# Investigating distributed practice as a strategy for school students learning mathematics

**Dissertation**

zur Erlangung des akademischen Grades

Doktorin der Philosophie (Dr. phil.)

Vorgelegt im Fachbereich Humanwissenschaften

der Universität Kassel

Von Katharina Barzagar Nazari

Kassel, September 2018

Datum der Disputation: 21. Dezember 2018

Erstgutachterin: Prof. Dr. Mirjam Ebersbach

Zweitgutachterin: Prof. Dr. Rita Borromeo Ferri

Drittgutachterin: PD Dr. Judith Schweppe

# Table of contents

# CHAPTER I: Introduction and aims

## 1. Synopsis

Memory and learning rank among the most fascinating phenomena in psychology, filling a countless number of books. However, theories and empirical studies on learning oftentimes focus on rote memory of isolated items (e.g. words, syllables or facts). From a psychological view, less is known about the retention of more complex material, like mathematical concepts and procedures. This also applies to a prominent learning strategy that is known as spacing or *distributed practice*. With distributed practice, a given learning duration is interrupted by at least one break of variable length. In contrast, with massed practice the same total time is spent learning, but without interruption. Although the theoretical explanation for the positive effect of distributed practice is still a matter of debate, there is a rich body of empirical evidence proving the positive effect on the retention of verbal material. Beyond rote memory, however, the empirical grounds regarding the effect of distributed practice are sparse.

In my dissertation I address this research gap by investigating the effect of distributed practice on the mathematical performance of school students. The dissertation is structured as follows: The subsequent introductory sections provide a compilation of theories on learning and empirical results regarding the positive effect of distributed practice. The main purpose is to expose distinctive features of mathematical learning and practice and to review the few available studies on distributed practice with mathematical material, especially considering their contribution to research on distributed practice that goes beyond rote memory. The three ensuing chapters cover the three studies I conducted on distributed mathematical practice in school, including experiments with third and seventh graders as well as high school students. In the final chapter of my dissertation, the results of the three studies are summarized and discussed.

## 2. Memory

In Cowan's *Embedded processes model of working memory* (Cowan, 1999), three levels of memory are distinguished: First, the *long-term memory* holds all information that is stored for more than just a few moments. At any given time, only a part of this information is activated, either consciously or unconsciously. This *activated part* of long-term memory constitutes the second level of memory and can also be considered as short-term memory. If information is highly activated and a person additionally is aware of this information, it is in the *focus of attention*, which constitutes the third level of memory. Cowan (1999) assumes a time limit for the activation of an item, that is, the level of activation of an item fades with time, and a capacity limit for the focus of attention, that is, only a limited amount of items or information can be in the focus of attention at once. Generally, no such limits are assumed for the long-term memory store per se (Bjork & Bjork, 1992; Cowan, 1999).

The allocation of attention can happen either automatically, that is, it can be the result of exceptional occurrences (like a loud noise or a flash of light), or attention can be allocated consciously. This deliberate allocation of attention is controlled by the *central executive*, which can also be described as the control center of the focus of attention (Cowan, 1999). Via the central executive, attention is directed and processing of information is regulated (Cowan, 1999; Norman & Shallice, 1986). The functions of the central executive can be further specified. Often, three executive functions are distinguished (Miyake et al., 2000): *Shifting* describes the ability to alternate between different cognitive demands. *Updating*, sometimes also referred to as *working memory*, means the constant monitoring of the information at hand with the objective to hold relevant information at disposal and disregard information that is not relevant (anymore). *Inhibition* is the ability to suppress automatic but unwanted reactions (Miyake et al., 2000).

The long-term memory store is generally assumed to have an associative structure (Anderson, 2013; Lefrancois, 2015). That is, separate items or information chunks are not stored independently from each other but linked, similar to a network. According to Cowan (1999), different processes influence the structure of this memory network, namely encoding, maintenance and retrieval, which will be described in further detail now: Items and features of memory can be activated by new or known stimuli, which are then encoded and result in new or adapted memory traces. Cowan (1999) states that the *encoding* of a stimuli is defined by the specific composition of simultaneously activated features at the time of the encoding. That is, items are not stored isolated and as blueprints of a stimulus, but rather the storage of an item is influenced by the specific combination of activated information at the time of the encoding, and it is stored intertwined with these activated features in long-term memory. Importantly, a stimulus does not have to be in the focus of attention to be encoded (that is, even stimuli that a person is not consciously aware of can result in changes in or creation of memory traces in long-term memory). However, in a case where a stimulus is encoded while not being attended to, less features or other items are activated compared to a case where the stimulus is encoded while being in the focus of attention. That is, in the latter case encoding happens to be more fragmented (Cowan, 1999). *Maintenance* is considered to describe the process of keeping an item in the focus of attention. This act of staying attended to an item is governed by the central executive and is related to the executive function updating that was described above. Maintenance or updating can be achieved by different strategies (e.g. encoding, rehearsal, mental search…) (Cowan, 1999). *Retrieval* describes the process of entering the correct item (that is, the item that is needed/searched for in the respective situation) into the focus of attention. There are different initial situations from which the item must be retrieved: If the item is already activated, the process of retrieval should be easier. However, this advantage has a time limit because the activation fades rather quickly. No such time limit is given if the item must be retrieved from long-term memory (besides possible practical limits like a time limit in a test). However, retrieval from long-term memory is only possible if the memory trace is sufficiently strong (Cowan, 1999). The processes involved in the encoding, maintenance and retrieval of an item are assumed to be relatively independent of its modality, the same is true for the decay of memory representations over time (i.e., the same principles apply for example for visual and verbal information) (Cowan, 1999).

In their *New Theory of Disuse*, Bjork and Bjork (1992) state several assumptions concerning the relationship between storage and retrieval of items in long-term memory: (1) The (current) retrieval strength of an item describes how accessible the item is at a given moment. Retrieval strength is assumed to be independent from the (more stable) storage strength, that describes the more general strength of the items memory trace (Bjork & Bjork, 1992). Reconsidering the terms introduced above, retrieval strength should be closely related to the level of activation of an item. (2) The storage strength of an item in long-term memory (i.e., the strength of its memory trace) is not limited and once established, the storage strength of an item cannot be lost again, as already stated within the model of Cowan (1999). However, according to Bjork and Bjork (1992) the gain in memory strength of an item is negatively correlated with both its storage and retrieval strength at learning. That is, the higher the current retrieval strength is in a learning situation, the easier it can be retrieved, but the memory trace or storage strength is not enhanced much by this easy retrieval. The same is assumed for the relationship between the *current* storage strength and the *gain* in storage strength: If an item already has a high storage strength, another learning opportunity will increase the storage strength to a smaller degree than when the item still has a low storage strength (Bjork & Bjork, 1992). (3) The retrieval capacity of memory, contrary to storage strength, is limited, that is, only a limited amount of information can be retrieved at a given moment (Bjork & Bjork, 1992). This could at least partly be explained by the assumption stated above that the focus of attention constitutes a limited capacity. (4) Both rehearsal and retrieval of an item strengthen its storage and retrieval strength, with retrieval having an even bigger effect than rehearsal (Bjork & Bjork, 1992). (5) When attention is directed from one item to another, the retrieval strength of the former item declines and it declines even faster, when the retrieval strength was on a high level and the storage strength is low (Bjork & Bjork, 1992).

## 3. Changes to memory: Forgetting and learning

Above it was stated that the long-term memory per se does not have a limited capacity and that the storage strength for an item, once established, cannot be lost again. However, it happens regularly that items or information that were once known cannot be retrieved, a phenomenon that can be labelled with the term *forgetting* (Tulving, 1974). The process of forgetting is described by a negatively accelerated power function, that is, shortly after the last retrieval, the probability of another successful retrieval drops faster than when a lot of time has already passed (Anderson, 2013).

Ultimately, the goal of learning is to create a robust and strong trace for the item to be learned, which should prevent or at least decelerate forgetting and increase the probability that the respective information is available when retrieval is demanded for. The idea that merely repeating an item over and over is sufficient to create a strong and sustainable memory trace has been outdated for quite some time (Anderson, 2013; Craik & Lockhart, 1972). In the following sections, several theories and hypotheses on first the mechanisms involved in forgetting and second successful learning strategies to prevent or at least minimize forgetting are described.

### 3.1 Forgetting

There are several theories attempting to describe why information is forgotten and how the process of forgetting works (Lefrancois, 2015). Following the *fading theory*, unused memory traces of stored items simply decay with time and finally vanish completely, if too much time passes. This theory, however, is criticized because its explanatory power is limited: "Time" itself will not cause the decay of memory traces and nothing is said about the factors that come with time that will actually cause the memory trace to decay (Lefrancois, 2015; McGeoch, 1932). Interference theories are more precise in this regard. According to this branch of theoretical approaches, learning different items or chunks of information in succession impedes the consolidation of each affected memory trace (Anderson, 2013; Lefrancois, 2015). With *proactive interference* it is assumed that former learned information hinders the learning of subsequent items (Lefrancois, 2015; Wixted, 2004). However, Wixted (2004) proposed, that *retroactive interference* is far more relevant regarding realistic learning situations. With retroactive interference, it is assumed that after learning an item, it takes some time – say, up to minutes – to consolidate or strengthen the respective trace. While the trace is still in the process of consolidation, learning other, independent items, may interfere (Lefrancois, 2015; Wixted, 2004). Thus, engaging in cognitive processes while a memory trace is still being consolidated impairs the trace. One explanation for interference effects may be the limited capacity of the focus of attention that was proposed by Cowan (1999). That is, following items displace items from the focus of attention that still were in the process of consolidation, potentially resulting in weaker, more fragmented memory traces. However, it is assumed that there are mechanisms that alleviate these interference effects – otherwise it would not be possible to learn more than one item at a time (Anderson, 2013; Lefrancois, 2015). For example, it is assumed that these interferences do not occur when there is a causal relationship between the two learned items (Anderson, 2013). That is, especially content-wise independent memory traces are prone to interference.

Altmann and Gray (2002) propose to consider decay of activation and interference to be two connected processes that both affect forgetting. The authors assume that the decay of activation actually is a functional process, in that it prevents or at least reduces the threat of interference with newly learned information. The authors assume that the rate at which the decay of activation progresses is adaptive, that is, items do not lose their level of activation at a stable rate but rather in dependency on the rate at which new information is acquired; the more new information is processed, the faster formerly activated parts of memory fade from activation (Altmann & Gray, 2002). However, as stated above this decrease in activation could also be the reason for retroactive interference in that it results in weaker, less elaborate encoding.

### 3.2 Learning and remembering

Successful learning strategies on the one hand should help to create a sustainable memory trace in long-term memory and on the other hand improve the ability to retrieve the relevant information when demanded, by practicing retrieval skills (Bjork, 1994). In their *Levels of Processing* framework, Craik and Lockhart (1972) propose that the strength of a specific memory trace is substantially defined by the depth of processing that occurs during learning. Level or depth of processing here is defined as the “degree of semantic or cognitive analysis” (Craik & Lockhart, 1972, p.675) that occurs within the learning process. Several learning strategies use this mechanism by initiating deeper processing than pure rehearsal of an item. One way to initiate deeper processing is *elaboration*. This means that newly learned information should be enriched with self-generated or self-chosen additional information. This can be achieved by thinking about examples or consciously using information from the long-term memory to connect the newly learned information to. These self-generated associations support a reliable trace in long-term memory (Anderson, 2013; Lefrancois, 2015). A similar strategy is the *organization* of learned information. This means, that newly learned information should be grouped and structured, for example, by creating categories or defining concepts (Lefrancois, 2015). The deliberate use of *memory cues* is another possibility to enhance learning. These cues can be content related, but they can also concern the physical or emotional context information from the learning situation. For example, if the retrieval of learned items happens in the same environment or in the same emotional state as the learning itself, the probability for a successful retrieval is higher than if the context of learning and retrieval differ (Anderson, 2013). All the described learning strategies are encoding strategies, that is, they change the way information is processed during learning. There are several mechanisms suspected to explain why more elaborate encoding strategies promote sustainable learning (see for example Atkinson & Shiffrin, 1968).

### 3.3 Desirable difficulties

A bunch of additional learning strategies were formally introduced by Bjork (1994) as being effective by making the learning process itself more difficult. The idea of the mechanism behind these learning strategies (introducing difficulties in learning to boost long-term performance) was already mentioned in the description of Bjork and Bjork's (1992) *New Theory of Disuse*. Here, it was stated that the gain in memory strength of an item is negatively correlated with its current retrieval strength, that is, that a low retrieval strength during learning leads to higher benefits on the memory trace (Bjork & Bjork, 1992). It is important to note that the positive effect of learning strategies relying on desirable difficulties is generally expected to emerge especially regarding the intermediate- and long-term performance, while performance during the learning may suffer from desirable difficulties (Bjork, 1994). That is, especially when the activation of the learning situation faded and the strength of the memory trace together with the retrieval skills are essential for successful retrieval, the learning strategies relying on desirable difficulties are expected to show their highest effect.

What distinguishes the strategies that will be described shortly from the strategies described above, is that they mostly not only influence the way items are encoded (influencing the strength of the memory trace), but also promote practice of the retrieval process itself. According to Bjork (1994), both these mechanisms are essential for successful learning. The most prominent strategies relying on desirable difficulties are introduced in the following: *Interleaved practice* means that related but different concepts, for example in mathematics, are learned alternately instead of consecutively (Rohrer, 2012; Rohrer & Taylor, 2007). The *generation effect* describes the phenomenon that if the information to be learned is self-generated, the intermediate and long-term retention is usually higher than when the same information is consumed rather “passively” as with simple rehearsal of a given item (Bertsch, Pesta, Wiscott, & McDaniel, 2007). Learning with *tests* means that tests are not only utilized to capture the current performance but also to learn and memorize information (Roediger III & Butler, 2011). Finally, spaced learning or *distributed practice* means that a given learning duration is spread across several learning sessions that are spaced apart in time, instead of learning the same total time in one session (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012).

Hitherto many empirical studies have proven learning strategies that rely on desirable difficulties to be an effective tool for long-term learning performance. The focus of the following sections will be the question under which circumstances the strategy of *distributed practice* proves a useful strategy.

