. To the best of our knowledge, only two studies have specifically investigated distributed practice effects on mathematical learning within the school context. In the study by Schutte et al. (2015), third graders practiced addition problems each day for 4 min over the course of 19 school days. They either practiced massed each day for 4 min, or they distributed the practice duration over two 2-min sessions, or over four 1-min sessions on the same day. Students in both distributed practice groups outperformed those who practiced each day in a massed fashion in the immediate posttest and a maintenance test conducted 10 days after the end of the intervention. These results notwithstanding, essentially all students were in a distributed practice condition, because practice in all groups was distributed across 19 days, and differences only reflect the effects of distribution within each day. In two experiments conducted by Chen, Castro-Alonso, Paas, and Sweller (2017), fourth and fifth graders who practiced a mathematical procedure distributed over 3 days achieved higher performance scores than students who practiced massed on 1 day. However, for theoretical reasons, the authors varied the retention interval between the conditions: The massed-practice students were tested immediately following the practice phase, whereas the distributed-practice students were tested the day after their last practice session. Additionally, intermediate and long-term retention was not tested. Thus, more research is needed to further develop the paradigm and to corroborate these promising first results on the effect of distributed practice on mathematical learning in school. To address this need, we investigated distributed practice effects on the learning of mathematical procedures in the classroom.

### 1.2 | Research objective and hypotheses of our study

The purpose of our study was to examine the effect of distributed practice on procedural knowledge (i.e., learning with exercises) in mathematics in the classroom. A brief mathematical lesson was developed for third and seventh graders based on their regular curricula. Students were not only required to learn facts taught in the classroom

by heart but also worked on slightly different but conceptually similar math exercises either by massed or distributed practice. We expected that students in the distributed practice condition would outperform students in the massed practice condition in the test conducted 1 week after the last practice session and that this difference would be even more marked in the 6-week follow-up test, because the positive effect of distributed practice becomes usually more evident in the long run (Küpper-Tetzl, 2014).

In sum, our central hypotheses were as follows: For both grades included in our study, we expected that students practicing a mathematical topic in a distributed fashion would perform better in a test conducted 1 week after the last practice than students practicing with the same exercises in a massed fashion. Furthermore, the positive effect of distributed practice compared with massed practice on the test performance should be even bigger after 6 weeks than after 1 week in both grades.

## 2 | METHOD

### 2.1 | Participants

In total, 141 third graders from five elementary schools and 171 seventh graders from four secondary schools in Germany were recruited. These age groups were selected, because they cover two school levels (i.e., primary and secondary school). The schools were located in or around a middle-sized city in Germany in neighborhoods with a medium socioeconomic status. Parents were informed about the aims of the study and signed a consent form to allow the participation of their children. Participation was voluntary and could be terminated by the children at any time in the study. One elementary school class with 16 students was excluded because the students failed to follow the instructions sufficiently and frequently talked during practice and test sessions, even though talking was forbidden. The decision to remove this specific group was made while the practice sessions took place and before any data were examined.

To be included in the analyses, students were required to be present during all lesson, practice, and test sessions. After exclusions, the data from 95 third graders (51 females, 44 males;  $M_{\text{age}} = 9$  years, 6 months; age range: 8–11 years) and 118 seventh graders (66 females, 52 males;  $M_{\text{age}} = 13$  years, 5 months; age range: 12–14 years) were analyzed.

The seventh graders were attending mathematical classes of different course levels, aimed at different graduation qualifications when finishing high school. The study included 55 students attending an intermediate level math class (aimed at the "Mittlere Reife" certificate) and 63 students attending higher level math classes (aimed at the "Abitur" qualification to study at a university). No such differentiation occurred for the third-grade classes. All participating students spoke fluent German.

### 2.2 | Design

Practice schedule was manipulated between subjects. Students either practiced the learning content (i.e., mathematical procedures) *massed*

on 1 day or *distributed* over the course of 3 consecutive days. The total practice duration was the same in both conditions.

