



No Robust Effect of Distributed Practice on the Short- and Long-Term Retention of Mathematical Procedures

Mirjam Ebersbach^{††} and Katharina Barzagar Nazari[†]

Division of Developmental Psychology, Department of Psychology, University of Kassel, Kassel, Germany

OPEN ACCESS

Edited by:

Lionel Brunel,
Paul Valéry University, Montpellier III,
France

Reviewed by:

Emilie Gerbier,
Université Côte D'Azur, France
Peter Verkoefen,
Erasmus University Rotterdam,
Netherlands

Edgar Erdfelder,
Universität Mannheim, Germany

*Correspondence:

Mirjam Ebersbach
mirjam.ebersbach@uni-kassel.de

[†]These authors share first authorship

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 25 September 2019

Accepted: 01 April 2020

Published: 29 April 2020

Citation:

Ebersbach M and
Barzagar Nazari K (2020) No Robust
Effect of Distributed Practice on
the Short- and Long-Term Retention
of Mathematical Procedures.
Front. Psychol. 11:811.
doi: 10.3389/fpsyg.2020.00811

We investigated the effect of distributed practice and more specifically the “lag effect” concerning the retention of mathematical procedures. The lag effect implies that longer retention intervals benefit from longer inter-study intervals (ISIs). University students ($N = 235$) first learned how to solve permutation tasks and then practiced this procedure with an ISI of zero (i.e., massed), one, or 11 days. The final test took place after one or five weeks. All conditions were manipulated between-subjects. Contrary to our expectations, the analyses revealed no effect of distributed practice and therewith also no lag effect, even though the sample size was sufficiently large. The only significant effect was that test performance was poorer after 5 weeks than after 1 week. In view of the present results and those of other studies, we assume that distributed practice works differently for declarative and procedural knowledge, with less robust of even absent effects when procedural skills are practiced with ISIs compared to massed practice.

Keywords: desirable difficulties, distributed practice, spacing, lag effect, mathematics learning, procedural knowledge

INTRODUCTION

One central aim of research in cognitive and educational psychology is to identify mechanisms that improve short- and long-term retention. So called *desirable difficulties* refer to mechanisms that make the learning process subjectively harder but contribute to a longer maintenance of knowledge (Bjork, 1994). Distributing the total learning time across several sessions instead of learning the same contents in only one session in a massed fashion—also called *distributed practice* or *spacing*—is one of these desirable difficulties. In general, distributed practice yielded robust effects in experimental settings (e.g., Cepeda et al., 2006; Carpenter et al., 2012) and can therefore be considered as a powerful learning tool (Gerbier and Toppino, 2015; Kang, 2016).

Several theoretical accounts try to explain the effect of distributed practice (for overviews see Küpper-Tetzel, 2014; Toppino and Gerbier, 2014). It is, for instance, assumed that distributed practice can impede *deficient processing* because it prevents the metacognitive impression in learners that the to be learned items are already familiar and therewith consolidated (i.e., feeling of knowing). Accordingly, a higher attentive maintenance is devoted to the processing of spaced items compared to massed items (Craik and Lockhart, 1972; Hintzman, 1974; Crowder, 1976; Challis, 1993). Another account assumes that *working memory depletion* occurs in massed practice but not in distributed practice due to the intervals between the practice sessions (Chen et al., 2018).

Furthermore, distributed practice fosters *study-phase retrieval processes* because learners have to remember what they have practiced in the session(s) before due to the intervals between practice sessions (Toppino et al., 2018). Such retrieval practice fosters the consolidation of knowledge by increasing the storage strength of the learned material (Bjork, 1975; Thios and D'Agostino, 1976; Bjork and Bjork, 1992, 2006; Braun and Rubin, 1998; Smith and Scarf, 2017). Distributed practice can also enhance the *variability of the learning contexts* (i.e., mental, physical, and experimental contexts). This variability facilitates recall because it provides more potential retrieval cues (Estes, 1955; Tulving and Thomson, 1973; Glenberg, 1979; Raaijmakers, 2003; Maddox, 2016). All of these mechanisms are not excluding each other and may work simultaneously.

Besides the general effect of distributed practice, some studies were concerned with the optimal length of the intervals between the single learning sessions (i.e., inter-study interval, ISI) to yielding maximal benefits of distributed practice for retention. Cepeda et al. (2008), using verbal material (i.e., true trivia facts), showed systematically that the optimal ISI depends on the timing of the final test (i.e., "lag effect"; Glenberg, 1976). Generally, longer retention intervals (RI) between the last learning session and a final test require longer ISI, although this relationship is considered to be non-monotonic (cf. Donovan and Radosevich, 1999; Janiszewski et al., 2003; Cepeda et al., 2009). Küpper-Tetzel and Erdfelder (2012) manipulated the ISI (none – which is equivalent to massed practice; 1 day; and 11 days) and RI (one; five weeks) systematically among adults who had to learn word pairs. Recall after 1 week was best when the ISI was 1 day (see also Cepeda et al., 2008), while recall after 5 weeks was best with an ISI of 11 days (see also Küpper-Tetzel et al., 2014a). Moreover, a contracting ISI (i.e., inter-study intervals becoming shorter between multiple learning sessions) yielded larger effects on the successful recall of word pairs within a period up to 1 week, compared to a constant or expanding ISI. In contrast, an expanding ISI yielded larger effects for a RI of 5 weeks (Küpper-Tetzel et al., 2014b). Moreover, an expanding ISI also yielded larger effects on the retention tested after 2 weeks, compared to a constant or contracting ISI, when participants had poor prior knowledge (i.e., a low-level initial training). No such effect was found for participants with prior knowledge (i.e., following a high-level initial training; Toppino et al., 2018). Thus, the length and timing of the ISI in distributed practice seem to be central factors for successful retention.