## 4. Distributed practice

Although Bjork (1994) was one of the first to introduce desirable difficulties as the common feature of many successful learning strategies, research on distributed practice specifically dates back more than a century (e.g. Ebbinghaus, 1913). As already described, with spaced or distributed practice a given learning duration is distributed over more than one session, with shorter or longer interruptions in between the sessions, while with massed practice the same total time is spent learning straight, without interruption. Often, massed practice results in superior practice and short-term performance compared to spaced or distributed practice. However, with a longer temporal lag between learning and test time, the positive effect of spaced practice as compared to massed practice emerges (Bjork, 1994; Dempster, 1989). In most studies on spaced and distributed practice, subsequent learning sessions are used to repeat the material to be learned (Toppino & Gerbier, 2014). A slightly different branch of research pursues the question whether input of new material should be distributed (e.g., Budé, Imbos, Wiel, & Berger, 2010; Collins, Halter, Lightbown, & Spada, 1999; Rohrer, 2015). However, this is usually not meant with the term spaced or distributed practice. That is, the effect of distributed practice is usually investigated regarding the question of how *repetitions* of the material to be learned should be scheduled.

According to Cepeda, Pashler, Vul, Wixted and Rohrer (2006) I will refer to the lag between separated learning occasions as *interstudy interval* and the lag between the last learning opportunity and the test as *retention interval*. If there are more than two learning sessions in distributed practice, different options to schedule the learning sessions exist: All interstudy intervals can be of equal length (equal schedule), the interstudy intervals can increase over time (expanding schedule) or they can decrease over time (contracting schedule) (Küpper-Tetzel, 2014). Research on the question of which schedule is superior is still sparse, but empirical findings indicate that the optimal schedule could depend on the lag length itself (e.g., seconds or days between sessions), if there was feedback involved or not, test specifics (e.g., cued or free recall), the retention interval and even the learning material (Küpper-Tetzel, Erdfelder, & Dickhäuser, 2014). If there is any effect of the interstudy interval-schedule, there tends to be an advantage for expanding schedules, especially when longer retention intervals are involved (Cepeda et al., 2006; Gerbier & Koenig, 2012; Gerbier, Toppino, & Koenig, 2014; Küpper-Tetzel et al., 2014). However, there is reason to believe that the schedule does not crucially influence the effect of distributed practice at all (Balota, Duchek, & Logan, 2007; Roediger III & Karpicke, 2011).

The vast majority of studies on spaced and distributed practice involve the learning of verbal material (Cepeda et al., 2006). In a typical experiment, participants are presented word pairs that are learned either in a massed or spaced fashion (e.g., Son, 2010). Performance is tested after a broad range of retention intervals, varying from seconds to even years. However, most studies implement retention intervals of less than a day (Cepeda et al., 2006). Although the majority of studies is on simple verbal material, positive spacing effects have been found with a great variety of learning material and are investigated for example in advertising (Appleton-Knapp, Bjork, & Wickens, 2005), grammar learning (Bird, 2010; Miles, 2014), surgical training or other motor skills (Lee &

Genovese, 1988; Moulton et al., 2006) and learning of science and mathematics (Grote, 1995; Rohrer & Taylor, 2006).

**4.1 The lag effect in distributed practice**

Within spaced or distributed practice, the effect of different interstudy intervals on the performance after a fixed retention interval is labelled with the term lag effect. Research on the lag effect led to the conclusion that given a fixed retention interval, the positive effect of spaced learning first shows a steep increase with an increasing interstudy interval before it reaches a maximum, indicating the optimal lag. Extending the interstudy interval beyond this optimal lag reduces the positive effect of distributed practice, but slower than it increased before the optimal lag (i.e., a little too long lag is better than a lag a little too short). That is, for a given retention interval the effect of the length of the interstudy interval follows an inverted and non-symmetrical u-shape, with both too short and too long lags decreasing the positive spacing effect (Cepeda et al., 2009; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Toppino & Gerbier, 2014). Additionally, it is assumed that the optimal interstudy interval depends on the retention interval, with longer retention intervals requiring longer interstudy intervals (Cepeda et al., 2008; Küpper-Tetzel, 2014). However, this relationship is presumably not linear. More precisely, comparing six different interstudy intervals for four different retention intervals, Cepeda et al. (2008) came to the conclusion that the ratio of interstudy interval and retention interval should decrease the longer the retention interval is. For example, based on their experiment the authors assume an optimal interstudy interval of about 1 day when the retention interval is 7 days (ratio of 0.14) and an optimal interstudy interval of about 21 days when the retention interval is 350 days (ratio of 0.06) (Cepeda et al., 2008).

It should be noted that the terminology concerning the effect of distributed versus massed practice sessions is not consistent. While the lag effect always refers to the comparison of different interstudy intervals within spaced or distributed practice, some authors use the term spacing effect in case of very short interstudy intervals and "massed practice-distributed practice" when longer lags are involved (Toppino & Gerbier, 2014). Other authors use the term spacing as opposed to massing and the term distributed practice as general term for both spacing and lag effects (Cepeda et al., 2006; Küpper-Tetzel, 2014). In the following, the term distributed practice will be used when the total learning time is distributed over more than one session and the term massed practice will be used, when the total learning time is spent learning without interruption.

### 4.2 Theoretical accounts for the distributed practice effect

In contrast to the tremendous amount of empirical support for a positive effect of distributed practice compared to massed practice on long-term performance, the theoretical foundation for the effect is surprisingly vague (Küpper-Tetzel, 2014). Nevertheless, there are mainly three approaches to explain the distributed practice effect, which are described in the following section.

The engagement with an item leads to its (more or less deep) cognitive processing which then strengthens the respective memory trace – these mechanisms already were described in the sections on memory and learning. Based on *deficient processing theories*, however, to engage with an item in a massed fashion leads to a shallow processing during the repetitions because the learner is not able or not willing to process the same item twice in succession (Crowder, 1976; Hintzman, 1974; Toppino & Gerbier, 2014). Thus, if an item is learned repeatedly at a time, the respective memory trace experiences less gains than if the same item is learned as often, but at different times. With an increasing lag between the learning situation and the repetition, as in distributed practice, the repeated processing of the item increases in depth and the memory trace is increasingly strengthened by the additional learning. In massed practice, on the other hand, only the first engagement with the item uses the most effective processing, while the following repetitions suffer from deficient processing and do not maximally strengthen the memory trace (Craik & Lockhart, 1972; Hintzman, 1974; Toppino & Gerbier, 2014).

When a chunk of information or item is studied for the second time, ideally, the memory trace that was initiated the first time the specific item was studied is retrieved (that is, either consciously or unconsciously re-entered into the focus of attention) and strengthened. However, if the item is not recognized, the item is encoded independently from the original trace, resulting in a new (second) memory trace. One important assumption in the frame of *study phase retrieval theories* is that the more time has passed since the last activation of an item, the harder it is to retrieve it from memory, which means that the probability for independent memory traces for identical items increases. However, as stated for example within the context of the *New Theory of Disuse* (Bjork & Bjork, 1992), a harder (successful) retrieval attempt is associated with a bigger gain in memory trace strength. The basic idea of study phase retrieval theories combines these ideas with respect to distributed practice: The more time has passed between learning sessions, the harder the retrieval attempts become, while also boosting their effect on the memory trace strength – as long as the retrieval succeeds (Bjork, 1975; Küpper-Tetzel, 2014; Thios & D'Agostino, 1976). Thios and D'Agostino (1976) propose different explanations for the fact that a harder (successful) retrieval should improve long-term retention more than an easy retrieval. First, it could be that the harder retrieval *provokes deeper processing* because more cognitive resources have to be engaged in the retrieval attempt. Deeper processing, as already mentioned several times, then leads to stronger memory traces (Craik & Lockhart, 1972). A second explanation by Thios and D'Agostino (1976) is that the harder retrieval is better training for the later test situation, where retrieval without help is usually hard as well. That is, in *closely representing the test situation*, the harder retrieval in distributed practice is a better preparation for the actual test than the easier retrieval in massed practice. According to both explanations massed practice should favor good performance in the learning situation but prevent optimal performance on later occasions – either, because processing was not deep enough and/or because the learner is not sufficiently prepared for the (difficult) retrieval at the test (Küpper-Tetzel, 2014; Toppino & Gerbier, 2014).

For the explanation of the effect of distributed practice via *encoding variability theories*, the most important feature of the long-term memory is the abovementioned network-like structure, where items are never stored isolated but encoded with information that was activated during the processing of the learned material (Cowan, 1999; Lefrancois, 2015). Information that is closely related to the searched item, then, can work as cues in later retrieval situations in that the activation of related material also increases the activation level of the target information, which should facilitate retrieval (Crowder, 1976; Glenberg, 1979; Toppino & Gerbier, 2014). The crucial proposition of encoding-variability-approaches to the explanation of the distributed practice effect, is that with an increasing lag between two repetitions of an item, the variability of encodings of the item increases, resulting in a greater number of retrieval cues, which in turn should increase the probability of successful retrieval in a later test (Crowder, 1976; Glenberg, 1979). Put in other words, it is assumed that with massed practice more or less the same features of memory are activated throughout the complete learning process, resulting in similar encodings each time the item is repeated and a more or less constant amount of retrieval cues. With distributed practice, on the other hand, because of the temporal lag since the last encoding, different features of memory are activated each time the item is repeated which in turn results in a greater variability of available retrieval cues and facilitated retrieval.

Unfortunately, all three theoretical approaches to explain the positive effect of distributed practice have seldom been tested directly (Küpper-Tetzel, 2014) and though at first glance they all seem capable to explain the effect, it became evident that none of the theories alone can account for all the specific features of the distributed practice effect. Hence, in recent theoretical publications on the distributed practice effect, different approaches are combined, generally study phase retrieval and encoding variability accounts (Küpper-Tetzel, 2014; Toppino & Gerbier, 2014; Verkoeijen, Rikers, & Schmidt, 2004; Young & Bellezza, 1982). In the following section, the approach proposed by Delaney, Verkoeijen and Spirgel (2010) will be described in greater detail.

For their combination of study phase retrieval and encoding variability accounts, Delaney et al. (2010) seize up on an approach by Malmberg and Shiffrin (2005). Malmberg and Shiffrin (2005) proposed that when an item is studied, a fixed amount of context information is stored for that item within the first one or two seconds of processing and after that, the context information is not enriched any more. That is, only "one-shot" of available *context information* is stored in each learning situation, regardless of the study duration. According to the authors, information on the *item content* (i.e., the items storage strength), however, continues to be strengthened with study duration. Context information, here, is similar to the memory cues described above and can refer to the emotional or physical state of the learner in the learning situation, physical surroundings and similar features. The authors suppose that in most study situations, context information is not in the focus of attention, though it should be activated to be transferred to long-term memory. By contrast with context information, the item content information itself should be in the focus of attention in the study situation (Malmberg & Shiffrin, 2005).

Delaney et al. (2010) extended this idea by adding another layer of information. They proposed that similar to context information, *associative information*, that is, associations between related items, is established right when the studied item enters the focus of attention and afterwards, a longer study duration does not result in more or stronger associations for the studied item (Delaney et al., 2010). In sum, while for context and associative information one-shot per study occasion is assumed, only the item content itself is strengthened with an increasing learning duration. This "one-shot hypothesis" concerning context and associative information implicates that with a fixed learning duration, massed and distributed practice equally strengthen the item content information or storage strength (Delaney et al., 2010; Malmberg & Shiffrin, 2005). However, with massed practice there is only "one shot" of context as well as associative information while with distributed practice, with each additional learning occasion context and associative information is strengthened (Delaney et al., 2010; Malmberg & Shiffrin, 2005). Furthermore, Delaney et al. (2010) assume, that with an increasing lag between two repetitions of an item, the additional strength in context and associative information that is achieved by the repetition increases. This assumption is similar to the one of Bjork and Bjork (1992) that was described above, in that with a *decreasing retrieval strength* a *higher increase* in context and associative information is achieved. As common for study phase retrieval accounts, this only applies when retrieval in subsequent learning situations is successful. If not, a new memory trace is initiated and the original trace and its cues are not strengthened or improved (Delaney et al., 2010).

Finally, Delaney et al. (2010) assume that the lag effect in distributed practice depends on the type of test that is applied. More precisely, they state that in free recall, distributed practice compared to massed practice especially strengthens the link between context and item content information, as there are no cues involved that could boost associative information (Delaney et al., 2010). However, with long retention intervals, the probability increases that study and test contexts differ substantially. In this situation, the richer connection between context and item content information should be of little advantage and the effect of distributed practice is supposed to be small (Delaney et al., 2010). That is, regarding free recall the authors expect a decreasing effect of distributed practice with an increasing retention interval. In cued recall, in contrast, distributed practice compared to massed practice especially strengthens the association between the studied item and its cue. As this cue is presented again at the test, the positive effect of distributed practice should be particularly pronounced after long retention intervals (Delaney et al., 2010). To sum it up, Delaney et al. (2010) assume that in free recall tasks, distributed practice compared to massed practice especially strengthens memory features that present an advantage for shorter retention intervals (context information) while in cued recall tasks, distributed practice compared to massed practice is assumed to especially strengthen memory features that present an advantage for longer retention intervals (associations between item and cue).

### 4.3 Distributed practice in the classroom

Given the reliable results regarding a positive effect of distributed practice it is hardly surprising that there is a growing interest in the question of whether the effect can be transferred to real classroom settings (Seabrook, Brown, & Solity, 2005). It could be shown that distributed practice can be implemented to improve vocabulary learning of fifth and sixth graders (Küpper-Tetzel et al., 2014; Sobel, Cepeda, & Kapler, 2011) and reading ability of first graders (Seabrook et al., 2005). Distributed practice can also improve more complex skills like grammar editing skills of undergraduate college students (Miles, 2014). In another study with college students, Kapler, Weston and Wiseheart (2015) could show that an online review seven days after a science lecture improved performance five weeks after the review compared to a review only one day after the lecture. Though no real massed condition was included in this study, it is an example for a possibility to utilize distributed practice in authentic learning settings. Similar to this study, many of the studies investigating the effect of distributed practice in the classroom do not use a pure massed practice vs. distributed practice condition. Often, there is no real massed practice as control condition (e.g., Goossens et al., 2016), distributed practice is combined with other learning strategies like interleaved practice (e.g., Hopkins, Lyle, Hieb, & Ralston, 2015), or the retention interval is confounded with condition (e.g., Chen, Castro-Alonso, Paas, & Sweller, 2017; Grote, 1995). Nonetheless, the empirical foundation for the effect of distributed practice hitherto suggests that it is a strategy that can be used in real-world learning environments, too.