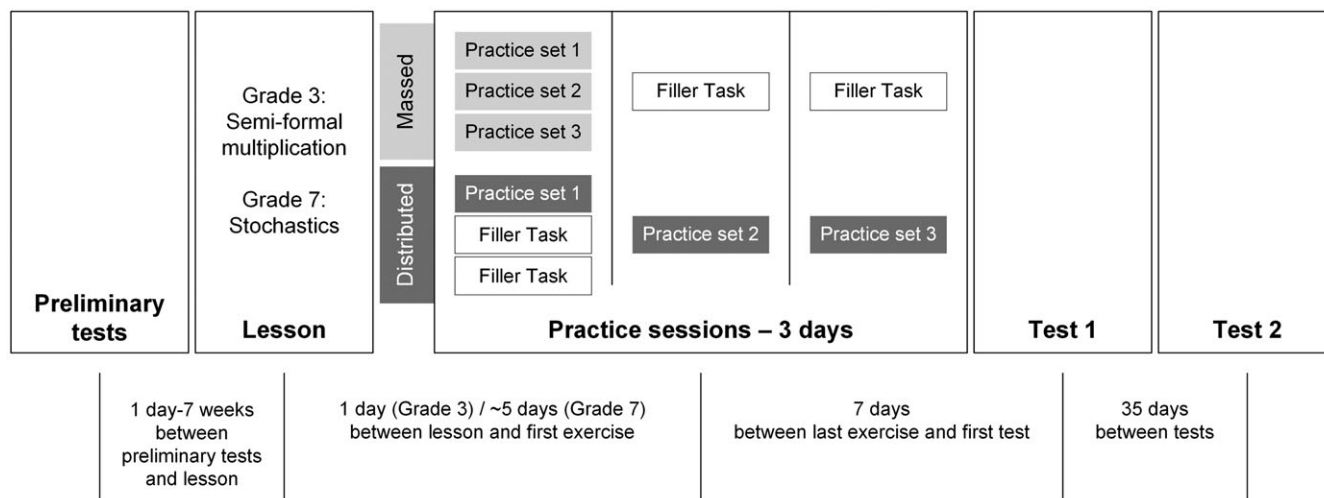
A randomized block design was used to assign the students to one of the two conditions, based on the students' prior mathematical performance. They were ranked by their last grade in mathematics and were then consecutively assigned to the massed or distributed condition within each grade level. In addition, about one half of the students in each class participated in the distributed condition and the other half in the massed condition. This procedure ensured that class and teacher effects were minimized: If a complete class had been assigned to the distributed practice condition and another class to the massed practice condition, differences in performance could be due to the experimental condition or to an overall better performance in one class (class effect) or to the experimental teacher (teacher effect), because not all classes were taught by the same person. Both of the latter effects were ruled out by the randomized block design.

Dependent measures were the intermediate and long-term math performance results, tested 1 and 6 weeks after the last practice session. In addition to the manipulated variable of practice schedule, several cognitive and motivational variables and individual characteristic variables were assessed to examine whether they potentially moderate the effect of the distributed practice schedule on learning performance. However, no interactions with the practice condition were detected. Information on these variables and the corresponding exploratory analyses can be found in the Supporting Information ([osf.io/vfcgz](https://osf.io/vfcgz)).

### 2.3 | Procedure

In Grade 3, the introductory lesson lasted 90 min and was used to introduce a semiformal multiplication method. In Grade 7, the introductory lessons consisted of two 90-min lessons, separated by 2 days, introducing basic probability calculations. Before the respective topics were introduced, the students worked on a short test assessing prior knowledge. In Grade 3, this test referred to precursors of semiformal multiplication that should have been acquired previously at least to some extent (e.g., semiformal addition and mental multiplication). The overall high performance in the prior knowledge test confirmed that the students of both conditions possessed the required knowledge. In Grade 7, the prior knowledge test assessed whether students already had some understanding of the probability concept. Contrary to the third graders, the topic should generally have been new, and it was expected that the students would not perform well in the prior knowledge test. This was confirmed for the students of both conditions. The lessons were held by student teachers without degree but with teaching experience who were supervised by the authors. As stated above, both conditions were realized within each class to minimize potential class and teacher effects. In each seventh grade class, both lessons were taught by the same teacher.

Following the introductory lesson(s), the students were required to work through three practice sheets. In the *massed condition*, students worked all three practice sheets consecutively on the first day of the practice sessions, whereas students in the *distributed condition*



**FIGURE 1** Schematic of the experimental procedure

solved only the first sheet. On the following 2 days, students in the distributed condition worked through the two remaining practice sheets (one per day). Thus, the overall number of exercises was the same in both conditions, but they were either massed in one session or distributed over 3 days. Breaks between the practice sheets were not scheduled, but students of the massed practice condition who finished their exercises early were required to wait until all students were ready to start the new sheet.