Most studies so far referring to distributed practice addressed declarative knowledge (e.g., facts, foreign vocabulary, names of objects or pictures, and understanding statistics concepts), with verbal recall as dependent variable (cf. Dempster, 1988; Cepeda et al., 2006; Carpenter et al., 2012; Kang, 2016). This is a clear shortcoming because knowledge acquisition in formal and informal learning contexts often includes procedural knowledge. For instance, in mathematics, it is central to be able to execute procedures fluently and accurately. However, much fewer studies considered the effect of distributed practice with regard to procedural knowledge.

In one of these studies, students watched consecutive, video-taped lessons on descriptive, and inferential statistics either

massed on 1 day or distributed over 4 days (Smith and Rothkopf, 1984). In a test after 5 days, students' factual knowledge (as assessed by "general recall," "cued recall," and "matching," including for instance the naming three measures of central tendency) was better in the distributed compared to the massed condition. However, their procedural knowledge (as assessed by "problems," that is, executing calculations) showed no such effect.

Rohrer and Taylor (2006, 2007) asked college students to learn and practice the solving of permutation tasks (i.e., identifying the number of unique orderings of a sequence of letters using a simple formula) in either a massed (i.e., two sessions successively on the same day) or distributed manner (i.e., two sessions with an ISI of 1 week). Students in the distributed condition performed better than participants in the massed condition on a final test after 1 week (Rohrer and Taylor, 2007) and after 4 weeks (Rohrer and Taylor, 2006)—but not after 1 week in the latter study.

Hopkins et al. (2016) spaced or massed quizzes in the context of a pre-calculus course for engineering students. The target skills addressed in the quizzes and the final exams involved declarative knowledge (e.g., understanding the definition of a logarithmic function) and procedural skills (e.g., solving quadratic equations by factoring and the zero product property). The spacing of quizzes referring to previously learned pre-calculus units across the whole duration of the course led to a better performance in the subsequent exams than massing the quizzes (for a similar study, see Lyle et al., 2019).

In another study with third-graders practiced addition problems over a period of 19 days in total (Schutte et al., 2015), either massed (4 min) or distributed (i.e., 2 min in the morning and 2 min in the afternoon, or four times for 1 min each, distributed over the whole day). Distributed practice outperformed massed practice in an immediate test and in a test after 10 days. However, actually all children practiced in a distributed manner across the period of 19 days, and a pure massed condition was not realized in this study.

Barzagar Nazari and Ebersbach (2019) directly compared distributed with massed practice concerning the learning of mathematical procedures among third and seventh graders in school. Pupils were first introduced to the procedure (i.e., semiformal multiplication or basic probability calculation). Thereafter, they practiced these procedures on 1 day for 45 min or on three consecutive days for 15 min each. The distributed practicing students outperformed the massed practicing students after 1 week and after 6 weeks, except for third graders, where the effect disappeared after 6 weeks. Chen et al. (2018) also revealed an effect of distributing practice across three consecutive days, including worked examples of adding fractions, on their task solving performance, tested at the fourth day.

Thus, even as there is some evidence that distributed practice might also enhance procedural skills in mathematics, the results are not always as straightforward (e.g., Smith and Rothkopf, 1984; Rohrer and Taylor, 2006; Barzagar Nazari and Ebersbach, 2019). One reason for the inconsistent findings might be the lag effect: The ISI might have been too short or too long for the particular RI realized in the respective studies. However, the lag effect has, so far, not been tested with regard to science-related procedural skills (e.g., mathematical procedures). Given

that the acquisition and long-term maintenance of mathematical procedures is central for mathematics learning and that studies examining the effect of distributed practice are scarce in this regard, further research is needed.

The present study¹ seizes this issue by examining the effect of distributed practice on the retention of mathematical procedures in a controlled experimental study with university students. We addressed the lag effect by manipulating the ISI as well as the RI systematically between-subjects, assessing short- and long-term effects. In line with the lag effect, we expected an interaction between ISI and RI. For a RI of 1 week, an ISI of 1 day should yield the largest effect, followed by an ISI of 11 days, and both ISIs should be better than no ISI (i.e., massed condition). For a RI of 5 weeks, in contrast, an ISI of 11 days should yield the largest effect, followed by an ISI of 1 day, and both ISIs should yield better results than no ISI (i.e., massed condition; cf. Cepeda et al., 2008).

Furthermore, we examined exploratorily whether individual learner characteristics moderate the effect of distributed practice. This issue has largely been neglected in previous research addressing desirable difficulties in learning. However, in order to provide learners and teachers with recommendations concerning optimal learning strategies, individual differences need to be addressed. We assumed that learners with a poorer *working memory capacity* might be over-challenged by distributed practice because they may stronger be affected by interferences that might occur between the practice sessions and, therewith, face more problems to maintain the primary task goals (Bui et al., 2013; but see Seabrook et al., 2005; Delaney et al., 2018). Furthermore, in line with deficient processing accounts (e.g., Craik and Lockhart, 1972), we expected participants with *difficulties to concentrate* on longer tasks to benefit especially from distributed practice because here, the duration of the single learning sessions is shorter and puts less demands on (longer-term) concentration. In addition, learners' *performance avoidance goals*, *work avoidance*, and *effort motivation* were assessed as potential moderators of the effect of distributed practice. We hypothesized that performance avoidance reduces learners' motivation to engage in the distributed practice condition as this condition requires more retrieval than the massed condition, and unsuccessful retrieval might be discouraging for this group of learners in particular. Furthermore, distributed practice is considered as a desirable difficulty in learning (Bjork, 1994), it is more effortful than cramming, and is therefore rarely used in the context of self-regulated learning (Barzagar Nazari and Ebersbach, 2018). Thus, learners scoring high in work avoidance and effort motivation might be less engaged in the distributed practice condition and, therefore, might benefit less or even not at all from the distribution. These moderator hypotheses were tested exploratorily with regard to the general effect of spacing; no specific moderator hypotheses were postulated concerning the interplay between different ISIs and RIs (see preregistration).