### 4.4 Moderators of the distributed practice effect

Features of the study material and learner characteristics that may influence the magnitude of the distributed practice effect have rarely been studied systematically. Hence, empirical findings on potential moderators of the effect are sparse. However, one important factor may be *task complexity*. In their meta-analysis, Donovan and Radosevich (1999) come to the conclusion that the magnitude of the positive effect of distributed practice decreases with task complexity. Some studies compared the effect of distributed practice among different *age* groups, resulting in inconsistent results: Studies by Maddox, Balota, Coane, and Duchek (2011) and Simone, Bell, and Cepeda (2013) found that the positive effect of distributed practice is smaller for older adults compared to the effect for younger adults. Investigating the effect of distributed practice among children and adolescents, however, Seabrook, Brown and Solity (2005) found no difference of the effect between different age groups. This could mean that the effect of age on the distributed practice effect only occurs in older adults while in younger age ranges no interaction of age and practice condition exists. *Working memory capacity* has also been investigated by Seabrook et al. (2005) as potential moderator for the distributed practice effect. However, the authors found no indication that the magnitude of the distributed practice effect depends on working memory capacity. A recent study by Delaney, Godbole, Holden and Chang (2018) led to the same conclusion. Another investigated moderator is the general performance level or *domain specific ability* of the learner. This has mainly been investigated regarding a mix of interleaved and distributed practice and not regarding distributed practice alone. However, the empirical findings from this mixed review approach can serve as a first indicator for differential effects regarding distributed practice by itself. Studies with ninth graders (Saxon, 1982) and college students (Hirsch, Kapoor, & Laing, 1982) on the effect of interleaved and distributed practice in mathematics led to the conclusion that students with lower mathematical performance benefit more from interleaved and distributed practice than students with higher mathematical performance. In contrast, Rea and Modigliani (1985) found no effect of mathematical ability on the positive effect of distributed practice in a multiplication task. However, the participants in this study learned multiplication facts by heart, potentially reducing the importance of mathematical ability. That is, in this study mathematical ability may not have been a proper indicator for the domain specific ability.

## 5. Distributed practice in mathematical learning

The material that is usually considered in studies on distributed practice can be described as rather unidimensional: Words or word-pairs have to be learned by heart through repetitions and the different item chunks that have to be learned are – content-wise – more or less incoherent. That is, typically only rote memory of isolated items is targeted. Regarding mathematics, however, repetitions are often used to practice a learned concept, but learning by heart is one component at most. That is, when a new concept or procedure is learned, exercises can help to practice the new topic or procedure, yet normally students do not practice with identical exercises but rather with conceptually similar but slightly varying tasks (e.g., to practice mental multiplication of two-digit numbers, students do not repeatedly solve 17 • 23, 17 • 23, 17 • 23; instead they would practice with different tasks like 17 • 23, 28 • 13, 27 • 23). In the following sections several important concepts regarding mathematical knowledge and practice are introduced, before research on the effect of distributed practice in mathematical learning is reviewed in greater detail.

Depending on the specific topic, two different types of mathematical knowledge can be distinguished: Following a slightly adjusted version of a definition proposed by Baroody, Feil and Johnson (2007), *conceptual knowledge* can be defined as “knowledge about facts, generalizations, and principles" (Rittle-Johnson & Schneider, 2015, p. 1119). On the other hand, procedures are defined as "series of steps, or actions, done to accomplish a goal" (Rittle-Johnson, Schneider, & Star, 2015, p. 2), and "the ability to execute action sequences (i.e. procedures) to solve problems" can be labelled *procedural knowledge* (Rittle-Johnson & Schneider, 2015, p. 1120). Generally, both conceptual and procedural knowledge do not have to be richly connected in long-term memory, but expertise is often characterized by richly connected and highly organized knowledge (Baroody et al., 2007; Rittle-Johnson & Schneider, 2015). That is, both conceptual and procedural knowledge can range from highly isolated to richly connected in long-term memory and expertise is often characterized by both richly connected conceptual and procedural knowledge. It is assumed that conceptual and procedural knowledge in mathematics develop iteratively and that gains in one of the two knowledge types also strengthens the other (Rittle-Johnson & Schneider, 2015).

Another important dimension regarding the learning of mathematics is the distinction between syntactical and semantical knowledge. Within mathematic didactics, learning on a *syntactical level* refers the ability to apply rules, for example in order to solve equation systems, whereas learning on a *semantical level* refers to the ability to recognize the connection between a mathematical symbol or representation and the aspect of reality it constitutes (e.g., understanding what a particular variable stands for) (Tietze, 1988). Syntactical knowledge can be roughly equated with rather isolated procedural knowledge. Semantical learning, on the other hand, implies a deeper understanding of the respective matter and assumes at least some degree of *connected knowledge* (conceptual and/or procedural, depending on the specific content).

Because the term procedural knowledge comprises syntactical knowledge, in the following the terms conceptual and procedural knowledge will be used to describe the *knowledge type* and the degree of semantical understanding involved in the tasks will be used to describe the *complexity* of the material. The dimensions knowledge type and complexity do not apply to material typically used in studies on distributed practice because they concern forms of learning that exceed memorization of isolated items. However, by now positive effects of distributed practice have been shown with mathematical material as well. The respective studies will be reviewed in the following, taking the dimensions knowledge type (i.e., is conceptual and/or procedural knowledge required) and complexity (i.e., the degree of semantical understanding that is targeted) into account. The focus will be on studies that specifically investigated the effect of distributed practice, excluding studies that used a combination of distributed and interleaved practice (often labelled mixed or cumulative review).

**5.1 Research on distributed practice of mathematical material**

In two early experiments, Gay (1973) investigated the effect of the temporal position of reviews on the ability to apply four different mathematical rules from algebra and geometry. Examples for the concepts that were taught are determining the exponent of a product ($a \bullet a^2 \bullet a^3 = ?$) and how the measure of the third angle of a triangle is determined (Angle A = 70, Angle B = 50, Angle C = ?). The rules were explained verbally and through examples and students worked similar but not identical examples to learn the material. In Experiment 1, 53 eighth graders received an initial learning session and participated in a review session either one day, one week or two weeks after the learning session. In the review sessions the students had the opportunity to practice the different rules again. A fourth group received no review at all. A delayed retention test was conducted three weeks after the initial learning session. That is, the interval between the last review session and the test (i.e., the retention interval) was not the same for the four experimental conditions, but the interval between the *first* learning and the test was. The analysis of the performance in the delayed retention test confirmed that all three groups that received a review session performed significantly better than the group without review. However, no significant difference between the three review groups was found, implying that the specific temporal position of the review did not have an effect on the ability to perform the mathematical rules (Gay, 1973). The author conducted a second experiment with 67 seventh graders with material very similar to the material from Experiment 1. Different from the first experiment, however, the participants received not one but two reviews (or no review). That is, after the initial learning session, the material was reviewed either one and two days later, one and seven days later, or six and seven days later. Again, a fourth group did not receive any review at all and the test was conducted three weeks after the initial learning. Similar to Experiment 1, all three groups who reviewed the learned material performed significantly better at the retention test than the group without review. Concerning the difference between the different review schedules, only the difference between the group with an early and a late review (one and seven days after learning) and the group with two early reviews (one and two days after learning) turned out to be significant, with the one and seven days-group performing significantly better than the one and two days-group (Gay, 1973). In sum, the experiments conducted by Gay (1973) confirmed that reviewing learned material improves performance up to three weeks after the learning session. Further conclusions should be drawn cautiously, because the study imposes several limitations: First, as stated above, the retention interval was confounded with condition, because the interval between the last review and the test depended on the review schedule. Second, there was no massed condition, because all review groups had interstudy intervals of at least one day. That is, the focus of the experiments lied on the lag effect instead of the effect of distributed (as compared to massed) practice. Third, the group sizes were rather small (between twelve and 18 students per group). Considering the dimensions introduced above, the material targeted rather isolated procedural knowledge, because the students had to learn rules more or less by heart and applied them, independent from a deeper understanding – though students who were able to recognize relationships to conceptual knowledge from long-term memory and hence reach a higher level of semantical understanding may have definitely profited from their insight. That is, it would have been sufficient to *know that* the exponents can be added to obtain the exponent of the product in the example “$a \cdot a^2 \cdot a^3 = ?$”, but students who *understood why* the exponents can be added may have had an advantage. However, it should be noted that this hypothesis has not been tested in the experiments.

In a study conducted by Rea and Modigliani (1985), 44 third graders learned multiplication facts either massed or distributed. The students had to learn five different multiplication facts, for example $8 \cdot 5 = 40$. In order to learn the five facts, the students were repeatedly presented ($8 \cdot 5 = 40$) and tested ($8 \cdot 5 = ?$) with the same fact (Rea & Modigliani, 1985). In the massed practice condition, presentations and tests of the same fact occurred consecutively without interruption. In the distributed practice condition, presentations and tests were interrupted with distractors. After a retention interval of one minute, all five multiplication facts were tested. Comparisons of the final performance revealed a significantly better performance in the distributed practice condition compared to the massed practice condition. Importantly, the distractors were similar to the learned material because they were also multiplication tasks (e.g. $2 \cdot 2 = ?$), which could have resulted in an additional effect of interleaved practice. In contrast to the study by Gay (1973), the material used in this study hardly targeted procedural knowledge, because the students had to learn the facts by heart. Using the dimensions introduced above, this kind of mathematical knowledge could be categorized as highly isolated conceptual knowledge (knowledge about facts which is only loosely connected to other knowledge in the long-term memory), requiring low semantical understanding, because the students are not explicitly taught how to solve the equations. In fact, this study is more similar to typical studies on distributed practice with verbal material (learning word pairs by heart is quiet similar to learning an equation by heart, neither requires any level of understanding or procedural practice).

Rickard, Lau and Pashler (2008) investigated the effect of distributed practice with similar material in that the participants had to practice mental multiplication by repeatedly solving the same problems. The participants learned 15 different multiplication problems, each consisting of the multiplication of one two-digit and one one-digit number, and the authors varied the degree of distribution by varying the set size, that is, in larger sets more intervening tasks had to be solved before the same task was repeated. There was no pure massed condition as the authors ensured that the same problem was never presented in succession. The test was conducted one week after the learning session (Rickard et al., 2008). Different from most other studies on distributed practice, the dependent variable was not performance in terms of correctly solved problems, but rather in terms of the reaction time the participants needed to solve the problem in the test. It was revealed that problems that were practiced in larger sets (high distribution due to more intervening problems) required significantly lower reaction times at the test than problems that were practiced in smaller sets (low distribution due to less intervening problems) (Rickard et al., 2008). In an additional experiment the authors investigated if the lower reaction times at the test in the high distributed practice group could be attributed to a higher use of direct retrieval of the correct solution, which would indicate that the participants had learned the results by heart. The authors argue, that the participants relied on different strategies during the learning session, depending on the set size: For problems in the small sets (low distribution), participants quickly started to retrieve the correct solution directly from the activated memory, instead of calculating the solution each time the problem was presented again in the learning session. For problems in the large sets (high distribution), the high number of intervening problems may have prevented this process for most of the items. However, problems for which the participants started to retrieve the solution directly from activated memory despite the larger number of intervening items in the high distribution condition, a strong and reliable long-term memory trace was created for the specific task and its solution, which facilitated quick retrieval in the test (Rickard et al., 2008). In their second experiment the authors confirmed that in the test, for problems in larger sets retrieval was used significantly more often than for problems in the smaller sets (Rickard et al., 2008). Taken together, the results of Rickard et al. (2008) imply that the higher degree of distribution particularly improved retrieval performance through stronger long-term memory traces rather than arithmetic ability. That is, not procedural knowledge was improved, but mainly rote memory performance. Compared to the study conducted by Rea and Modigliani (1985), the degree of procedural knowledge required in this study nevertheless was a little higher, because each time the participants were presented with a multiplication problem, they were prompted to solve it instead of simply being shown a problem and its solution. Additionally, the participants were explicitly told how to solve the problems (multiplying the one-digit number with the tens place and the ones place separately and then adding the two results). That is, in the study by Rickard et al. (2008), a higher degree of procedural knowledge was required, though the participants still repeatedly were confronted with the same problems that ultimately could be (and, based on the second experiment, actually were) learned by heart.

More complex tasks were used in two studies conducted by Rohrer and Taylor (2006, 2007). In the first experiment, 116 college students learned a formula to solve permutation problems (i.e., to calculate the number of unique orderings of a given sequence like aabccc) and solved a total of ten problems to practice the procedure. In the massed practice condition, all ten problems were worked in one session, while in the distributed practice condition, students had two practice sessions with five problems each and an interstudy interval of one week. Performance was tested one week or four weeks after the last practice session occurred. For the one-week test, no significant difference was found between massed and distributed practicing students. However, four weeks after the last practice session, distributed practicing students performed significantly better than massed practicing students (Rohrer & Taylor, 2006). In a second, very similar experiment, a significant advantage for distributed practice compared to massed practice was found even one week after the last practice session (Rohrer & Taylor, 2007). Here, the students again learned how to solve permutation problems either distributed (two practice sessions à two problems and two examples, interstudy interval one week) or massed (one practice session à four problems and four examples) or lightly massed (one practice session à two problems and two examples). Interestingly, there was no significant test performance difference between the two massed practice conditions, indicating that the number of practice tasks is less important than the timing of the practice sessions (Rohrer & Taylor, 2007). The material used by Rohrer and Taylor (2006, 2007) targets procedural knowledge (being able to correctly apply a formula in order to solve a problem) as well as isolated conceptual knowledge (knowing the correct formula). Semantical understanding was not targeted, because for one thing, it was not explained to the participants why the formula worked and for another thing, the mathematical proof is rather complex, which makes it unlikely that participants could recognize it by themselves.