In third grade, all students started practicing the day after their lesson. Each of their sheets consisted of four exercises, and students were given 10 min per sheet. In seventh grade, the lag between the second lesson and the first practice session ranged from 5 to 7 days. The greater time lag compared with the third graders was necessary because the students were scheduled to practice on three consecutive days, and the seventh graders already had two sessions in the first week for the introductory lesson. In Grade 7, each sheet consisted of three exercises, and students were given 15 min per sheet. For a schematic overview of the complete procedure, see Figure 1.

While working on the practice sheets, each student had access to a summary sheet containing examples, including solutions, from the introductory lesson. We ensured that examples and exercises were different to prevent students from simply copying the solutions. The summary sheet was handed to the students, which simulated a homework situation in which all classroom material is usually available. Students were asked at the end of each practice sheet whether they had used the summary to solve the exercises.

To examine effects on learning performance, students were handed tasks similar but not identical to those that were worked on during the practice sessions. Similar to the practice sessions, each test session lasted about 10 min in Grade 3 and 15 min in Grade 7. Test sessions differed from practice sessions in that the summary sheet was not provided. Thus, students were required to complete the test tasks completely on their own. Each test additionally included one of two transfer tasks to investigate whether the effect of distributed practice could be generalized to exercise problems that had not been previously worked on. However, because the performance in both grades was very poor in these transfer tasks in both

learning conditions, they were not further analyzed. Information on the transfer tasks is provided in the Supporting Information (osf.io/vfcgz).

The first test was regularly conducted 7 days after the last practice opportunity. However, in one school, 26 students of the distributed practice condition in seventh grade were already tested after 6 days because of a school holiday. In addition, one single seventh grade student of the distributed practice condition was already tested after 5 days due to an organizational mistake. These deviations from the planned schedule will be accounted for in Section 3. The second test regularly took place 5 weeks after the first test (6 weeks after the last practice opportunity). Two seventh grade classes were tested 1 week earlier because of the summer holidays. Consequently, the lag between the first and the second test was 28 days for 29 students and between 35 and 37 for the other 88 students of Grade 7. In Grade 3, one class with 10 participating students was tested 10 to 12 days later than scheduled, due to organizational reasons. However, given that these procedural deviations regarding the second test in both grade levels occurred for both the massed and the distributed practice students of the respective classes, we expected that the results are only marginally affected, if at all. Nonetheless, like the deviations regarding the first test, these deviations will be accounted for when the results are presented.

No substantial questions were answered by the experimenters during practice and test sessions. Only strategic help was provided (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?”). To prevent cheating and mutual help, students from the distributed and the massed conditions in each class were seated alternately in all sessions, beginning with the introductory lesson. Students and teachers were not given any individual feedback on practice or test performances. Teachers were told not to continue working on the topics relevant for the study (third grade: semiformal multiplication; seventh grade: stochastics) before the end of the last test. During every test session, the teachers were asked whether they had worked on the topics. Unfortunately, one teacher of a third grade class stated that she reviewed semiformal multiplication after the

first but before the second test was conducted. This will be considered in Section 3, too.<sup>1</sup>

## 2.4 | Material

The lessons and all practice sheets were developed specifically for this study, based on typical school lessons and exercises. Several didactics experts, including mathematics education specialists for elementary and secondary schools, supervised this development and evaluated or revised the lessons and exercises. The topics for both grades were chosen from the regular curriculum and only classes in which teachers had not already taught the topics during the current school year could participate.

In the *introductory lessons*, the third graders were taught a semi-formal multiplication method and the seventh graders were taught basic probability calculations. Semiformal multiplication is a procedure to solve multiplication tasks including multidigit numbers. Within this study, only tasks with one one-digit number and one two- or three-digit number were taught. In this task, students were required to split the two- or three-digit number into hundreds (in case of three-digit numbers), tens, and single units, and then multiply these by the one-digit number. The results had to be summed up to obtain the result. The teaching content for the seventh graders included the calculation of simple probabilities and drawing tree diagrams. In the first lesson, they were introduced to the basic concepts of probability calculations and how to draw one-stage tree diagrams. The second lesson built on the first lesson by teaching the students how to draw simplified and multiple-stage tree diagrams. The lessons of both grades provided explanations of the relevant mathematical procedures and example tasks, which were solved together in the class. In the seventh-grade classes, the lessons additionally included short probability experiments, such as coin tossing and a few exercises that were worked on by the students individually. The third graders were not required to work on exercises alone during the lessons. Thus, we minimized individual practice during the lessons in both grades so that the practice effect could be attributed to the practice sessions, not the lessons.