¹This study was pre-registered: <https://osf.io/y79s5/>. It was carried out in accordance with the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2017). The local ethics committee does not claim a general check of studies in psychology.

MATERIALS AND METHODS

Design

The study followed a 3 (ISI: none vs. 1 day vs. 11 days) × 2 (RI: one vs. five weeks) between-subjects design. The dependent variable was the performance in the final test tapping procedural knowledge (i.e., solving permutation tasks; Rohrer and Taylor, 2006, 2007).

Sample

The initial sample consisted of 273 students. Thirty-eight students had to be excluded because they had terminated the study before the end ($n = 18$), had taken part in a similar study before involving the same content ($n = 12$), or for other reasons ($n = 8$). The number of drop-outs in each condition was small and unsystematic, ranging between one to five participants. The final sample consisted of 235 students² (mean age: $M = 24$ years 5 months; 176 women, 59 men) of a wide range of study programs, being randomly assigned to each of the six experimental conditions ($n = 38$ – 40 per condition). Students took part with informed consent and could terminate the experiment at all times. They received 15 Euro when they completed all three sessions or part of the total sum if they dropped out beforehand.

Material

Participants received a computer-based tutorial introducing how to calculate permutations, and worked the practice tasks and test tasks on the computer, too. The tasks were adopted from Rohrer and Taylor (2007; e.g.: "In how many unique ways can the letters *abbccc* be arranged?"; for an overview of all practice and test tasks, see section "Appendix A"). Individual characteristics were assessed via computerized questionnaires and tasks: Working memory capacity by the Corsi Block Task (Kessels et al., 2000) and the Digit Span Backwards test (Woods et al., 2011), learners' ability to concentrate by the Sustained attention test CPT-AX (Rosvold et al., 1956), learners' effort motivation by the LIST (Wild and Schiefele, 1994; Cronbachs alpha: $\alpha = 0.74$), and learners' performance avoidance goals and work avoidance by two scales of the SELLMO (Spinath et al., 2012; split-half reliability: $r = 0.73$ – 0.78), both adapted to mathematics learning.

Procedure

The study was realized as computerized experiment in the laboratory. Students practiced and were tested in small groups up to four persons but the tasks were performed individually on the computer. Each student was randomly assigned to one of the six experimental groups comprising between 38 and 40 students (see section "Design"). Students' prior knowledge on the learning subject (i.e., permutations) was not assessed because no systematic effects were expected due to the random assignment of participants to the experimental conditions. However, students'

²According to GPower (Faul et al., 2009; version 3.1), a sample size of $N = 138$ would be required, assuming a mean effect size of $f^2 = 0.15$ and a power of $1 - \beta = 0.95$, when the data were analyzed by a multiple linear regression, including five predictors (i.e., RI35 compared to RI7, ISI1 compared to ISI0, ISI11 compared to ISI1, and plus the two interaction terms).

performance in the first practice set served as a control variable in the analyses. In the first session, the individual characteristics of all students were assessed³, and students in the distributed condition also received a tutorial explaining the formula to compute permutations and completed the first four of eight practice tasks. In the second session (exactly 1 or 11 days later), students in the distributed condition worked the remaining four practice tasks, whereas students in the massed condition received the tutorial but now and thereafter completed all eight practice tasks in this second session. The practice tasks had to be worked without the tutorial, but students were provided with the full solution path of the first and second task as well as of the fifth and sixth task after they had completed the tasks. This type of incomplete feedback was chosen to save time and because it was assumed that presenting only half of the solution paths would be sufficient to support students' comprehension of permutation. In the third session (exactly one or five weeks later), the unannounced final test was administered that was introduced as further practice session, including five permutation tasks (each with a time limit of 45 s). Each correct solution in the practice and test trials was scored with one point (i.e., max. 4 points in each of the two the practice sets, max. 5 points in the test). Students were informed at the beginning of the study about the schedule of the sessions but not about what would happen exactly in the sessions to prevent them from additional practicing. After the test, students were informed about the aim of this study and that the final session included a performance test.

RESULTS

Table 1 shows the mean performance score in each practice set (max. 4) and the mean score for the first task of practice set 2 (max. 1) as indicator of study-phase retrieval, separately for each condition. An exploratory analysis was conducted to test whether the performance between the first and second practice set changed. A repeated-measures ANOVA with practice set as within-subjects variable and ISI as between-subjects variable revealed no main effects of practice set and ISI, $ps > 0.13$, but an interaction of the two variables, $F(2, 232) = 3.88$, $p = 0.022$, and $\eta_p^2 = 0.03$. *Post hoc* tests showed that the practice performance in groups who practiced with an ISI of 11 days declined between the first and second practice set, $t(77) = 2.49$, $p = 0.045$ (Bonferroni-Holm corrected), whereas practice performance in the other groups did not decline, $p > 0.50$. Furthermore, to check for performance differences between the study and test conditions already in the first practice set, another exploratory ANOVA was computed with ISI and RI as independent variables. This analysis yielded no significant differences between the conditions, $ps > 0.36$.

Descriptive statistics of the test performance are shown in **Figure 1**. As outlined in the introduction, we expected

³For the sake of using time efficiently within the three sessions, the working memory of students in the massed condition was tested in the first session, while it was tested in the second session with students in the distributed condition – in both cases after the practice trials. Furthermore, concentration ability was assessed after the test in the third session in both learning conditions.

TABLE 1 | Mean performance in the practice sets 1 and 2, and in the first task of practice set 1 only, separately for each ISI and RI.

Retention interval	Practice set	Inter-study interval		
		ISI 0	ISI 1	ISI 11
RI 7	1	1.95 (1.69)	1.43 (1.50)	1.85 (1.59)
	2	2.18 (1.30)	1.45 (1.41)	1.43 (1.15)
	1st task of set 2	0.68 (0.47)	0.43 (0.50)	0.40 (0.50)
RI 35	1	1.79 (1.56)	1.82 (1.57)	1.55 (1.61)
	2	1.87 (1.34)	1.89 (1.43)	1.32 (1.23)
	1st task of set 2	0.62 (0.49)	0.61 (0.50)	0.34 (0.48)

Mean scores (max. 4) of the performance in each practice set and in the first task of the second practice set (max. 1); standard deviations in parentheses.