In a relatively recent study that was already mentioned above, Schutte et al. (2015) investigated the effect of distributed practice on math fact fluency regarding basic addition problems. The study was conducted in school with 53 third graders who practiced addition with sums up to 18 for 19 school days either massed (four minutes each day), distributed (two minutes twice each day) or distributed (one minute four times each day). The final performance was tested at the end of the 19 days of practice (posttest) and again ten days after the last practice day (maintenance test). Performance was measured as correct digits per minute, that is, the dependent variable reflects not only addition accuracy but also addition speed. No significant difference was revealed for the posttest- and maintenance performance of the students of the two distributed practice conditions, but students of both distributed practice conditions performed significantly better than students of the massed practice condition (Schutte et al., 2015). It should be noted that similar to most of the presented studies there was no real massed practice condition, because students of all conditions practiced each day for 19 school days. The results are promising nonetheless because practicing non-identical addition problems actually targets procedural knowledge in that the correct solutions could not be learned by heart and performance could not be attributed to rote memory. However, the level of semantic understanding required for the addition tasks can be considered relatively low.

The last study on distributed practice with mathematical material is a classroom study conducted by Chen et al. (2017). In their first experiment, 54 fourth graders learned how to add positive fractions with different denominators by means of worked examples and exercises that had to be worked individually. In the massed practice condition, three worked examples, each followed by an exercise that had to be solved individually, were worked consecutively in one session. After the completion of a working memory test, the posttest was conducted. In the distributed practice condition, on the other hand, only one worked example and one individually worked exercise was presented at a day, three days in a row. On the fourth day, again following a working memory test, the posttest was conducted (Chen et al., 2017). That is, the retention interval differed between the conditions by one day, because massed practicing students were tested the day of practice and distributed practicing students were tested the day after the last practice session. Comparing the performance in the posttest task, students of the distributed practice condition performed significantly better than students of the massed practice condition. In a second experiment, 61 fifth graders learned two topics, one with distributed practice and one with massed practice. The topics were calculation with negative numbers and solving fractional equations. The procedure was similar to the first experiment, that is, three worked examples, each followed by an individually worked exercise, were presented either massed on one day or distributed over three days. As in Experiment 1, the retention interval for the distributed practice condition was one day while for the massed practice condition performance was tested the day of the learning session. Again, practicing in a distributed fashion resulted in significantly higher posttest performance scores than practicing in a massed fashion (Chen et al., 2017). In comparison to the other studies reviewed in the present section, the material used by Chen et al. (2017) is among the ones with the highest complexity. The exercises required procedural knowledge and additionally, both solid conceptual knowledge and semantical understanding could improve the understanding of the matter and with that the final performance. For example, the fourth graders could simply follow the instructions on how to solve addition problems of fractions with different denominators even without really understanding them (relying only on rather isolated procedural knowledge), but recognizing basic principles of fractions or maybe even relating the material to real-world problems (i.e., including conceptual knowledge by connecting the newly learned procedure to knowledge from long-term memory and thereby increasing semantical understanding) could have potentially improved the performance of the students.

### 5.2 Distributed practice in mathematical learning – conclusion

To sum up, although there are several studies investigating distributed practice with mathematical material, they differ in important aspects, which makes a systematic and clear conclusion difficult. Most importantly, many of the reviewed studies did not include a real massed condition but rather compared different degrees of distribution (Gay, 1973; Rickard et al., 2008; Schutte et al., 2015), which basically means they investigated the lag effect instead of the effect of distributed practice. Another common problem was that in several experiments different retention intervals where imposed for different conditions (Chen et al., 2017; Gay, 1973), resulting in a confounding of practice condition and retention interval. Finally, though all studies were on learning of mathematical material, the actual complexity differed severely. While the exercises used in some studies basically required rote memory of isolated mathematical facts and hence were very close to classical studies on distributed practice (Rea & Modigliani, 1985; Rickard et al., 2008), others required procedural knowledge (Chen et al., 2017; Gay, 1973; Rohrer & Taylor, 2006, 2007; Schutte et al., 2015). However, even regarding these studies it still could be questioned if the level of conceptual knowledge and semantical understanding that may have improved the performance was comparable (e.g., regarding simple multiplication facts semantical understanding may have less of an impact than regarding fractional equations). Astonishingly, despite all these differences in nearly all presented experiments distributing mathematical practice resulted in superior performance compared to massed practice, indicating that distributed practice is a promising strategy for different forms of mathematical practice, too. However, in order to draw more reliable conclusions concerning distributed practice in mathematical learning, more studies including massed practice conditions and with comparable retention intervals would be desirable. Additionally, using more complex material could help to pin down if distributed practice only improves the memory aspect of mathematical practice (i.e., memory of isolated conceptual knowledge like a specific equation and its solution or a correct formula) as indicated by the results of Rickard et al. (2008), or if procedural knowledge (i.e., the ability to perform a specific procedure) and even semantical understanding (i.e., understanding why the procedure is correct/works) benefit as well.

## 6. Summary and research questions

The previous sections served as an introduction to basic memory models and theoretical approaches to learning and forgetting. One important feature of successful learning strategies is that they enhance processing during learning, resulting in stronger, more reliable and richly connected memory traces. The other important feature is practice of the retrieval process that is required during test situations, when information must be re-entered into the focus of attention, either from activated or from non-activated long-term memory (Bjork, 1994). One learning strategy that serves both these purposes, that is, enhanced processing and practice of retrieval, is distributed practice – which is addressed in this dissertation.

In various empirical studies, distributed practice has been proven to improve retention of verbal material compared to massed practice, at least intermediate and long-term (Carpenter et al., 2012; Cepeda et al., 2006), yet it has seldom been investigated using mathematical material. However, mathematical practice differs from rote memory learning of verbal material in significant aspects: Most importantly, mathematical practice is usually applied to improve procedural knowledge (Rittle-Johnson & Schneider, 2015) – not by learning a fact by rote but by practicing a given procedure with similar, but varying exercises. Additionally, mathematical topics are often more complex than verbal material because they are not learned as isolated facts (like random word pairs – a regular content of studies on distributed practice), but can be connected to conceptual and procedural knowledge stored in long-term memory and thereby understood either superficially or on a deeper level (Tietze, 1988). Considering the theoretical model proposed by Delaney et al. (2010), it can be assumed that distributed practice compared to massed practice improves mathematical performance as well, because besides the positive effect on retention, the increase in associations (connections to other knowledge in long-term memory) that is associated with distributed practice in particular, could also have the potential to increase semantical understanding. To sum it up, learning mathematics by practice differs considerably from rote memory learning, which means that based on studies with verbal material alone, the positive effect of distributed practice cannot automatically be expected to emerge in mathematics, too. Although, a positive effect of distributed practice compared to massed practice on mathematical performance is likely, this should be verified empirically.

While there are a few studies on distributed practice with mathematical material, it became evident that they differ concerning several aspects. Apart from design issues (e.g., the confounding of practice condition and retention interval, Chen et al., 2017), a crucial aspect were the different levels of complexity. Some studies were basically identical to classical studies on distributed practice (e.g., instead of word pairs, multiplication facts were learned by heart, Rea & Modigliani, 1985), and only few used more complex material, and had the participants actually practice with non-identical exercises (e.g., Rohrer & Taylor, 2006, 2007). Based on the available studies, however, a positive effect of distributed practice on mathematical performance can be assumed.

An important factor for this lack of complex (mathematical) material in experiments on distributed practice may be that in order to be able to practice something, participants first have to be taught the respective concept or procedure, which increases the complexity of the experimental design. Taking the studies by Rohrer and Taylor (2006, 2007) as an example, before having been shown how to calculate the number of permutations, participants were not able to practice the respective formula. That is, studies on the effect of distributed practice on mathematical performances almost always need some kind of lecture or tutorial. A way to address this is to investigate the *effect of* *distributed mathematical practice in school*. On the one hand, it is commonplace for school students to first learn a concept and then practice it with exercises, so the experimental procedure can be implemented quite naturally. On the other hand, investigating the effect of distributed practice in school additionally offers the opportunity to validate the effect of distributed practice in a real, authentic learning context – something that has only rarely been considered in previous studies on distributed practice, especially regarding mathematics. For these reasons, in the studies presented in the following three chapters of this dissertation the effect of distributed practice on mathematical performance has been investigated in school.

### 6.1 Study overview

In each study, the topic was chosen from the regular mathematics curriculum and the material was prepared with the aim to create a learn- and practice-situation as authentic as possible. The participating classes had not covered the topic in the respective school year and first learned about it in lecture sessions conducted by the study team. Afterwards, the students practiced the covered topics with exercises either massed (all exercises worked consecutively in one session), or distributed over several days. The performance was tested one to six weeks after the last practice session in order to assess long-term effects, too. Because differential effects of distributed practice have only rarely been considered, additionally in each experiment several learner characteristics were examined as potential moderators.

In the first study, *Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom*, two experiments were conducted. The main purpose of the first study was to investigate whether distributed practice generally improves performance of school students learning mathematics. In the first experiment, third graders learned a semi-formal multiplication method, and in the second experiment, seventh graders learned basic stochastics. Afterwards, students in both grades practiced the respective topics with three practice sets, each containing multiple exercises on the covered topics. Students of the massed practice condition worked all three practice sets in one session, students of the distributed practice condition worked one practice set per day for three consecutive days. The students were provided summary sheets with examples during the practice sessions but received no correct solutions. The performance was tested twice, first one week after the last practice session and then again six weeks after the last practice session. All lecture, practice and test sessions took place in school and were supervised by the study team.

In the second study, *Distributed practice in mathematics: Recommendable especially for students on a medium performance level?,* one experiment was conducted. The main purposes of the second study were to a) substantiate the results regarding the seventh graders from the first study and b) investigate if the inclusion of correct solutions during practice results in an overall increased performance level and potentially also in a larger effect of distributed practice. Therefore, the experiment of the second study was conducted in Grade 7, with lecture sessions and practice sets very similar to the second experiment of the first study. However, this time the students received the correct solution as feedback each time they finished an exercise. An expanding interval schedule was applied for the distributed practice condition: Students of the massed practice condition again worked all three practice sets in one session, but students of the distributed practice condition had two days between the first and the second practice set, and five days between the second and the third practice set. The performance was tested two and six weeks after the last practice set. Again, the lecture, practice and test sessions took place in school and were supervised by the experimental team.

In the third study, *Distributed practice: Rarely realized in self-regulated mathematical learning*, one experiment in high school math courses was conducted with the aim to investigate whether distributed practice proves a successful learning strategy in a more self-regulated context, too. The main purposes of this study were to a) expand the considered age range and to b) investigate if online exercises may offer a simple way to implement distributed practice in school teaching and learning. As in the other studies, the students were taught a topic that was afterwards practiced. This time, however, only the lecture sessions took place in school and the practice and test sets were worked online at home, using the same expanding interval procedure that was used in the second study. This setting, that is, students receive some input in school and afterwards have to practice the content at home, is typical for high school where learning generally is more self-regulated than in the lower grades. The performance was tested two weeks after the last practice set was worked.

# CHAPTER II: Study 1

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

Katharina Barzagar Nazari & Mirjam Ebersbach

A version of this chapter is published as:

Barzagar Nazari, K., & Ebersbach, M. (2018). Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom. *Applied Cognitive Psychology*. doi: 10.1002/acp.3485

**Abstract**

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 the regular curriculum. Thereafter, they practiced in one of two conditions: In the massed condition, they worked three practice sets on one day. In the distributed condition, they worked one practice set per day for three consecutive days. Bayesian analyses of the performance in two follow-up tests one and six weeks after the last practice set revealed a positive effect of distributed practice as compared to massed practice in Grade 7. In Grade 3, a positive effect of distributed practice was supported by the data only in the test one 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.

## 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, 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 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 with *massed practice* (Bjork & Bjork, 2011). The effect of distributed practice is widely studied and considered to be robust, with the long-term retention performance after distributed practice usually exceeding performance after massed practice (Carpenter et al., 2012; Küpper-Tetzel, 2014).

Different theoretical approaches have been proposed 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.

**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), while 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. Instead, 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 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., Kapler, Weston, & Wiseheart, 2015), for example, by investigating the distributed practice effect using 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, & Spirgel, 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 pure memorization and can be defined as “the ability to execute action sequences (i.e., procedures) to solve problems”, whereas conceptual mathematical knowledge 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 four min over the course of 19 school days. They either practiced massed each day for four 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 on the immediate post-test 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 other experiments conducted by Chen, Castro-Alonso, Paas and Sweller (2017), fourth and fifth graders who practiced a mathematical procedure distributed over three days achieved higher performance scores than students who practiced massed on one 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.

**Research objective and hypotheses of our study**

The purpose of our study was to examine the effect of distributed practice on mathematical performance in the classroom. Brief mathematics lessons were developed for third and seventh graders based on their regular curricula. Students were not required to learn facts taught in the classroom by heart but instead 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 one week after the last practice session, and that this difference would be even more marked in the six-week follow up test, because the positive effect of distributed practice becomes usually more evident in the long run (Küpper-Tetzel, 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 one week after the last practice than students practicing with the same exercises in a massed fashion. Furthermore, the positive effect of distributed practice compared to massed practice on the test performance should be even bigger after six weeks than after one week in both grades.

## Method

**Participants**

In total, 141 third graders from five elementary schools and 171 seventh graders from four secondary schools in Germany were recruited [1]. These age groups were selected, because they cover two different 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 socio-economic 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

[1] This study was carried out in accordance with the recommendations 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.

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 have been present during all lesson, practice and test sessions. After exclusions, the data from 95 third graders (51 female, 44 male; $M_{age}$ = 9 years, 6 months, age range: 8-11 years) and 118 seventh graders (66 female, 52 male; $M_{age}$ = 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 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.

**Design**

Practice schedule was manipulated between subjects. Students practiced the learning content (i.e., mathematical procedures) either *massed* on one day or *distributed* over the course of three 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. Therefore, 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 one week and six 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 supplemental material provided online (https://osf.io/vfcgz/?view_only=21d3cbd9c2e04119864b2096a2f8ca10).