During the *practice sessions*, the students worked on tasks based on the introductory lesson. All practice sheets for third and seventh graders were conceptually similar but contained tasks with different numbers. Therefore, using solutions learned by heart was not possible. The maximum score on a practice sheet in third grade was 16. The maximum score on a practice sheet in seventh grade was 8. Examples of the semiformal multiplication tasks that were worked on by the third graders and the practice sheet containing all types of tasks worked on by the seventh graders can be found in the Appendix A. The complete material including the lesson scripts can be inspected in the Supporting Information ([osf.io/vfcgz](https://osf.io/vfcgz)).

The tests assessing intermediate and long-term performance were similar to the practice sheets. The maximum score on a test sheet in

both grades was 12. The numerical solutions of all tasks of the exercise and test sheets were scored by the same rater using a predefined scoring scheme. The scoring was unequivocal because each problem had only one correct (often numerical) solution. Consequently, a second rater was not needed. Depending on the task, receiving more than one point per task was possible, and partial points were granted (e.g., when a third grader correctly split the multiplicands but made a mistake when multiplying them).

## 3 | RESULTS

The data and R scripts for data preparation and analyses are provided in the Supporting Information ([osf.io/vfcgz](https://osf.io/vfcgz)).<sup>2</sup>

### 3.1 | Effect of distributed practice in Grade 3

The descriptive results of the third graders' performance in all practice and test sessions are shown in Table 1.

The descriptive results revealed that in the practice sheets, third graders in both conditions already scored near the maximum, and that this high performance carried over to the intermediate and long-term tests, suggesting a ceiling effect. In addition, no decrease in performance occurred between the first and second test for the distributed practicing third graders and even a small increase for the massed practicing third graders.

To test the two main hypotheses, two linear regression models were calculated, one addressing the performance 1 week after the last practice set and one addressing the performance change from first to second test. The models were performed as Bayesian regression models because the sample sizes were rather small, and Bayesian statistics allows to interpret the results as a range of values together with their respective probabilities instead of the more binary result of frequentist statistics, where an effect is labelled as either significant or not significant. With Bayesian statistics, the parameters of the models are represented as distributions rather than point estimates, and the effects can be assigned probabilities instead of significance levels (Bürkner, 2017; Kruschke, 2015). The Bayesian models were realized with the R package *brms* (Bürkner, 2017). Because there is only very little research on the effect of distributed practice on mathematical performance in a school context, we kept the default priors for the effect parameters, which is an improper flat prior over the reals (Bürkner, 2017). This way, the results are mainly defined by the data and hardly influenced by assumptions about the prior distributions. All of the following models include the practice condition (distributed vs. massed practice) as well as the performance score in the very first practice sheet as control variable for the performance before any experimental manipulation was introduced and were checked for autocorrelation and proper chain convergence.

<sup>1</sup>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.

<sup>2</sup>The data were analyzed with R (R Core Team, 2016), and the following R-packages were used for data preparation and analyses (in alphabetical order): *partykit* (Hothorn, Hornik, & Zeileis, 2015; Hothorn & Zeileis, 2015), *psych* (Revelle, 2016), *RStan* (Stan Development Team, 2018), *tidyverse* (Wickham, 2017).



**TABLE 1** Performance of third graders (mean percentage of correctly solved tasks)

Practice condition	Practice sheets			Tests	
	1	2	3	1 week	6 weeks
Massed	78% (72–85%)	86% (79–92%)	85% (79–92%)	82% (75–88%)	86% (79–92%)
Distributed	81% (75–87%)	90% (86–94%)	88% (84–92%)	91% (87–95%)	92% (85–95%)
N	95	95	95	95	95

Note. 95% Confidence intervals in parentheses.