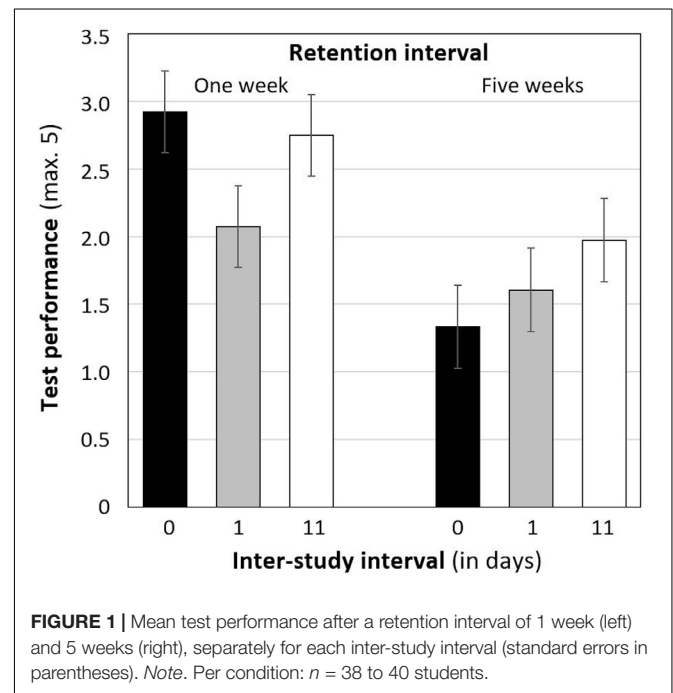


FIGURE 1 | Mean test performance after a retention interval of 1 week (left) and 5 weeks (right), separately for each inter-study interval (standard errors in parentheses). Note. Per condition: $n = 38$ to 40 students.

an interaction between ISI and RI, as indicated by the best performance in a test after 1 week when students practiced with an ISI of 1 day, followed by an ISI of 11 days, and by massed practice. In contrast, performance in a test after 5 weeks was expected to be best when students practiced with an ISI of 11 days, followed by an ISI of 1 day, and by massed practice.

To test our hypotheses, a multiple linear regression was computed with ISI and RI as well as their interaction terms as predictors for the test performance. Because the normal distribution of the residuals could not be assumed (Shapiro-Wilk test: $ps < 0.001$), a robust multiple regression was computed based on bootstrapping. The regression model was significant, $F(5, 229) = 4.21$, $p = 0.001$, and $R_{adj}^2 = 0.06$. As shown in **Table 2**, the performance in a test after 5 weeks was in general poorer ($M = 1.63$, $SD = 1.82$) than in a test after 1 week ($M = 2.53$, $SD = 2.01$). However, contrary to our expectations, there was no main effect of ISI nor an interaction between RI and ISI.

TABLE 2 | Results of a robust multiple regression testing main effects of ISI and RI and interactions.

Predictor	B	SE	t	β	p
ISI_A	-0.02	0.18	-0.11	-0.01	0.909
ISI_B	0.26	0.15	1.71	0.11	0.103
RI	-0.47	0.13	-3.79	-0.24	0.001**
RI \times ISI_A	0.32	0.18	1.84	0.12	0.078
RI \times ISI_B	-0.08	0.15	-0.50	-0.03	0.633

Based on 1000 bootstrap samples. ISI: inter-study interval; RI: retention interval. ISI_A: distributed practice with an ISI of 1 or 11 day compared to massed practice (ISI 0) as reference category; ISI_B: ISI 11 compared to ISI 1 as reference category; RI: RI 35 compared to RI 7 as reference category. RI \times ISI_A and RI \times ISI_B are the interaction terms. ** $p < 0.01$.

Exploratory Analyses

The sample size was not sufficient for computing moderator analyses to uncover potential moderating effects of the learners' characteristics (for descriptive statistics, see **Table 3**) on the effect of distributed practice. These effects were therefore assessed exploratorily by means of conditional inference tree models (CIT, Hothorn et al., 2006, 2015). CIT models are—compared to more classical analyses, like multiple regressions—a flexible tool to uncover linear and non-linear associations between a dependent variable and multiple independent variables as well as interactions between independent variables. The variables can have different scales of measurement. Furthermore, CIT models facilitate the interpretation of complex regression problems by means of the visualization of the fitted decision trees (Zeileis et al., 2008). CIT models are based on recursive binary partitioning and first test globally whether the null hypothesis (i.e., that the dependent variable is independent from all tested independent variables) can be rejected. If this is the case, the independent variable with the strongest relationship to the dependent variable is chosen to divide the sample, based on two (or more) categories of the independent variable, into subgroups that maximally differ from each other with regard to the dependent variable. This process is iterated until the null hypothesis cannot be rejected anymore. A potential moderating effect of the learners' characteristics would be reflected in such a pattern that the whole sample would be divided for instance in learners with lower and higher working memory capacity, and that only in the group of learners with higher working memory, an advantage of

distributed practice would emerge, as mirrored in an additional splitting of this subgroup by means of the practice conditions. Given the exploratory character of these analyses, the results should be treated with caution but may provide interesting hints for future research. However, the analyses revealed none of the expected interaction effects (for a complete information on the results of the CIT analyses, see section “**Appendix B**”).