**Procedure**

In Grade 3, the introductory lesson lasted 90 min and was used to introduce a semi-formal multiplication method. In Grade 7, the introductory lessons consisted of two 90-min lessons, separated by two 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 semi-formal multiplication that should have been acquired previously at least to some extent (e.g., semi-formal 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 a total of 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* solved only the first sheet. On the following two days, students in the distributed condition worked 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 three 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 five to seven days. The greater time lag compared to 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.

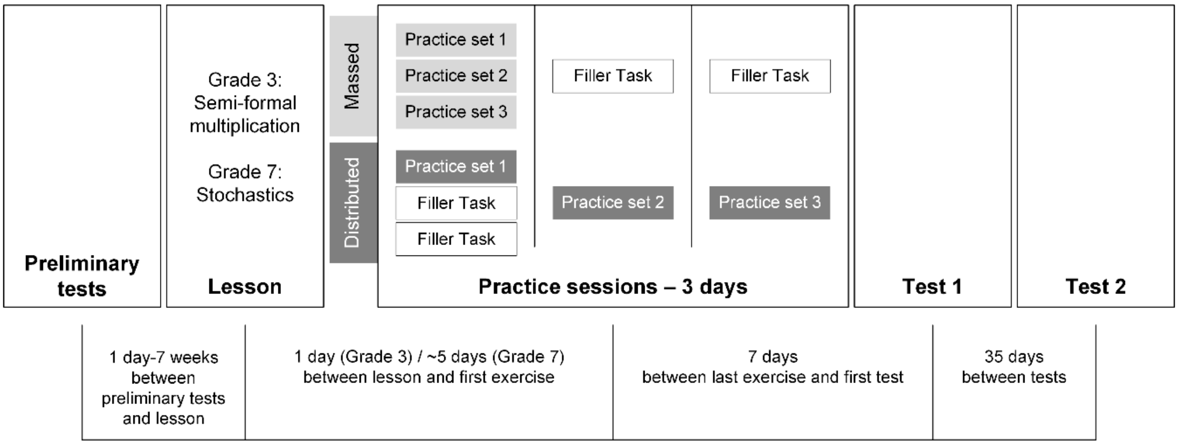



*Figure 1.* Schematic of the experimental procedure.

While working on the practice sheets, each student had access to a summary sheet containing examples, including solutions, from the introductory lesson. This procedure was supposed to simulate a homework situation in which all classroom material is usually available. We ensured that examples and exercises were different to prevent students from simply copying the solutions. 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 regardless of the learning condition, they were not further analyzed. Information on the transfer tasks is provided online in the supplemental material (https://osf.io/vfcgz/?view_only=21d3cbd9c2e04119864b2096a2f8ca10).

The first test was regularly conducted seven days after the last practice opportunity. In one school, 26 students of the distributed practice condition in seventh grade were already tested after six days because of a school holiday. One single seventh grade student of the distributed practice condition was already tested after five days due to an organizational mistake. These deviations that occurred in Grade 7 will be discussed in the Results section. The second test regularly took place five weeks after the first test (six weeks after the last practice opportunity). Two seventh grade classes were tested one 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 ten participating students was tested ten to twelve 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: semi-formal 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 semi-formal multiplication in class after the first but before the second test was conducted. This will be considered in the Results section, too.

**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. Semi-formal multiplication is a procedure to solve multiplication tasks including multi-digit numbers. Within this study, only tasks with one one-digit number and one two- or three-digit number were taught. The 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 semi-formal 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. The complete material including the lesson scripts can be inspected in the supplemental material provided online (https://osf.io/vfcgz/?view_only=21d3cbd9c2e04119864b2096a2f8ca10).

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 practice 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).

## Results

The data and R scripts for data preparation and analyses are provided online (https://osf.io/vfcgz/?view_only=21d3cbd9c2e04119864b2096a2f8ca10) [2].

### 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.

Table 1

*Performance of Third Graders (Mean Percentage of Correctly Solved Tasks)*

| | Practice sheets | | | Tests | |
|---|---|---|---|---|---|
| Practice condition | 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 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 one week after the last practice set and one addressing the

[2] The data was 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).

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 possible 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. That is, 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 conversion.

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 one 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 condition 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 one week after the last practice as compared to massed practice, than that distributed practice has no effect or a negative effect. 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 set unsurprisingly can also be expected to achieve higher scores in the first test, compared to third graders with a poorer performance in the first practice set.

The second model was calculated to analyze the third graders' performance change from the first to the second test, conducted six weeks after the last practice. It was hypothesized that the positive effect of distributed practice would be even more pronounced six 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, for students of the massed practice condition a higher performance gain is estimated). 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. More precisely, it is only 0.1 times as likely that distributed practice has a positive effect than that it has no effect or a negative effect, 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.

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 six 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 six weeks after the last practice as compared to 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 six weeks is much less conclusive compared to the effect one 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 to massed practice – has a positive effect on the performance after six weeks as that it has no effect or a negative effect. That is, six 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).

As was stated in the procedure there were two important points that have to be considered regarding the third grade: 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 semi-formal multiplication had been reviewed after the first, but before the second test was conducted. 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. Here, it could be confirmed that removing the students did not result in an effect of practice condition. 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.

**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.

Table 2

*Performance of Seventh Graders (Mean Percentage of Correctly Solved Tasks)*

| | Practice sheets | | | Tests | |
|---|---|---|---|---|---|
| Practice condition | 1 | 2 | 3 | 1 Week | 6 Weeks |
| Massed | 63%<br>(56-70%) | 75%<br>(67-81%) | 59%<br>(52-66%) | 44%<br>(36-52%) | 45%<br>(38-53%) |
| Distributed | 61%<br>(54-68%) | 80%<br>(75-85%) | 55%<br>(49-62%) | 51%<br>(44-58%) | 51%<br>(42-59%) |
| *N* | 118 | 118 | 118 | 118 | 118 |

*Note*. 95%-confidence intervals in parentheses.

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 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 0-12-point score range. The evidence ratio for the effect of distributed practice on the performance one 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. 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 as well. Again control analyses were performed because 23 of the 118 students were tested one or two 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 evidence ratio of 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 only 0.62 times as 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. Because an evidence ratio of 1 would mean no effect, this pretty much leads to the conclusion that contradicting our hypothesis the practice condition 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).

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, an additional model was conducted to analyze the performance in the second test directly. With performance six 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 one week after the last practice set. The evidence ratio for the positive effect of distributed practice was 10, that is, 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 six weeks after the last practice, compared to no effect or a negative effect, which is considered strong evidence (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 a higher performance in the first practice set is related to a higher score in the test. Finally, for the second test again a control analysis was performed because some students were tested one 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 evidence ratio (8 instead of 10). Again, the control analysis confirms the central results of the main analyses which therefore will be further referred to.

## Discussion

The main purpose of this study was to investigate the effect of distributed practice on the mathematical performance within an authentic educational setting. Students of elementary and secondary schools practiced mathematical procedures either massed in one day or distributed across three days. The total duration of practice was constant in both conditions. Performance was tested after retention intervals of one and six 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 one 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 ten 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 one week after the last practice (i.e., about 1-point performance gain for the distributed practice condition within a 0-12-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, while the seventh graders had two spaced introductory lessons and started practicing five to seven days after the second lesson. Nevertheless, both age groups benefitted from distributed practice in a similar manner in the test one 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 six 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 semi-formal 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.

**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 one week, but not six weeks, after the last practice set. In Grade 7, the effect emerged both after one and after six 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.

**Acknowledgments**

We would like to thank Wiebke Eden, Ulrike Haßfurther, Jessica Kühnel and Kristin Wenzel for their support in conducting the experiments and Dorothea Hentzschel, Kathrin Hillebrand, Damaris Keller, Klara Schelbert, Natascha Schupp and Maike Siegle for also teaching the participating classes.

## Appendix. Exercise examples

The complete material can be inspected in the supplemental material provided online (https://osf.io/vfcgz/?view_only=21d3cbd9c2e04119864b2096a2f8ca10).

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

**Exercise 1**

a) **13** · **6** =

· =

· =

b) **3** · **159** =

· =

· =

· =

c) **5** · **58** =

· =

· =

d) **243** · **4** =

· =

· =

· =

Example practice sheet for the seventh graders. Each practice sheet consisted of three exercises.

**Exercise 1**

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

Answer: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

**Exercise 2**

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

Answer: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

**Exercise 3**

There are 80 apples in one basket. 12 of them are unripe, 8 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.

**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 societ*y (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 Robert 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 doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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*, 1–19. doi: 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. doi: 10.1016/S0022-5371(72)80001-X

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, 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* (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. doi: 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. doi: 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. doi: 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 partytioning 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. doi: 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. doi: 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. doi: 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.

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. doi: 10.1002/acp.1266

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6), 481–498. doi: 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. doi: 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. doi: 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. doi: 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* (Vol. 60, pp. 113–189). Academic Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352–373. doi: 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

# CHAPTER III: Study 2

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

Katharina Barzagar Nazari & Mirjam Ebersbach

**Abstract**

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

## Introduction

One important aim of learning in school and other real-world learning contexts is to acquire knowledge and skills that can be retrieved not only after a short delay but also in the long run. However, many strategies usually applied by learners, such as repeatedly reading the learning content, promote rather short-term retention (Bjork, Dunlosky, & Kornell, 2013). Thus, the knowledge might be retrieved in the next exam but then quickly starts to decay. Several mechanisms have been proposed that aim at promoting long-term retention in particular, by aggravating the learning process for the learner. These mechanisms are labelled *desirable difficulties* (Bjork, 1994) and are related to a range of successful learning strategies, for instance, the *generation* of the information to be learned by the learner him- or herself instead of receiving the complete learning material in advance (McDaniel, Waddill, & Einstein, 1988). Furthermore, *testing* can be used as a learning strategy, too, when the learner tries to recall the to be learned content already in the learning phase – instead of repeatedly rehearsing the material (Karpicke & Roediger III, 2008). *Spacing* or *distributed practice* means that the total time for learning is distributed across at least two learning sessions – instead of massed in only one learning session (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Finally, with *interleaved practice* the learning contents of at least two related topics are covered alternately – instead of covering one topic first and the other topic thereafter (Rohrer & Taylor, 2007).

The positive effects of learning strategies based on desirable difficulties were demonstrated in a large number of studies that also have been entered into meta-analyses, suggesting in general robust effects for the mentioned strategies (generation: Bertsch, Pesta, Wiscott, & McDaniel, 2007; testing: Adesope, Trevisan, & Sundararajan, 2017; Rowland, 2014; interleaved practice: Brunmair & Richter, 2018; Rohrer, 2012). For distributed practice, which is the strategy central for the present study, meta-analyses yielded a significant advantage for distributed compared to massed practice, with medium to large effect sizes (Cepeda et al., 2006; Donovan & Radosevich, 1999; Janiszewski, Noel, & Sawyer, 2003). Moreover, longer retention intervals were shown to profit from longer temporal distances between the distributed practice sessions, although this relationship is not considered to be linear (Cepeda et al., 2006).

Several theoretical accounts try to explain the positive effect of distributed practice (for overviews see Küpper-Tetzel, 2014; Toppino & Gerbier, 2014). One account assumes *deficient processing*. Here, it is supposed that when a person learns an item by successive repetitions (as in massed practice), the person (intentionally or not) processes the item more poorly than when the repetitions are separated by time lags (as in distributed practice). Thus, deficient processing takes place in massed practice, whereas distributed practice enhances processing and thereby improves retention performance (Craik & Lockhart, 1972; Crowder, 1976; Hintzman, 1974). Another potential explanation for the superior performance after distributed practice compared to massed practice is *study phase retrieval*. The central assumption here is that the retrieval of a learned item or chunk of information supports long-term retention, especially when the retrieval process itself is hard, yet successful. When the lag between learning situations increases, as in distributed practice, retrieval is harder than with massed practice, but it maximally strengthens the memory trace if retrieval is successful. In massed practice, on the other hand, retrieval throughout the learning period should be fairly easy, because little time passed since the last retrieval attempt. Unfortunately, thereby the memory trace is hardly strengthened, resulting in poorer long-term performance (Bjork, 1975; Küpper-Tetzel, 2014; Thios & D'Agostino, 1976). Finally, *encoding variability* theories are based on the assumption that a learned item is stored together with retrieval cues. These retrieval cues can be of various kinds and consist, for example, of the emotional state during learning, physical information of the learning environment, or associations between the item and prior knowledge. It is assumed that in massed practice, where repetitions follow immediately one after another, the item is encoded repeatedly in roughly the same way, resulting in little growth of relevant retrieval cues. With distributed practice, in contrast, more time passes between the repetitions and chance is thus higher that the item is encoded in a different way, accompanied by different retrieval cues. The greater variability of retrieval cues after distributed practice should then boost performance compared to massed practice (Crowder, 1976; Glenberg, 1979; Toppino & Gerbier, 2014). By now, the different approaches proposed to explain the positive effect of distributed versus massed practice are assumed to be not exclusive. Instead, several accounts might work simultaneously. Most attempts to formulate such hybrid theories combine different variations of study phase retrieval and encoding variability approaches (Delaney, Verkoeijen, & Spirgel, 2010; Verkoeijen, Rikers, & Schmidt, 2004, 2005; Young & Bellezza, 1982).

**Distributed practice of mathematical procedures**

The large majority of studies on desirable difficulties in general, and on the distributed practice effect in particular, included adults as participants who were tested in the laboratory, involving rather incoherent learning contents that referred to rote memory of declarative knowledge, such as word lists or pictures (Carpenter, Cepeda, Rohrer, Kang,

& Pashler, 2012; Cepeda et al., 2006; Dempster, 1988; Kang, 2016). Although the interest in examining the effect of distributed practice in applied learning settings is increasing (e.g., Kapler, Weston, & Wiseheart, 2015), it is still a fairly open question of whether distributed practice yields positive effects if curriculum-based learning content is used, aiming for example at promoting the acquisition of procedural knowledge, such as in mathematics. The present study addresses this question by examining whether distributed practice of mathematical material in the classroom leads to better performance than massed practice in secondary school students.