The first model was calculated with the performance of the third graders in the first test as dependent variable, in order to analyze the performance differences between massed and distributed practicing students 1 week after the last practice session. The mean of the posterior distribution of the effect of distributed practice in this first model was 0.99 (95% credible interval = 0.17 to 1.82). That is, third graders of the distributed practice are expected to attain about one point more in the performance tests (max.: 12 points) than the third graders of the massed practice condition. An evidence ratio test for the posterior distribution of the effect of distributed practice revealed that it is about 100 times more likely that distributed practice has a positive effect on the performance 1 week after the last practice as compared with massed practice, than that distributed practice has no effect or a negative effect ( $BF_{10} = 103$ ). With reference to Lee and Wagenmakers (2013), this can be interpreted as extreme evidence in favor of a positive effect of distributed practice. The mean of the posterior distribution of the control variable (i.e., performance in the first practice set), was 0.35 (95% credible interval = 0.23 to 0.47), suggesting that third graders with a better performance in the first practice can also be expected to achieve higher scores in the first test, compared with third graders with a poorer performance in the first practice set ( $BF_{10} = \text{inf.}$ ).

The second model was calculated to analyze the third graders' performance change from the first to the second test, conducted 6 weeks after the last practice. It was hypothesized that the positive effect of distributed practice would be even more pronounced 6 weeks after the last practice set. The change was calculated as the difference between the scores in the second test and the first test. According to our hypothesis, this change score should be higher (or less negative) in the distributed practice condition than in the massed practice condition. However, the mean of the posterior distribution for the effect of distributed practice on the change score was  $-0.62$  (95% credible interval =  $-1.53$  to  $0.30$ ), which means that, contrary to our hypothesis, the third graders of the distributed practice condition are estimated to have a higher performance loss than third graders of the massed practice condition (or rather, considering the performances presented in Table 1, students of the massed practice condition had a higher performance gain). Correspondingly, according to the evidence ratio test, it is less likely that the effect of distributed practice has a positive effect on the performance change between first and second test than that it has no or even a negative effect ( $BF_{10} = 0.1$ ), which can be interpreted as moderate evidence against a positive effect of distributed practice (Lee & Wagenmakers, 2013). The performance in the first practice set did not seem to influence the performance change from first to second test: The mean of the posterior distribution of the effect is 0.06 with a 95% credible interval from  $-0.07$  to  $0.20$  ( $BF_{10} = 5$ ).

Because there was no empirical support for our second hypothesis, a third model was calculated to analyze the difference between distributed and massed practicing students 6 weeks after the last practice set and to check whether the performance advantage of distributed practice held for the second test. The posterior distribution of the effect of distributed practice on the performance 6 weeks after the last practice as compared with massed practice had a mean of 0.37 (95% credible interval =  $-0.50$  to  $1.20$ ). This range suggests that the effect of distributed practice after 6 weeks is much less conclusive compared with the effect 1 week after the last practice set. In fact, the evidence ratio test revealed that it is only four times as likely that distributed practice—compared with massed practice—has a positive effect on the performance after 6 weeks as that it has no effect or a negative effect ( $BF_{10} = 4$ ). That is, 6 weeks after the last practice set, there is only little evidence for a positive effect of distributed practice on the mathematical performance of third graders (Lee & Wagenmakers, 2013). The effect of the performance in the first practice set on the performance in the second test was similar to the effect on the performance in the first test with a mean of 0.42 (95% credible interval = 0.30 to 0.54;  $BF_{10} = \text{inf.}$ ).

As was stated in the procedure, there were two important points that have to be considered regarding the third graders: Ten students were tested considerably later for their long-term performance (45 to 47 days after the last practice set instead of 35 days) and one teacher (14 students) had stated that semiformal multiplication had been reviewed after the first, but before the second test was conducted. Therefore, control analyses were performed to make sure that the missing effect of distributed practice on the long-term performance could not be attributed to these special cases. It could be confirmed that removing these students did not reveal evidence for an effect of practice condition on the long-term performance. That is, neither the longer retention interval for one group nor the review of the experimental topic for another group was the reason for the missing effect in the long-term test performance in Grade 3.

### 3.2 | Effect of distributed practice in Grade 7

The descriptive results of the seventh graders' performance in all practice and test sessions are shown in Table 2.