DISCUSSION

We investigated the effect of distributed practice on the retention of procedural knowledge in mathematics. More specifically, we aimed at examining the lag effect systematically to figure out whether learners tested after a shorter RI would benefit more from a shorter interval between the practice sessions ISI, whereas learners tested after a longer RI would benefit more from a longer ISI. Based on previous research investigating the lag effect with regard to verbal material, we expected a shorter ISI of 1 day to be optimal for a RI of 1 week, and a longer ISI of 11 days to be optimal for a RI of 5 weeks (e.g., Cepeda et al., 2009; Küpper-Tetzl et al., 2014a). In addition, in line with the general effect of distributed practice (e.g., Cepeda et al., 2006), a lag between the practice sessions (i.e., ISI > 0, distributed conditions) was expected to result in a better test performance than no lag (i.e., ISI = 0; massed condition). However, these expectations were not confirmed in our experiment: There was no main effect of ISI, no interaction between ISI and RI, only a main effect of RI: Learners performed poorer in a test after 5 weeks than in a test after 1 week, independently of the ISI.

Our findings are consistent with results of other studies that failed to demonstrate a robust effect of distributed practice with regard to mathematics. Rohrer and Taylor (2006), for instance, found no benefit of distributed compared to massed practice for the retention of mathematical procedures (i.e., permutation tasks like in the present study) when learners were tested after 1 week. Similarly, Smith and Rothkopf (1984) did not find an effect of distributing statistics lessons over 4 days on the solving of mathematical problems in a test 5 days later, and third-graders showed no benefit from the distributed practice of mathematical skills in a test after 6 weeks (Barzagar Nazari and Ebersbach, 2019). Also concerning other rather procedural skills, such as second language syntax, there was no performance difference in

TABLE 3 | Descriptive statistics of the learner characteristics per condition.

Learner characteristics	Condition					
	ISI 0_RI 7	ISI 0_RI 35	ISI 1_RI 7	ISI 1_RI 35	ISI 11_RI 7	ISI 11_RI 35
Working memory (Corsi Block)	6.25 (0.98)	6.08 (1.13)	6.00 (1.28)	6.21 (0.99)	6.18 (1.36)	5.79 (0.78)
Working memory (Digit Span)	6.90 (1.17)	6.77 (1.75)	6.30 (1.22)	6.92 (1.58)	6.90 (1.55)	6.61 (1.46)
Concentration	3.42 (1.10)	3.76 (0.84)	3.35 (1.06)	3.27 (0.93)	3.55 (1.04)	3.40 (0.92)
Effort motivation	4.20 (0.94)	4.12 (0.61)	4.34 (0.81)	4.34 (0.67)	4.31 (0.69)	4.31 (0.67)
Performance avoidance	2.41 (0.89)	2.52 (0.87)	2.46 (0.91)	2.41 (0.87)	2.58 (0.99)	2.42 (0.92)
Work avoidance	2.16 (0.88)	2.19 (0.67)	4.38 (2.38)	2.12 (0.62)	2.00 (0.71)	2.10 (0.73)

Mean scores; standard deviations in parentheses.

a test after 1 week depending on whether practice was distributed with an ISI of 3 days or with an ISI of 14 days (Bird, 2010). Unfortunately, a pure massed condition was not realized in this study. Other studies also failed to reveal a benefit of distributed practice for foreign language acquisition, including vocabulary, grammar, listening and reading (e.g., Lapkin et al., 1998; Collins et al., 1999; Serrano and Munoz, 2007).

However, when the learning addressed the recall of verbal material, not the application of procedures, distributed practice did yield quite reliably positive effects, even in a test after 1 week (e.g., Cepeda et al., 2008, using trivia facts; Küpper-Tetzel and Erdfelder, 2012 and Küpper-Tetzel et al., 2014a, using word pairs or vocabulary; for a review see Cepeda et al., 2006).

One might thus assume that the effect of distributed practice works differently for procedural and declarative knowledge. While declarative (or conceptual) knowledge refers to the understanding of concepts and the ability to recall facts related to these concepts, procedural knowledge additionally involves the construction of schemata and automation processes (Anderson, 1982). The discrepancy between both knowledge types in mathematics becomes evident already when one considers the timing of the development of mathematical skills in children. Depending on the particular skill, procedural or declarative knowledge emerges earlier (see for instance Rittle-Johnson and Schneider, 2015; Rittle-Johnson et al., 2015, for reviews).

The assumption that distributed practice might affect declarative and procedural knowledge differently is confirmed by a meta-analysis of Donovan and Radosevich (1999) revealing that the effect of the distribution depends on the *complexity* of the learning content. More complex content, requiring a number of distinct behaviors or mental procedures, benefits less (or not) from the distribution of practice compared to less complex content. Mathematical procedures can be conceived as being more complex as they include not only recalling the underlying concept (e.g., the formula) but also the ability to apply this concept within a task solving procedure. Thus, working memory is demanded stronger when executing procedures than when simply recalling previously learned word lists (e.g., Ashcraft and Krause, 2007). This complexity might have contributed to the finding of no advantage of distributed practice on procedural knowledge.

This assumption is supported by the finding that a testing effect—that is, a benefit for retention when previously learned information is retrieved already in the learning phase—remains absent or is even reversed when problem solving skills were addressed instead of incoherent learning material (van Gog and Kester, 2012; van Gog et al., 2015; van Gog and Sweller, 2015). van Gog et al. (2015) assume that practicing a problem-solving procedure as long as the underlying schema or concept is not fully consolidated might yield no additional benefits for the learning outcome. Moreover, Ullman (2004) refers to *interferences between procedural memory and declarative memory* in that the retrieval of declarative knowledge can sometimes hinder the retrieval of procedural knowledge (an vice versa).

Finally, given that procedural skills often involve motor elements (i.e., the execution of a procedure), it seems worthwhile

to consider the effect of distributed practice on motor skills. A meta-analysis by Lee and Genovese (1988) revealed overall a medium effect on the retention of motor procedures. However, discrete and continuous procedures have to be differentiated with regard to this effect: Only continuous motor procedures, defined by an arbitrary beginning and end (e.g., rotary pursuit), benefit from distributed practice compared to massed practice. *Discrete motor procedures*, defined by a fixed beginning and end (e.g., throwing a ball), in contrast, benefit more from massed practice (e.g., Lee and Genovese, 1989; Garcia et al., 2008; Panchuk et al., 2013). If one conceives solving a mathematical task as involving a discrete (motor) task, this aspect might have additionally contributed to the absent benefit of distributed practice for this type of tasks.