Concerning distributed practice in mathematics, there are two studies with college students who practiced permutation problems either in a distributed fashion with a 7-days-delay or in a massed fashion on one day. The distributed practice group clearly outperformed the massed practice group after one week (Rohrer & Taylor, 2007) as well as after four weeks (Rohrer & Taylor, 2006). Yazdani and Zebrowski (2006) used a combination of distributed and interleaved practice for the geometry homework of high school students and showed that working on exercises of previous and current topics (spacing + interleaving) improved performance compared to working on exercises only of current topics (massing), up until six weeks after the learning period. A similar approach was already pursued by Hirsch, Kapoor, and Laing (1982) and Saxon (1982) who used a combination of distributed and interleaved practice for mathematics assignments in college (Hirsch et al., 1982) and school (Saxon, 1982) and found that this "mixed review homework strategy" led to better performance in students in particular in the lower and medium performance ranges (for a review on mixed review in mathematics see also Rohrer, 2009).

To our knowledge, there are only three studies investigating the isolated effect of distributed practice in mathematics in the school context. Schutte et al. (2015) had third graders practice basic addition problems by presenting them with four 1-min practice sessions (a) consecutively (i.e., massed), (b) distributed with two back-to-back sessions in the morning and two back-to-back sessions in the afternoon, and (c) all four sessions separated and distributed across one day. The three practice procedures were repeated across 19 days in total. Practice was explicitly timed and the number of correct digits per minute at two tests, reflecting basic math fact fluency, served as dependent variable. Third graders in both distributed practice conditions gained more basic math fact fluency across the 19 days than third graders in the massed practice condition. However, one might argue that all three groups practiced in a distributed manner as they practiced addition across multiple days.

Chen, Castro-Alonso, Paas, and Sweller (2017) compared the performance of 4th and 5th graders who practiced fraction addition (Grade 4), calculation with negative numbers, and solving fractional equations (both Grade 5) either massed on one day or distributed across three days. The authors found that distributing practice over three days significantly improved performance compared to massed practice in both grades. However, these results have to be interpreted with caution because the students of the massed practice condition were tested immediately after finishing their practice exercises while the students of the distributed practice condition were tested one day after they had finished their practice exercises. Thus, the retention interval was not held constant across both conditions, which might have caused an adverse effect on the attention and motivation of the massed practicing students.

Another direct comparison between the effects of massed and distributed practice of mathematical procedures in school was conducted by Barzagar Nazari and Ebersbach (2018). They taught third and seventh graders a mathematical procedure previously unknown to the students (i.e., semi-formal multiplication and probability calculation, respectively) that was then practiced in the classroom either massed for 45 min on one day or distributed across three consecutive days for 15 min each. The performance in similar but not identical tasks was assessed in two tests taking place one week and six weeks after the last practice session. Based on Bayesian analyses of the test results, there was very strong evidence for a positive effect of distributed practice compared to massed practice one week after the last practice session in Grade 3, but the effect diminished in the second test, conducted six weeks after the last practice session. It was assumed that while the third graders did not practice semi-formal multiplication in class until the last test was conducted, they still most likely practiced related skills within other topics that were covered (e.g., they could have practiced mental multiplication, which is needed for semi-formal multiplication), which could then have led to a decreasing performance difference between the two practice conditions in the second test. In Grade 7, in contrast, there was strong evidence for a positive effect of distributed practice compared to massed practice on both the intermediate and long-term test performance. However, contrary to the hypothesis that the positive effect of distributed practice would become more pronounced in the long run (Küpper-Tetzel, 2014), it diminished (as in Grade 3) or remained stable (as in Grade 7). One special feature of the described study was that the students of both grades did not receive feedback on their practice performance nor were they provided with the correct solutions. Though it is encouraging that there was evidence for a positive effect of distributed practice even without feedback or correct solutions, the question remains of whether the effect increases or is more pronounced in the long run with feedback during the practice sessions. Thus, students in the current study received feedback by being presented with the correct solution once they had completed a practice task.

**The influence of individual characteristics on the effectivity of desirable difficulties**

Another relevant – but so far rather neglected – aspect concerning the applicability of desirable difficulties in real learning contexts is whether the effects emerge for all learners in a similar manner or whether they depend on learners' *motivational and cognitive characteristics*. This is an important question in particular if one strives to give recommendations to teachers concerning the use of desirable difficulties in school. Such aptitude-treatment-interactions (Cronbach & Snow, 1977; Snow, 1989) have been established in many instructional areas and are supported by empirical findings (e.g., Hattie, 2008). However, with regard to desirable difficulties, such interactions were theoretically assumed, too (e.g., McDaniel & Butler, 2011), but rarely tested. With regard to distributed practice, which is the desirable difficulty that is in focus of the current study, indications of moderating variables are sparse. While the effect of distributed practice has also been demonstrated for a broad *age range* including children (e.g., Seabrook, Brown, & Solity, 2005), the role of individual differences with regard to motivational or cognitive variables is less clear. A few studies examined the effect of *prior knowledge* – some yielding a larger benefit by distributed and interleaved practice for students with low prior knowledge (Hirsch et al., 1982; Saxon, 1982), while others found no such interaction (Rea & Modigliani, 1985). Studies on the effect of working memory capacity seem to indicate that there is no interaction with the effect of distributed practice (Delaney, Godbole, Holden, & Chang, 2018; Seabrook et al., 2005).

Barzagar Nazari and Ebersbach (2018) addressed the aspect of individual differences concerning the effect of distributed practice more broadly by additionally assessing various cognitive variables (i.e., initial practice performance, working memory capacity, sustained attention, metacognitive monitoring) and motivational variables (i.e., mathematical self-efficacy, effort motivation, performance-avoidance goals, work avoidance) of third and seventh graders in their sample. However, none of these variables yielded a moderating effect. Given the sparse empirical basis concerning moderating variables on the effect of distributed practice, we exploratory addressed this aspect in the present study, too: One of the challenges of distributed practice compared to massed practice is that learners have to repeatedly engage in a topic that they may already have forgotten about, because the last learning or practice session was several hours or days ago. For students lacking (mathematical) self-efficacy, this may result in a complete surrender to the topic because they cannot keep up while repeatedly being confronted with their own failings (Zimmerman, 1995). A similar reasoning can be applied to individual effort motivation, with students displaying high effort motivation being more likely to keep working even when they face difficulties in distributed practice sessions. On the other hand, students displaying low effort motivation might be tempted to stop to engage with the topic in distributed practice, because they repeatedly have to retrieve information that has been forgotten again, eventually reducing the positive effect of this learning strategy. Concerning concentration abilites, however, it might be especially those with poor concentration who benefit most from distributed practice because it requires to concentrate for a shorter time compared to massed practice. Hence, *mathematical self-efficacy*, *effort motivation* and *self-rated concentration difficulty* were considered to be potential moderators for the effect of distributed practice in this study. Due to a programming error, however, only self-efficacy and concentration difficulty could be evaluated (see Design).

**Research questions and hypotheses**

With this study, we aimed at extending the empirical grounds regarding the usability of distributed practice for mathematical learning and regarding the question of whether all learners benefit from this learning strategy to a similar extend or not. Therefore, we examined the effect of distributed mathematical practice with feedback among secondary school students and investigated two potential moderating motivational variables. We expected that the inclusion of feedback in the practice sessions would enhance students' performance during practice and might help to boost the effect of distributed practice in particular in the long run. Students practicing mathematical procedures in a distributed fashion were expected to outperform students practicing the same procedures in a massed fashion in an intermediate test, conducted two weeks after the practice. In a long-term test, conducted six weeks after practice, the positive effect of distributed practice was expected to be even larger, because the effect of distributed practice is known to be especially pronounced in the long run (Küpper-Tetzel, 2014). In addition, we exploratory analyzed moderating effects of motivational variables of the learners that theoretically have the potential to explain individual differences regarding the effect of distributed practice.

## Method

### Participants

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

### Design

Practice condition served as independent variable and was manipulated between subjects: Within each class, one group of students practiced massed on one day and the other group practiced distributed on three different days. For the distributed practicing students, an expanding interval schedule was applied (see Material and Procedure; Küpper-Tetzel, Kapler, & Wiseheart, 2014). Before the students were assigned to one of the two practice conditions, they were ranked by their mathematics grade of their last school certificate, and then, within each grade group, they were randomly assigned to one of the two conditions, ensuring that the overall math performance was roughly equal in both conditions before manipulation. The final performance in the practiced content (i.e., probability calculation) served as dependent variable and was tested two and six weeks after the last practice set was finished. To examine if individual learner characteristics moderated the effect of distributed practice, we additionally included a questionnaire assessing three motivational characteristics. However, for the items on effort motivation accidentally the wrong answer scale was programmed and because of this, the items could not be evaluated. In Table 1, the variable scales of the two remaining measures are displayed along with translated sample items and information on their reliability based on our data (Cronbach’s alpha). Finally, after finishing an exercise the students were asked

[1] This study was carried out in accordance with the recommendations 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.

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

Table 1

*Instruments Used to Assess Potential Moderators*

| Potential moderators | Employed instrument | Reliability |
|---|---|---|
| Mathematical self-efficacy | 7 items of a German *Academic Self-Efficacy Scale for School Children* (Jerusalem & Satow, 1999), adapted to mathematics<br><br>(Sample item: "In math, I can solve even the difficult problems if I try hard.") | α = 0.87 |
| Concentration | 6 items of the German *Learning Strategies in College – LIST* (Boerner, Seeber, Keller, & Beinborn, 2005; Wild & Schiefele, 1994)<br><br>(Sample item: "When I'm learning, I'm easily distracted.") | α = 0.91 |

**Material and Procedure**

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

**Prior testing and introductory lessons.** In the first session (about 45 min), all students individually answered a questionnaire on tablets concerning two motivational characteristics (see Table 1). The questionnaire was programmed with the survey tool LimeSurvey (Project Team / Schmitz, 2012). Thereafter, students' prior knowledge on probability calculation was assessed via pen-and-paper to ensure that the students had in fact no significant experiences concerning this topic before the study started. The session was directly followed by the first introductory lesson on probability calculation (45 min). A second, longer introductory lesson on probability calculation (90 min) was held one or two days after the first introductory lesson, depending on the class schedule. Both introductory lessons were held by student assistants experienced in teaching children and adolescents, who were supervised by the authors. The introductory lessons were designed as normal school lessons and included examples and short experiments (e.g., coin tossing). They were very similar to the lessons used for the abovementioned study by Barzagar Nazari and Ebersbach (2018).

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

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

correct solution was shown, the students were asked how hard they considered the exercise on the previous page[3]. It was not possible to click back to previous pages.

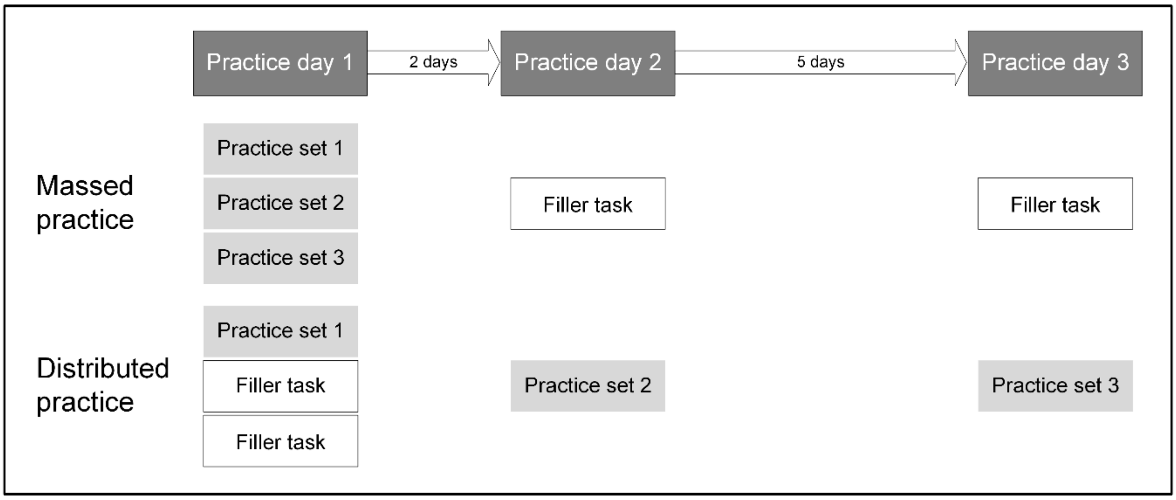



*Figure 2*. Illustration of the two practice schedules. Each practice set and filler task took no more than 15 minutes.

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

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

The students did not receive any individual feedback on their performance in practice sets (the same sample solution was provided regardless of the answer given by the student) and no feedback at all on the performance in test sessions. The student assistants supervising the practice and test sessions were instructed to respond only with strategic feedback to substantial questions concerning probability calculation (e.g., “Can you remember how we solved this kind of problem during the lesson?” or “Maybe you remember that there was a trick to solve this kind of problem?”). In addition, we asked the

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

regular math teachers not to elaborate the topic before the last test had been realized. At the last test session, all teachers confirmed that they did not work on probability calculation during the study period.

All practice exercises and test tasks were prepared with the research tool formr (Arslan & Tata, 2017) and worked in class on tablets in a self-paced manner. Students who finished first were asked to wait quietly until the rest of the class had finished too. On times when the students did not have to work any exercises, they were provided brain-teasers as filler tasks.

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

## Results

The data sets and R scripts for data preparation and analyses are provided online (https://osf.io/d542q/?view_only=7896f90d809140d08b777fba6d564454).

The prior knowledge test, conducted before the lessons, confirmed that the participating sample had a rudimentary knowledge of probability calculation, at most, with only seven of 81 students gaining scores of 50% or more. The performance of students in the massed and distributed condition in the practice and test sets are displayed in Table 2.