Like for the performance of the third graders, Bayesian linear regression models were used to analyze the effect of distributed practice on the intermediate and long-term performance of the seventh-grade students. Again, in each of the following models, performance in the first practice set was included as control variable. In the first model, with performance in the first test as dependent variable, the

**TABLE 2** Performance of the seventh graders (mean percentage of correctly solved tasks)

Practice condition	Practice sheets			Tests	
	1	2	3	1 week	6 weeks
Massed	63% (56–70%)	75% (67–81%)	59% (52–66%)	44% (36–52%)	45% (38–53%)
Distributed	61% (54–68%)	80% (75–85%)	55% (49–62%)	51% (44–58%)	51% (42–59%)
N	118	118	118	118	118

Note. 95% Confidence intervals in parentheses.

posterior distribution of the effect of distributed practice had a mean of 0.93 (95% credible interval =  $-0.18$  to  $2.06$ ). This result is rather similar to that of the third graders, who also had a positive effect of distributed practice of around 1 point within a 0- to 12-point score range. The evidence ratio for the effect of distributed practice on the performance 1 week after the last practice lead to the conclusion that it is 18 times more likely that distributed practice has a positive effect compared to massed practice, than that it has no effect or a negative effect ( $BF_{10} = 18$ ). With reference to Lee and Wagenmakers (2013), this can be considered strong evidence. The effect of the performance in the first practice set on the first test performance had a mean of 0.82 (95% credible interval =  $0.55$  to  $1.10$ ), which means that for seventh graders with a better performance in the first practice set, a higher performance in the first test is expected ( $BF_{10} = \text{inf.}$ ). Again, control analyses were performed because 23 of the 118 students were tested 1 or 2 days early (5 or 6 days after the last practice instead of 7 days). Restricting the sample to those who were tested as scheduled resulted in a similar mean estimate for the effect of distributed practice (0.86 instead of 0.93) but a lower Bayes factor ( $BF_{10} = 9$  instead of 18). Though this is only considered moderate evidence in favor of a positive effect of distributed practice, we suspect that this is mainly due to the smaller sample and as the evidence still supports the positive effect of distributed practice of almost 1 point we maintain our interpretation of the main analysis.

Again, the second model was computed with the performance change from first to second test as dependent variable. The mean of the posterior distribution of distributed practice was  $-0.15$  (95% credible interval =  $-1.10$  to  $0.79$ ), and the evidence ratio indicated that it is a little less likely that distributed practice has a positive effect as that it has no effect or a negative effect on the performance change from first to second test ( $BF_{10} = 0.6$ ). Because an evidence ratio of 1 would mean no effect, this pretty much leads to the conclusion that practice condition contradicting our hypothesis did not influence the performance development between first and second test. The mean of the control variable, performance in the first practice set, was 0.17 (95% credible interval =  $-0.06$  to  $0.39$ ;  $BF_{10} = 13$ ).

Because the second hypothesis—that students in the distributed practice condition would show a smaller performance decrease than students in the massed practice condition—was not supported in Grade 7, an additional model was conducted to analyze the performance in the second test directly. With performance 6 weeks after the last practice set as dependent variable, the mean of the posterior distribution of distributed practice was 0.79 (95% credible interval =  $-0.35$  to  $1.94$ ), that is, the posterior distribution for the effect of distributed practice in Grade 7 was roughly similar to the one for the effect on performance 1 week after the last practice set. Based

on the evidence ratio test, for the seventh graders even at the second test, it is 10 times more likely that distributed practice has a positive effect on the performance 6 weeks after the last practice, compared with no effect or a negative effect, which is considered strong evidence ( $BF_{10} = 10$ ; Lee & Wagenmakers, 2013). Finally, the posterior distribution of the first practice set performance had a mean of 0.99 in this model (95% credible interval =  $0.71$  to  $1.27$ ), confirming again that students with a higher performance in the first practice set also scored higher in the test ( $BF_{10} = \text{inf.}$ ). Finally, for the second test again a control analysis was performed because some students were tested 1 week earlier than scheduled. Removing these students from the sample, similar to the first test, resulted in a comparable mean estimate (0.83 instead of 0.79) and a slightly lower Bayes factor ( $BF_{10} = 8$  instead of 10). Again, the control analyses confirm the central results of the main analyses which therefore will be further referred to.

## 4 | DISCUSSION

The main purpose of this study was to investigate the effect of distributed practice on mathematics performance within an authentic educational setting. Students of elementary and secondary schools practiced mathematical procedures either massed in 1 day or distributed across 3 days. The total duration of practice was constant in both conditions. Performance was tested after retention intervals of 1 and 6 weeks.