To sum up, due to the procedural nature of solving arithmetic tasks, the effect of distributed practice on mathematical skills might be less pronounced or even absent (but see Chen et al., 2018, or Lyle et al., 2019, for positive effects). Further research is necessary to examine the effect of distributed practice on procedural skills, also including procedures from different subjects and a more fine-grained grading of ISI and RI. In addition, especially with regard to procedural skills, one should differentiate between the effects of distributed practice on practice performance (skill acquisition) and test performance (retention). Our data showed a significant decrease of the practice performance between practice set 1 and 2 when the two sets were separated by an ISI of 11 days but not for the shorter ISIs. It would also be interesting to examine the effect of increasing, decreasing or constant practice performance in the context of distributed practice with regard to retention.

Concerning the learner characteristics, the explorative analyses in our study provided no hints for moderation effects. However, the results have to be interpreted cautiously given the relatively small sample size and the explorative character of the analyses. Thus, future studies concerned with distributive practice in particular—and with desirable difficulties in learning in general (Bjork, 1994)—should consider such individual aspects before recommending particular learning strategies to all learners in the same way.

To conclude, the effect of distributed practice on the retention of procedural skills requires further clarification before distributed practice can be recommended as an effective learning strategy in mathematics. Moreover, if no such effect exists, theoretical approaches are needed to explaining why the effect is existent for declarative knowledge but not for procedural knowledge. First considerations have been presented here (i.e., complexity; inferences between procedural and declarative memory; discrete motor procedures), but more research is required to check these assumptions.

DATA AVAILABILITY STATEMENT

The dataset for this study and the list of variables can be found in the OSF (<https://osf.io/f6jqp/> and <https://osf.io/b4fz7/>).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Both authors contributed in equal parts to this manuscript. ME provided the general idea of this research, supervised its

realization, and wrote large parts of the manuscript. KB specified the design of this study, collected and analyzed the data, and wrote the results section. Both authors revised the previous versions of this manuscript.

FUNDING

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

REFERENCES

- American Psychological Association (2017). *Ethical Principles of Psychologists and Code of Conduct*. Available online at: <https://www.apa.org/ethics/code/ethics-code-2017.pdf> (accessed August 2, 2019).
- Anderson, J. R. (1982). Acquisition of a cognitive skill. *Psychol. Rev.* 89, 369–406. doi: 10.1037/0033-295X.89.4.369
- Ashcraft, M. H., and Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychon. Bull. Rev.* 14, 243–248. doi: 10.3758/bf03194059
- Barzagar Nazari, K., and Ebersbach, M. (2018). Distributed practice: rarely realized in self-regulated mathematical learning. *Front. Psychol.* 9:2170. doi: 10.3389/fpsyg.2018.02170
- Barzagar Nazari, K., and Ebersbach, M. (2019). Distributing mathematical practice of third and seventh graders: applicability of the spacing effect in the classroom. *Appl. Cogn. Psychol.* 33, 288–298. doi: 10.1002/acp.3485
- Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Appl. Psycholinguist.* 31, 635–650. doi: 10.1017/S0142716410000172
- Bjork, R. A. (1975). “Retrieval as a memory modifier: an interpretation of negative recency and related phenomena,” in *Information Processing and Cognition: The Loyola Symposium*, ed. R. L. Solso (New Jersey, NJ: Lawrence Erlbaum), 123–144.
- Bjork, R. A. (1994). “Memory and meta-memory considerations in the training of human beings,” in *Metacognition: Knowing about Knowing*, eds J. Metcalfe and A. Shimamura (Cambridge: MIT Press), 185–205.
- Bjork, R. A., and Bjork, E. L. (1992). “A new theory of disuse and an old theory of stimulus fluctuation,” in *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes*, Vol. 2, eds A. Healy, S. Kosslyn, and R. Shiffrin (Hillsdale, NJ: Erlbaum), 35–67.
- Bjork, R. A., and Bjork, E. L. (2006). “Optimizing treatment and instruction: Implications of a new theory of disuse,” in *Memory and Society: Psychological Perspectives*, eds L.-G. Nilsson and N. Ohta (New York, NY: Psychology Press), 116–140.
- Braun, K., and 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, 37–65. doi: 10.1080/741941599
- Bui, D. C., Maddox, G. B., and Balota, D. A. (2013). The roles of working memory and intervening task difficulty in determining the benefits of repetition. *Psychon. Bull. Rev.* 20, 341–347. doi: 10.3758/s13423-012-0352-5
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., and Pashler, H. (2012). Using spacing to enhance diverse forms of learning: review of recent research and implications for instruction. *Educ. Psychol. Rev.* 24, 369–378. doi: 10.1007/s10648-012-9205-z
- Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., and Pashler, H. (2009). Optimizing distributed practice: theoretical analysis and practical implications. *Exp. Psychol.* 56, 236–246. doi: 10.1027/1618-3169.56.4.236
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol. Bull.* 132, 354–380. doi: 10.1037/0033-2909.132.3.354
- Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., and Pashler, H. (2008). Spacing effects in learning a temporal ridge line of optimal retention. *Psychol. Sci.* 19, 1095–1102. doi: 10.1111/j.1467-9280.2008.02209.x
- Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *J. Exp. Psychol.* 19, 389–396. doi: 10.1037/0278-7393.19.2.389
- Chen, O., Castro-Alonso, J. C., Paas, F., and Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: evidence from the spacing effect. *Educ. Psychol. Rev.* 30, 483–501. doi: 10.1007/s10648-017-9426-2
- Collins, L., Halter, R. H., Lightbown, P. M., and Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Q.* 33, 655–668.
- Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X
- Crowder, R. G. (1976). *Principles of Learning and Memory*. Hoboken, NJ: John Wiley & Sons.
- Delaney, P. F., Godbole, N. R., Holden, L. R., and Chang, Y. (2018). Working memory capacity and the spacing effect in cued recall. *Memory* 26, 784–797. doi: 10.1080/09658211.2017.1408841
- Dempster, F. N. (1988). The spacing effect: a case study in the failure to apply the results of psychological research. *Am. Psychol.* 43, 627–634. doi: 10.1037/0003-066X.43.8.627
- Donovan, J. J., and Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: now you see it, now you don't. *J. Appl. Psychol.* 84, 795–805. doi: 10.1037/0021-9010.84.5.795
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychol. Rev.* 62, 369–377. doi: 10.1037/h0046888
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Garcia, J. A., Moreno, F. J., Reina, R., Menayo, R., and Fuentes, J. P. (2008). Analysis of effects of distribution of practice in learning and retention of a continuous and a discrete skill presented on a computer. *Percept. Mot. Skills* 107, 261–272. doi: 10.2466/pms.107.1.261-272
- Gerbier, E., and Toppino, T. C. (2015). The effect of distributed practice: neuroscience, cognition, and education. *Trends Neurosci. Educ.* 4, 49–59. doi: 10.1016/j.tine.2015.01.001
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *J. Verbal Learn. Verbal Behav.* 15, 1–16. doi: 10.1016/S0022-5371(76)90002-5
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Mem. Cogn.* 7, 95–112. doi: 10.3758/BF03197590
- Hintzman, D. L. (1974). “Theoretical implications of the spacing effect,” in *Theories in Cognitive Psychology: The Loyola Symposium*, ed. R. L. Solso (Oxford: Lawrence Erlbaum).
- Hopkins, R. F., Lyle, K. B., Hieb, J. L., and Ralston, P. A. S. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educ. Psychol. Rev.* 28, 853–873. doi: 10.1007/s10648-015-9349-8
- Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 15, 651–674. doi: 10.1198/106186006X133933