Table 2

*Mean Performance Scores of the Students in Practice and Test Sets*

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

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

Bayesian regression models were used to test the hypotheses. Among others, a reason to choose Bayesian statistics over the frequentist approach was the relatively small sample size. While the binary results (significant or not significant) of the frequentist approach can make reliable conclusions complicated or even impossible when small sample sizes are involved, Bayesian statistics allows to assign each parameter a range of values together with the respective probability for the effect, even with small sample sizes (Kruschke, 2015). Each of the models included the practice condition (distributed vs. massed practice) to test the respective hypothesis and the score in the very first practice set as control variable. The control variable was included to ensure that the effect of distributed practice was not due to performance differences that existed prior to the experimental manipulation. The models were estimated in R (R Core Team, 2016) using the brms package[4] (Bürkner, 2017) and checked for autocorrelation and proper chain conversion. Because there are only few existing results on the effect of distributed practice on mathematical performance in school, no priors were specified. That is, for the two independent variables an improper flat distribution over the reals served as prior distribution (Bürkner, 2017).

To test the first hypothesis that distributing mathematical practice improves performance two weeks after the last exercises more than massing the same amount of practice, a Bayesian linear regression model with the performance in the first test as dependent variable was computed. Practice condition and initial practice performance served as independent variables. Contradicting the hypothesis, there was only little evidence for a positive effect of distributed practice: The mean of the posterior distribution for the effect of distributed practice was 0.31 (95% credible interval = -0.79 to 1.40). That is, the expected performance difference between the students of the distributed and massed practice condition is most likely between -0.79 and 1.4 (with the test score ranging from 0 to 9.5 points). The evidence ratio of 2.41 confirms that it is only about twice as likely that distributed practice has a positive effect than that it has no effect or a negative effect on the performance two weeks after the last practice set, compared to massed practice. Regarding Lee and Wagenmakers (2013) this is only “anecdotal evidence” for a positive effect. The mean for the effect of the performance in the first practice set was 0.58 (95% credible

[4] Further R-packages we used for data preparation and analysis were (in alphabetical order): BayesFactor (Morey & Rouder, 2015), gridExtra (Auguie, 2017), partykit (Hothorn & Zeileis, 2015), psych (Revelle, 2016), rstan (Stan Development Team, 2018) and tidyverse (Wickham, 2017).

interval = 0.37 to 0.78), that is, a higher score in the first practice set is related to a higher score in the first test.

The second hypothesis was that the positive effect of distributed practice would be more pronounced in the long run, that is, the difference between distributed and massed practicing students should be larger in the second test, conducted six weeks after the last practice session. To test this hypothesis, the performance change between the second and first test was calculated. According to the hypothesis, this change score should be higher (or less negative, in the case of a performance decrease) in the distributed practice group than in the massed practice group. Indeed, the mean of the posterior distribution for the effect of distributed practice was 0.63 (95% credible interval = -0.33 to 1.57) and the evidence ratio in favor of a positive effect of distributed practice on the change score was 10, which is considered strong evidence for a positive effect of distributed compared to massed practice (Lee & Wagenmakers, 2013). The performance in the first practice set, however, did not seem to influence the performance change between first and second test: The posterior distribution for the initial practice performance was -0.01 (95% credible interval = -0.19 to 0.16).

Because the first hypothesis could not be supported by the data, the performance difference between distributed and massed practicing students in the second test was additionally compared directly. For the second test, the mean of the posterior distribution of distributed practice was 0.93 (95% credible interval = -0.21 to 2.07) and the evidence ratio for the positive effect of distributed practice compared to massed practice on the long-term test performance was 18, which is considered strong evidence (Lee & Wagenmakers, 2013). In practical terms this means that the performance advantage for the seventh graders of the distributed practice condition compared to the seventh graders of the massed practice condition is most likely about one point (with a maximum test score of 9.5 points). Similar to the first test performance, the posterior distribution for the effect of the initial practice performance was 0.56 (95% credible interval = 0.35 to 0.78) confirming that students who scored higher in the first practice set on average can be expected to also achieve a higher long-term test performance.

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

**Exploratory analyses**

Conditional inference tree models, as proposed by Hothorn, Hornik and Zeileis (2006, 2015) were used to exploratory analyze the potentially moderating effects of mathematical self-efficacy and concentration difficulty on the effect of distributed practice. These models are based on recursive binary partitioning and can be used to exploratory detect relationships between a dependent and multiple independent variables which can be measured on different scales. Basically, the model first checks globally whether the null hypothesis (i.e., that the dependent variable is independent from all tested independent variables) can be rejected. If this is the case, the independent variable with the strongest association to the dependent variable is selected. Based on the selected independent variable, the observations are split into two groups with the criterion to minimize the *p*-value. The results are two different subsamples – distinguished based on the selected independent variable – that show maximally different distributions of the dependent variable. This process is repeated for each resulting subsample until the null hypothesis of independence cannot be rejected any longer.

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

Two additional conditional inference tree models were calculated (again, one for each test time) to investigate whether the effect of distributed practice depends on the initial performance level, as results by Hirsch et al. (1982) and Saxon (1982) suggested. In these models, practice condition and performance in the first practice set served as independent variables and the respective test performance as dependent variable. These analyses revealed that in the second test, conducted six weeks after the last practice session, distributed practicing students performed better than massed practicing students in particular if their performance in the first practice set was in a medium score range (i.e., between 3 and 7 points out of 9.5 points, s. Figure 2). In the low or high score ranges, the test performance was rather independent from the practice condition, indicating that the effect that was supported by the Bayesian regression model above may be mainly due to the students in a medium performance range. This interaction did not appear for the first test.

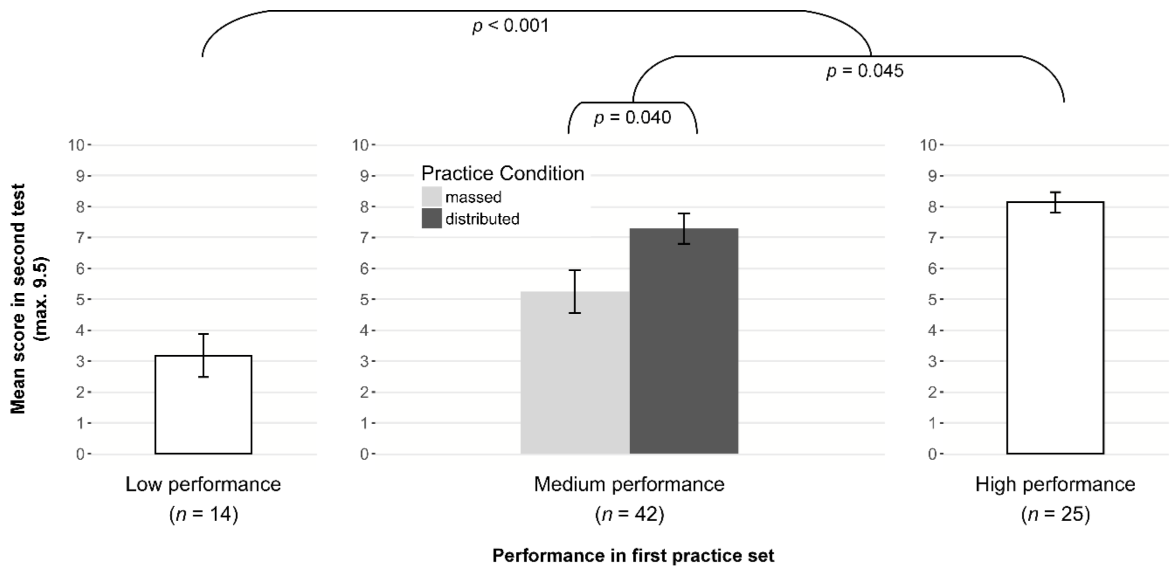



*Figure 3*. Interaction between students' performance in the first practice phase and learning condition concerning their long-term test performance (i.e., six weeks after the last practice session).

## Discussion

Despite the strong empirical indicators that suggest a positive effect of distributed practice on the long-term memory of verbal content (e.g., Cepeda et al., 2006), little is known about the effect of distributed practice on the learning of mathematics. Though few studies found positive effects of distributed practice using mathematical content (e.g., Rohrer & Taylor, 2006, 2007; Schutte et al., 2015; Barzagar Nazari & Ebersbach, 2018), the question if and how distributed practice affects performance in more complex domains like procedural mathematical learning still lacks sufficient empirical support.

In the current study, the effect of distributed practice with feedback on mathematical learning was examined in seventh grade. Contrary to the central hypotheses and the results by Barzagar Nazari and Ebersbach (2018), there was no evidence for a positive effect of distributed practice compared to massed practice on the intermediate performance of the students, tested two weeks after practice. However, regarding the long-term performance tested six weeks after the last practice session, strong evidence for a positive effect of distributed practice was revealed. The latter finding is similar to the one reported by Barzagar Nazari and Ebersbach (2018). Though it is theoretically assumed that the effect of distributed practice emerges especially in the long term, it is not clear why there was no empirical evidence for a positive effect two weeks after the last practice set in the present study. Most of the previous studies only investigated the effect of distributed with similar retention intervals of up to one or two weeks and still found positive effects of distributed practice. Perhaps, feedback boosted the performance in both conditions so that the performance was still similar after two weeks, overshadowing the effect of practice condition that became apparent only in the long term. Future studies should try to implement distributed practice of complex material with feedback with a greater variance of retention intervals to further investigate its effect.

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

These results, though exploratory, are interesting because they suggest that regarding mathematical learning there might be subgroups of learners whose performance is rather unaffected by the distributed or massed practice strategy, possibly because they are either not able to grasp the underlying concept at all (i.e., the low performance group) or because they grasp it that quickly that the practice strategy does not matter, too (i.e., the high-performance group). In each case, the performance is relatively stable and independent from practice, and hence independent from the practice strategy. It is only the students in the medium performance range who might benefit from distributed practice. Interestingly, the aforementioned studies examining the effect of mixed review of mathematical homework by Hirsch et al. (1982) and Saxon (1982) point into a similar direction since the authors found that low to medium performing students profited more from distributed (and interleaved) practice than high performing students. More studies on distributed practice of mathematical content should be conducted to further investigate the relationship between ability and the effect of the practice strategy.

Another issue raised in the current study is whether distributed practice is in fact a difficulty for learners, as assumed for instance by Bjork (1994). Descriptively, there seemed to be no differences concerning the practice performance of massed and distributed practicing students in the present study (see Table 2). Additionally, Bayesian t-tests did not indicate that the perceived difficulty of the practice and test sets differed between the conditions at all. However, indicators that distributed practice is regarded as effortful strategy by learners is reflected by the fact that the majority of students does not think that it is an effective study strategy (McCabe, 2011) and therefore tends to resort to massed practice and devote their time for learning on the two days prior to the test (Hartwig & Dunlosky, 2011; Taraban, Maki, & Rynearson, 1999). Thus, students show hardly any insights into the actual benefit of distributed practice and feel more confident with massed practice (Bjork et al., 2013; Kornell, 2009). Studies like the present one may serve to undeceive learners and teachers about the effectivity of distributed practice for learning in school.

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

Additionally, as in the previous study by Barzagar Nazari and Ebersbach (2018), only one specific topic was covered in the current study and generalizations concerning other topics based on this study should be drawn with caution. The topic stochastics was picked for different reasons. First, the topic is part of the curriculum for Grade 7, but regularly covered at the end of the school year or not at all. That is, with this topic the chance was high to find enough courses who had not already covered the topic at the beginning of the study period. Second, stochastics as a mathematical topic is relatively independent from other domains such as analysis or geometry. That is, the courses were roughly comparable in this specific topic as their performance was hardly influenced by the topics covered in the regular mathematics class before the start of the study. Third, the goal was to expand the results of the previous study on the same topic, so roughly the same material was used. Nevertheless, research on distributed practice with mathematical content would benefit from more studies using a broader range of material.

**Conclusions**

The current study contributes to answer the question why and under which circumstances distributed practice proves a useful learning strategy in realistic learning contexts, even beyond learning of rather simple verbal content. We argue that while the effect of distributed practice emerges for roughly all learners if simple verbal content or mathematical routines are practiced (Carpenter et al., 2012; Rohrer & Taylor, 2006, 2007), the effect of distributed practice on learning of more complex content, like the acquisition and application of more advanced mathematical procedures as in the present study, might especially be effective for students in the medium performance range (see also Hirsch et al., 1982; Saxon, 1982). These assumptions, however, are based on exploratory analyses and need further empirical confirmation. In sum, distributed practice remains a promising learning strategy and more studies on a broad range of content and learners could help to deepen our understanding of when and why it works.

**Acknowledgments**

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

## Appendix A. Exercise examples

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

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

**Exercise 1**

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

Answer: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

**Exercise 2**

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

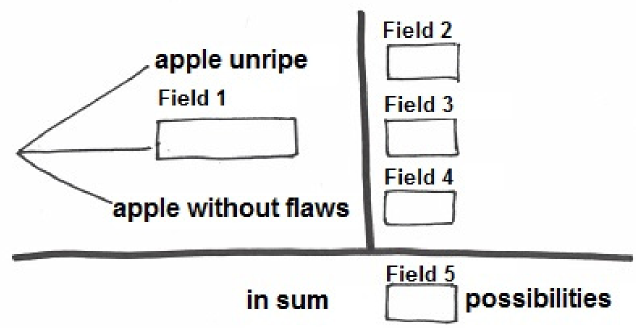
a) …draw one of the four jacks? Answer: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

b) …draw one of the two red queens? Answer: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

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

**Exercise 3**

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




Field 1: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

Field 2: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

Field 3: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

Field 4: ___________________________

Field 5: ___________________________

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

## Appendix B. Exclusion procedure

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

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

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

**References**

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. doi: 10.3102/0034654316689306

Arslan, R. C., & Tata, C. S. (2017). formr.org survey software (Version 0.16.13).

Auguie, B. (2017). gridExtra: Miscellaneous functions for “grid” graphics (R package version 2.3). Retrieved from https://CRAN.R-project.org/package=gridExtra

Barzagar Nazari, K., & Ebersbach, M. (2018). *Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom*. Manuscript submitted for publication.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35(2), 201–210. doi: 10.3758/BF03193441

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., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64(1), 417–444. doi: 10.1146/annurev-psych-113011-143823

Boerner, S., Seeber, G., Keller, H., & Beinborn, P. (2005). Lernstrategien und Lernerfolg im Studium: Zur Validierung des LIST bei berufstätigen Studierenden. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37(1), 17–26. doi: 10.1026/0049-8637.37.1.17

Brunmair, M., & Richter, T. (2018). *Similarity matters: A meta-analysis of interleaved learning and its moderators*. Manuscript submitted for publication.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. doi: 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. doi: 10.1007/s10648-012-9205-z

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. doi: 10.1037/0033-2909.132.3.354

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*, 1–19. doi: 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. doi: 10.1016/S0022-5371(72)80001-X

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Oxford, England: Irvington.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum.