The results of the present study provide a strong indication that distributing mathematical practice across several days improves mathematical performance of students in elementary and secondary school at least up to 1 week after the last practice session. These results are in line with the findings of Schutte et al. (2015), who reported a positive effect of distributed practice on mathematical performance for third graders until 10 days after the intervention. However, all children in their study, in essence, practiced in a distributed manner, because they solved addition problems across 19 days. The intensity of the distribution was varied only within each day, and children solved the tasks either in one, two, or four sessions. Thus, a pure massed condition, as in our study, was not involved.

Interestingly, the results for third and seventh graders were surprisingly similar 1 week after the last practice (i.e., about 1-point performance gain for the distributed practice condition within a 0- to 12-point score range), despite differences in the procedure: First, different topics were covered as they were oriented towards the actual curricula of the two grades. Second, the schedule of the introductory and practice sessions differed: The third graders had only one introductory lesson and started practicing the next day, whereas the seventh graders had two spaced introductory lessons and started

practicing 5 to 7 days after the second lesson. Nevertheless, both age groups benefitted from distributed practice in a similar manner in the test 1 week after the last practice.

Contrary to our hypotheses, the advantage of distributed practice did not become stronger in the long run. Especially in Grade 3, the opposite seemed to be the case, because the advantage of distributed practice was no longer present 6 weeks after the last practice session. In Grade 7, however, the effect appeared to be more stable. An explanation for this difference between the two grades could be the kind of tasks the students worked on. Although most of the third graders had not practiced semiformal multiplication in class outside our visits, they may have practiced mental multiplication and formal addition methods within other topics in mathematics (e.g., when learning about money, students add and subtract sums of money and thereby practice addition). Because students of both conditions potentially practiced these skills, differences between the conditions may have been diminished in Grade 3. Stochastics, practiced in Grade 7, on the other hand, is a rather unique and isolated mathematical topic. That is, the performance differences of the seventh graders were likely not influenced by other topics covered in class between the two tests.

One important feature of the present study should be noted: The students did not receive any feedback or correct solutions during the practice sessions. The only external source available to them was the example sheet with exercises and solutions similar to the ones that had to be practiced. This practice environment is comparable to homework exercises for which all material is available, but solutions are provided only in the next class. In future studies, it could be investigated if the positive effect of distributed practice is also revealed or even boosted when feedback is included in the practice sessions. Additionally, the effects of distributed practice with and without feedback could be directly compared.

## 4.1 | Conclusions

Our data suggest that even in complex applied settings, distributed practice improves mathematical performance more than massed practice. In the present study, this effect was more stable for the seventh graders compared to the third graders: In Grade 3, where other topics were covered that might have diminished the effect of distributed practice, the effect was present 1 week, but not 6 weeks, after the last practice set. In Grade 7, the effect emerged both after 1 and after 6 weeks.

To sum up, distributed practice remains a promising strategy in the context of classroom learning and research should continue to explore the conditions under which distributed practice improves the performance of learners in natural educational settings.

## ACKNOWLEDGEMENTS

We would like to thank Wiebke Eden, Ulrike Haßfurth, Jessica Kühnel, and Kristin Wenzel for their support in conducting the experiments and Dorothea Hentzschel, Kathrin Hillebrand, Damaris Keller, Klara Schelbert, Natascha Schupp, and Maïke Siegle for also teaching the participating classes. Our thanks for proofreading and valuable comments on this manuscript goes to Maj-Britt Isberner.

## 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).

## ORCID

Katharina Barzagar Nazari  <http://orcid.org/0000-0003-4909-272X>

## REFERENCES

- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York: Worth Publishers.
- Bjork, R. A. (1975). 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* (pp. 123–144). Oxford, England: Lawrence Erlbaum.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: The MIT Press.
- Bjork, R. A. (2013). Desirable difficulties perspective on learning. In H. Pashler (Ed.), *Encyclopedia of the mind*. Thousand Oaks: Sage Reference.
- Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval, and time in working memory: Evidence from once-presented words. *Memory*, 6(1), 37–66. <https://doi.org/10.1080/741941599>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- 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. <https://doi.org/10.1007/s10648-012-9205-z>
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438–448. <https://doi.org/10.3758/MC.36.2.438>
- 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. <https://doi.org/10.1037/0033-2909.132.3.354>
- Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 389–396. <https://doi.org/10.1037/0278-7393.19.2.389>
- Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2017). Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educational Psychology Review*, 30, 1–19. <https://doi.org/10.1007/s10648-017-9426-2>
- 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. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum.
- Delaney, P. F., Verhoeven, P. P. J. L., & Spiguel, A. (2010). 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* (ed., Vol. 53) (pp. 63–147). London: Academic Press.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology.



- Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7(2), 95–112. <https://doi.org/10.3758/BF03197590>
- Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., Bouwmeester, S., & Zwaan, R. A. (2016). Distributed practice and retrieval practice in primary school vocabulary learning: A multi-classroom study. *Applied Cognitive Psychology*, 30(5), 700–712. <https://doi.org/10.1002/acp.3245>
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium*. Oxford, England: Lawrence Erlbaum.
- Hothorn, T., Hornik, K., & Zeileis, A. (2015). ctree: Conditional inference trees. The Comprehensive R Archive Network. Retrieved from <https://rdrr.io/rforge/partykit/f/inst/doc/ctree.pdf>
- Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partitioning in R. *Journal of Machine Learning Research*, 16, 3905–3909.
- Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, 3, 12–19. <https://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. <https://doi.org/10.1016/j.learninstruc.2014.11.001>
- Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). London: Academic Press.
- Küpper-Tetzel, C. E. (2014). Understanding the distributed practice effect. Strong effects on weak theoretical grounds. *Zeitschrift für Psychologie*, 222(2), 71–81. <https://doi.org/10.1027/2151-2604/a000168>
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Revelle, W. (2016). psych: Procedures for personality and psychological research (R package version 1.6.12). Northwestern University, Evanston, Illinois, USA. Retrieved from <https://CRAN.R-project.org/package=psych>
- Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. C. Kadosh, & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 1118–1134). Oxford: Oxford University Press.
- Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, 107(3), 900–908. <https://doi.org/10.1037/edu0000001>
- 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. <https://doi.org/10.1002/acp.1266>
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6), 481–498. <https://doi.org/10.1007/s11251-007-9015-8>
- Schutte, G. M., Duhon, G. J., Solomon, B. G., Poncy, B. C., Moore, K., & Story, B. (2015). A comparative analysis of massed vs. distributed practice on basic math fact fluency growth rates. *Journal of School Psychology*, 53(2), 149–159. <https://doi.org/10.1016/j.jsp.2014.12.003>
- Son, L. K., & Simon, D. A. (2012). Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review*, 24(3), 379–399. <https://doi.org/10.1007/s10648-012-9206-y>
- Stan Development Team (2018). RStan: the R interface to Stan (R package version 2.17.3). Retrieved from <http://mc-stan.org/>
- Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, 15(5), 529–536. [https://doi.org/10.1016/0022-5371\(76\)90047-5](https://doi.org/10.1016/0022-5371(76)90047-5)
- Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (1st ed., Vol. 60) (pp. 113–189). Waltham: Academic Press.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352–373. <https://doi.org/10.1037/h0020071>
- Wickham, H. (2017). tidyverse: Easily install and load “tidyverse” packages (R package version 1.2.1). Retrieved from <https://CRAN.R-project.org/package=tidyverse>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Barzagar Nazari K, Ebersbach M. Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom. *Appl Cognit Psychol*. 2019;33:288–298. <https://doi.org/10.1002/acp.3485>

**APPENDIX A**  
**EXERCISE EXAMPLES**

The complete material can be inspected in the Supporting Information (osf.io/vfcgz).

Example practice sheet for the third graders. One practice sheet consisted of one exercise with four tasks.

**A.1 | Exercise 1**

(a)  $13 \cdot 6 =$

1	3	·	6	=
		·		=
		·		=

(b)  $3 \cdot 159 =$

3	·	159	=
	·		=
	·		=
	·		=

(c)  $5 \cdot 58 =$

5	·	58	=
	·		=
	·		=

(d)  $243 \cdot 4 =$

2	4	3	·	4	=
			·		=
			·		=
			·		=

Example practice sheet for the seventh graders. One practice sheet consisted of three exercises [Colour figure can be viewed at wileyonlinelibrary.com]

**A.2 | 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: \_ \_ \_ \_ \_

**A.3 | Exercise 2**

What are the chances of drawing one of the two red queens out of a deck with 52 cards?

Answer: \_\_\_\_\_

**A.4 | Exercise 3**

There are 80 apples in one basket. Twelve of them are unripe, eight others contain a worm. Draw a tree diagram to determine the probability of grabbing an unripe apple, of grabbing an apple with a worm inside and of grabbing a good apple.