- Hothorn, T., Hornik, K., and Zeileis, A. (2015). *ctree: Conditional Inference Trees. The Comprehensive R Archive Network*. Available online at: <https://rdrr.io/rforge/partykit/f/inst/doc/ctree.pdf> (accessed July 27, 2019).
- Janiszewski, C., Noel, H., and Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: implications for research on advertising repetition and consumer memory. *J. Consum. Res.* 30, 138–149. doi: 10.1086/374692
- Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: policy implications for instruction. *Policy Insights Behav. Brain Sci.* 3, 12–19. doi: 10.1177/2372732215624708
- Kessels, R. P., van Zandvoort, M. J., Postma, A., Kappelle, L. J., and de Haan, E. H. (2000). The corsi block-tapping task: standardization and normative data. *Appl. Neuropsychol.* 7, 252–258. doi: 10.1207/S15324826AN0704_8
- Küpper-Tetzel, C. E. (2014). Understanding the distributed practice effect. Strong effects on weak theoretical grounds. *Zeitschrift für Psychol.* 222, 71–81. doi: 10.1027/2151-2604/a000168
- Küpper-Tetzel, C. E., and Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: a multinomial processing tree analysis. *Memory* 20, 37–47. doi: 10.1080/09658211.2011.631550
- Küpper-Tetzel, C. E., Erdfelder, E., and Dickhäuser, O. (2014a). The lag effect in secondary school classrooms: enhancing students' memory for vocabulary. *Instruct. Sci.* 42, 373–388. doi: 10.1007/s11251-013-9285-2
- Küpper-Tetzel, C. E., Kapler, I. V., and Wiseheart, M. (2014b). Contracting, equal, and expanding learning schedules: the optimal distribution of learning sessions depends on retention interval. *Mem. Cogn.* 42, 729–741. doi: 10.3758/s13421-014-0394-1
- Lapkin, S., Hart, D., and Harley, B. (1998). "Case study of compact core French models: Attitudes and achievement," in *French Second Language Education in Canada: Empirical Studies*, ed. S. Lapkin (Toronto, ON: University of Toronto Press), 3–31.
- Lee, T. D., and Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. *Res. Q. Exerc. Sport* 59, 277–287. doi: 10.1080/02701367.1988.10609373
- Lee, T. D., and Genovese, E. D. (1989). Distribution of practice in motor skill acquisition: different effects for discrete and continuous tasks. *Res. Q. Exerc. Sport* 60, 59–65. doi: 10.1080/02701367.1989.10607414
- Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., and Ralston, P. A. S. (2019). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educ. Psychol. Rev.* 87:659. doi: 10.1007/s10648-019-09489-x
- Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: a case for encoding variability and study-phase retrieval. *J. Cogn. Psychol.* 28, 684–706. doi: 10.1080/20445911.2016.1181637
- Panchuk, D., Spittle, M., Johnston, N., and Spittle, S. (2013). Effect of practice distribution and experience on the performance and retention of a discrete sport skill. *Percept. Mot. Skills* 116, 750–760. doi: 10.2466/23.30.PMS.116.3.750-760
- Raaijmakers, J. (2003). Spacing and repetition effects in human memory: application of the SAM model. *Cogn. Sci.* 27, 431–452. doi: 10.1016/S0364-0213(03)00007-7
- Rittle-Johnson, B., and Schneider, M. (2015). "Developing conceptual and procedural knowledge of mathematics," in *The Oxford Handbook of Numerical Cognition*, eds R. C. Kadosh and A. Dowker (Oxford: Oxford University Press), 1102–1118.
- Rittle-Johnson, B., Schneider, M., and Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educ. Psychol. Rev.* 27, 587–597. doi: 10.1007/s10648-015-9302-x
- Rohrer, D., and Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Appl. Cogn. Psychol.* 20, 1209–1224. doi: 10.1002/acp.1266
- Rohrer, D., and Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instruct. Sci.* 35, 481–498. doi: 10.1007/s11251-007-9015-8
- Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D. Jr., and Beck, L. H. (1956). A continuous performance test of brain damage. *J. Consult. Psychol.* 20, 343–350. doi: 10.1037/h0043220
- Schutte, G. M., Duhon, G. J., Solomon, B. G., Poncy, B. C., Moore, K., and Story, B. (2015). A comparative analysis of massed vs. distributed practice on basic math fact fluency growth rates. *J. Sch. Psychol.* 53, 149–159. doi: 10.1016/j.jsp.2014.12.003
- Seabrook, R., Brown, G. D. A., and Solity, J. E. (2005). Distributed and massed practice: from laboratory to classroom. *Appl. Cogn. Psychol.* 19, 107–122. doi: 10.1002/acp.1066
- Serrano, R., and Munoz, C. (2007). Same hours, different time distribution: any difference in EFL? *System* 35, 305–321. doi: 10.1016/j.system.2007.02.001
- Smith, C. D., and Scarf, D. (2017). Spacing repetitions over long timescales: a review and a reconsolidation explanation. *Front. Psychol.* 8:962. doi: 10.3389/fpsyg.2017.00962
- Smith, S. M., and Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cogn. Instruct.* 1, 341–358. doi: 10.1207/s1532690xcio103_4
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C., and Dickhäuser, O. (2012). *Skalen zur Erfassung der Lern- und Leistungsmotivation - SELMO 2 (überarbeitete und neu normierte Auflage)*. Göttingen: Hogrefe.
- Thios, S. J., and D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *J. Verbal Learn. Behav.* 15, 529–536. doi: 10.1016/0022-5371(76)90047-5
- Toppino, T. C., and Gerbier, E. (2014). "About practice: repetition, spacing, and abstraction," in *The Psychology of Learning and Motivation*, Vol. 60, ed. B. H. Ross (Waltham, MA: Academic Press), 113–189.
- Toppino, T. C., Phelan, H.-A., and Gerbier, E. (2018). Level of initial training moderates the effects of distributing practice over multiple days with expanding, contracting, and uniform schedules: evidence for study-phase retrieval. *Mem. Cogn.* 46, 969–978. doi: 10.3758/s13421-018-0815-7
- Tulving, E., and Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80, 352–373. doi: 10.1037/h0020071
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270. doi: 10.1016/j.cognition.2003.10.008
- van Gog, T., and Kester, L. (2012). A test of the testing effect: acquiring problem-solving skills from worked examples. *Cogn. Sci.* 36, 1532–1541. doi: 10.1111/cogs.12002
- van Gog, T., Kester, L., Dirx, K., Hoogerheide, V., Boerboom, J., and Verhoeven, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educ. Psychol. Rev.* 27, 265–289. doi: 10.1007/s10648-015-9297-3
- van Gog, T., and Sweller, J. (2015). Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases. *Educ. Psychol. Rev.* 27, 247–264. doi: 10.1007/s10648-015-9310-x
- Wild, K. P., and Schiefele, U. (1994). Lernstrategien im Studium: ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie* 15, 185–200.
- Woods, D. L., Kishiyama, M. M., Lund, E. W., Herron, T. J., Edwards, B., Poliva, O., et al. (2011). Improving digit span assessment of short-term verbal memory. *J. Clin. Exp. Neuropsychol.* 33, 101–111. doi: 10.1080/13803395.2010.493149
- Zeileis, A., Hothorn, T., and Hornik, K. (2008). Model-based recursive partitioning. *J. Computat. Graph. Stat.* 17, 492–514.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ebersbach and Barzagar Nazari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX A

Practice and Test Tasks

In each task, participants were asked to calculate in how many unique ways the presented letters can be arranged. Correct solutions are indicated in parentheses.

Practice Set 1

abbc (12); abbbcc (60); aaaabc (30); and aaccccc (21).

Practice Set 2

accc (4); aaaacccc (70); aabbcc (90); and aacccc (15).

Test

aacc (6); aaacc (10); abbbbbc (56); aaaaabc (42); and aaccccc (28).

APPENDIX B

Results of the Conditional Inference Tree Analyses

Regarding the test performance after *one week*, the performance in the first practice set (max. 5 points) yielded an effect, with poor performers (i.e., ≤ 2 points; $n = 73$) demonstrating a poorer test performance than medium performers (i.e., 3 points; $n = 20$), $p < 0.001$, who performed worse than good performers (i.e., > 3 points, $n = 27$), $p = 0.033$. Working memory, assessed by the digit span, had an effect on the test performance after one week, with learners with a shorter digit span (i.e., ≤ 7 ; $n = 89$) achieving lower test scores than learners with a longer digit span (i.e., > 7 ; $n = 31$), $p = 0.032$. Work avoidance had an effect, too, with learners with a higher work avoidance (i.e., > 2.75 ; $n = 23$) yielding lower test scores than learners with a lower work avoidance, i.e., ≤ 2.75 ; $n = 97$), $p = 0.044$. Working memory, assessed by the Corsi block, sustained attention, effort motivation, and concentration yielded no effects and no interactions between the individual learners' variables and the learning conditions (i.e., ISI) were revealed.

Regarding the test performance after *five weeks*, there was an effect of the performance in the first practice set, too. Poor performers (i.e., ≤ 2 points; $n = 73$) demonstrated a poorer test performance than better performers (i.e., > 2 points; $n = 40$), $p < 0.001$. In addition, there was an effect of working memory, as assessed by the backward digit span: Students with a shorter to normal digit span (i.e., ≤ 8 ; $n = 106$) showed a poorer performance in the final test than students with a large digit span (i.e., > 8 ; $n = 8$), $p = 0.004$. None of the other variables yielded a main effect on the final test performance after 5 weeks and no interactions could be assumed with the learning conditions (i.e., ISI).