Delaney, P. F., Godbole, N. R., Holden, L. R., & Chang, Y. (2018). Working memory capacity and the spacing effect in cued recall. *Memory*, 26(6), 784–797. doi: 10.1080/09658211.2017.1408841

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, 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* (Vol. 53, pp. 63–147). London: Academic Press.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, 43(8), 627–634. doi: 10.1037/0003-066X.43.8.627

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, 84(5), 795–805. doi: 10.1037/0021-9010.84.5.795

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7(2), 95–112. doi: 10.3758/BF03197590

Hartwig, M. K., & Dunlosky, J. (2011). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, 19(1), 126–134. doi: 10.3758/s13423-011-0181-y

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

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.

Hirsch, C. R., Kapoor, S. F., & Laing, R. A. (1982). Alternative models for mathematics assignments. *International Journal of Mathematical Education in Science and Technology*, 13(3), 243–252. doi: 10.1080/0020739820130301

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674. doi: 10.1198/106186006X133933

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 partytioning in R. *Journal of Machine Learning Research*, 16, 3905–3909.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research*, 30(1), 138–149. doi: 10.1086/374692

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

Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, 3, 12–19. doi: 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. doi: 10.1016/j.learninstruc.2014.11.001

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. doi: 10.1126/science.1152408

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, 23(9), 1297–1317. doi: 10.1002/acp.1537

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. doi: 10.1027/2151-2604/a000168

Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition*, 42(5), 729–741. doi: 10.3758/s13421-014-0394-1

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, 39(3), 462–476. doi: 10.3758/s13421-010-0035-2

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–198). New York: Psychology Press.

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language*, 27(5), 521–536. doi: 10.1016/0749-596X(88)90023-X

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (R package version 0.9.12-2). Retrieved from https://CRAN.R-project.org/package=BayesFactor

Project Team / Schmitz, C. (2012). LimeSurvey: An open source survey tool. Hamburg: LimeSurvey Project. Retrieved from http://www.limesurvey.org

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning: Journal of Practical Research & Applications*, 4(1), 11–18.

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

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, 40(1), 4–17.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, 24(3), 355–367. doi: 10.1007/s10648-012-9201-3

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. doi: 10.1002/acp.1266

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6), 481–498. doi: 10.1007/s11251-007-9015-8

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. doi: 10.1037/a0037559

Saxon, J. (1982). Incremental development: A breakthrough in mathematics. *The Phi Delta Kappan*, 63(7), 482–484.

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. doi: 10.1016/j.jsp.2014.12.003

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology*, 19(1), 107–122. doi: 10.1002/acp.1066

Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *A series of books in psychology. Learning and individual differences: Advances in theory and research* (pp. 13–59). New York: W H Freeman/Times Books/ Henry Holt & Co.

Stan Development Team. (2018). RStan: The R interface to Stan (R package version 2.17.3). Retrieved from http://mc-stan.org/

Taraban, R., Maki, W. S., & Rynearson, K. (1999). Measuring study time distributions: Implications for designing computer-based courses. *Behavior Research Methods, Instruments, & Computers*, 31(2), 263–269. doi: 10.3758/BF03207718

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. doi: 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* (Vol. 60, pp. 113–189). Academic Press.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 796–800.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2005). Limitations to the spacing effect: Demonstration of an inverted u-shaped relationship between interrepetition spacing and free recall. *Experimental Psychology*, 52, 257–263.

Wickham, H. (2017). tidyverse: Easily install and load "tidyverse" packages (R package version 1.1.1). Retrieved from https://CRAN.R-project.org/package=tidyverse

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

Yazdani, M. A., & Zebrowski, E. (2006). Spaced reinforcement: An effective approach to enhance the achievement in plane geometry. *Journal of Mathematical Sciences and Mathematics Education*, 1, 37–43.

Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(6), 545–559.

Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202–231). New York: Cambridge University Press.

# CHAPTER IV: Study 3

## *Distributed practice:*
## *Rarely realized in self-regulated mathematical learning*

Katharina Barzagar Nazari & Mirjam Ebersbach

A version of this chapter is published as:

Barzagar Nazari, K., & Ebersbach, M. (2018). Distributed practice: Rarely realized in self-regulated mathematical learning. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.02170

**Abstract**

With distributed practice, a fixed learning duration is spread over several sessions, whereas with massed practice, the same time is spent learning in one uninterrupted session. Numerous studies have proven distributed practice to be an effective tool for improving long-term retention of verbal material. For simple procedural knowledge in mathematics, distributed practice seems to be a promising learning strategy as well, at least when the practice schedule is externally guided. The purpose of the present study was to investigate distributed mathematical practice in the context of self-regulated learning in high school. In total, 158 students were invited to participate. In a preliminary phase, motivational and cognitive characteristics of the students were assessed. Afterwards, the students were introduced to basic statistics, a topic of their regular curriculum. At the end of the introduction, the students could sign up for the study to further practice this content. Eighty-seven students did so and were randomly assigned either to the distributed or to the massed practice condition. In the distributed practice condition, they received three practice sets on three different days. In the massed practice condition, they received the same three sets, but all on one day. All exercises were worked in the context of self-regulated learning at home. Performance was tested two weeks after the last practice set. Only 44 students finished the study, which hampered the analysis of the effect of distributed practice. To examine the individual characteristics of the students who completed the exercises in the context of their self-regulated learning, exploratory analyses were conducted. The proportion of students who finished all exercises was significantly higher in the massed than in the distributed practice condition, suggesting particular difficulties in the latter condition. Within the distributed practice condition, a significantly larger proportion of female students completed the exercises compared to male students. Additionally, among these female students, a larger proportion showed lower concentration difficulty. No such differential effects were revealed in the massed practice condition, indicating that when implementing distributed practice in self-regulated learning, individual learner characteristics are particularly important.

## Introduction

Teachers at school and university should generally be interested in learning techniques that promote long-term retention of the learned contents. The reason for this is that most topics taught in school or at university are rather complex and many advanced topics rely – sometimes more, sometimes less so – on prior knowledge. In mathematics, for example, in order to grasp stochastics, solid knowledge of fractional arithmetic is helpful. That is why in most cases it is important to use learning strategies that help to store knowledge in a way that facilitates long-term retention, rather than learning strategies that result in knowledge that can be retrieved only for a short period after it was taught and acquired.

One branch of learning strategies that promote long-term retention is based on so-called *desirable difficulties* (Bjork, 1994). Desirable difficulties are mechanisms that slow down the learning process and make it harder for the learner, but boost long-term performance. Learning strategies that make use of these desirable difficulties are, among others, the *generation* of (part of) the content that has to be learned (McDaniel, Waddill, & Einstein, 1988), (self-)*testing* as part of the learning process (Karpicke & Roediger III, 2008), the *interleaving* of similar but not identical topics during the learning process (Rohrer, 2012), and the spacing of a given learning time across more than one learning session, which is also called *distributed practice* (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006).

Most of these learning strategies have been studied extensively in the laboratory and are known to result in robust and strong performance improvements as compared to the respective control conditions (Adesope, Trevisan, & Sundararajan, 2017; Bertsch, Pesta, Wiscott, & McDaniel, 2007; Brunmair & Richter, 2018; Rohrer, 2012; Rowland, 2014). However, less is known about the effects of these strategies in real learning contexts because field studies are relatively rare so far. Moreover, it is still an open question to which degree learning strategies related to desirable difficulties are used in the context of self-regulated learning and which learner characteristics are associated with their use. The present study addresses this question by examining the effect of distributed practice as a learning strategy in a real learning context involving self-regulated mathematical learning of high school students.

### Distributed practice in mathematical learning and in school

Distributed practice means that a given learning or practice duration is distributed across more than one learning session, whereas in massed learning, the same time is spent in one learning session only (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012). The effect of distributed practice has been explored for decades in numerous studies, which suggest robust and large effects on long-term performance in particular, as compared to massed practice (Cepeda et al., 2006). The effect of distributed practice has been demonstrated under a great range of circumstances: with different materials (e.g., Carpenter et al., 2012), in different age groups (e.g., Toppino, Kasserman, & Mracek, 1991), and using different lags between the learning sessions (e.g., Cepeda et al., 2009).

However, there are still contexts and conditions that have been considered less so far. These include the effect of distributed practice on mathematical learning in general and, in particular, mathematical learning in school. An exception are the studies by Rohrer and Taylor (2006, 2007) on distributed practice of mathematical procedures in college. The students in these studies practiced permutation problems either massed in one session or distributed across two sessions with a lag of seven days. The number of problems that were practiced and the total practice time were the same in each condition, yet students of the distributed practice condition outperformed students of the massed practice condition one week after practice (Rohrer & Taylor, 2007) as well as four weeks after practice (Rohrer & Taylor, 2006). Although in these studies the effect of distributed practice was investigated with material other than wordlists – which were often used in the classical studies (e.g., Cepeda et al., 2006) – the learning material was still rather narrow in content (i.e., the application of only one specific formula was practiced) and not very complex (i.e., the correct procedure was learned by heart and not explained to the participants).

Another branch of studies investigated a combination of interleaved and distributed practice, termed “mixed review”, in school (Saxon, 1982; Yazdani & Zebrowski, 2006) and in college (Hirsch, Kapoor, & Laing, 1982). This means that in every practice situation, all topics that have previously been taught are practiced and not only the topic that was just covered. Thereby, practice is distributed across several sessions, and different topics are covered in an interleaved (but accumulating) manner. Mixing exercises on current and previous topics improved performance more than only reviewing the most recent topic in each session (for a review on “mixed review” in mathematical learning, see Rohrer, 2009). In addition, Hirsch et al. (1982) and Saxon (1982) found that the positive effect on mathematics performance was particularly large for students in lower to medium performance ranges.

Only few studies have investigated the isolated effect of distributed practice on mathematics learning in school: Schutte et al. (2015) asked third graders to practice basic addition problems four minutes each day for nearly three weeks. Students who distributed their four minutes practice time across the day (1-minute practice session four times a day or 2-minutes practice sessions two times a day) outperformed students who practiced massed each day for four minutes. However, because all students practiced each day for nearly three weeks, even those of the "massed practice" condition where technically practicing in a distributed manner. In another study by Chen, Castro-Alonso, Paas, & Sweller (2017), students of Grades 4 and 5 practiced different mathematical topics of their regular curriculum either massed in one session or distributed across three sessions on three consecutive days. The results indicated that distributing practice across three days significantly improved performance as compared to one massed practice session. However, generalizations based on these results are limited because the lag between the last practice session and the test was not the same for distributed and massed practicing students. Furthermore, in a study by Barzagar Nazari and Ebersbach (2018a), strong evidence for a positive effect of distributed practice one week after the last practice session was found for third graders who practiced semi-formal multiplication in three sessions distributed across three consecutive days, as compared to students who worked the same number of practice exercises massed in one session. Five weeks later, however, this effect disappeared. The reason for this vanishing effect might have been that between the test sessions the students practiced content that was related to the study topic in their regular classes, thereby obliterating differences between the practice conditions. In a similar study conducted in Grade 7 with students practicing stochastics, a topic that is more dissociated from other mathematical topics, evidence for a strong effect of practice condition on the final performance was revealed, both one and six weeks after the last practice session (Barzagar Nazari & Ebersbach, 2018a). In another study comparing the effects of distributed and massed practice of stochastics exercises that was conducted in Grade 7, no evidence for an effect of practice condition was found in the first test two weeks after the last practice. However, a later test conducted six weeks after the last practice again revealed strong evidence for a positive effect of distributed practice (Barzagar Nazari & Ebersbach, 2018b). An additional, exploratory result of this second study in Grade 7 indicated that the positive effect of distributed practice occurs especially for students in the medium performance range (similar to the results of Hirsch et al., 1982, and Saxon, 1982).

In sum, although only few studies have investigated the isolated effect of distributed practice on mathematics learning in school, most findings suggest a positive impact. However, in previous studies, sessions including distributed practice were highly structured by the experimenters or teachers and took place solely in the classroom. In real-world learning settings, especially older students at school and students at university have to do a lot of learning and practice outside the classroom in a more self-regulated manner. The aim of the present study was to investigate whether distributed practice can also be implemented using online exercises in order to improve mathematical performance in a real-world learning context, in which practice relies more heavily on self-regulation. In addition, we examined which learners actually followed the distributed practice schedule.

**Differential effects of distributed practice**

One question that has hardly been investigated in studies examining the effects of desirable difficulties in general, and of distributed practice in particular, is whether all learners profit from such learning strategies in a similar way or whether individual differences moderate the effectivity of distributed practice (Delaney, Verkoeijen, & Spirgel, 2010). However, this is a central question in order to decide whether distributed practice can be recommended in general for educational contexts or only for particular contexts or learners. Previous studies described earlier showed larger effects of distributed practice for students with low or medium prior knowledge (Hirsch et al., 1982; Saxon, 1982; Barzagar Nazari & Ebersbach, 2018b). In the two previous studies by Barzagar Nazari and Ebersbach (2018a, 2018b), several other motivational (e.g., mathematical self-efficacy) and cognitive learner characteristics (e.g., concentration difficulty) were also considered as potential moderators of the effect of distributed practice. Except for the abovementioned baseline performance, no interactions with the effect of distributed practice on test performance were found. However, the sample sizes in these studies might have been too small to reveal moderator effects.

As noted previously, most of the aforementioned studies employed a highly teacher-guided practice procedure. In a more self-regulated learning scenario, the question of whether and how individual learner characteristics affect the use and effectivity of distributed practice actually becomes even more important: Individual motivational and cognitive characteristics could not only affect the effect of distributed practice on final test performance, but also determine if and how the students follow the respective practice schedule. Distributed practice requires learners to repeatedly engage with a topic or procedure, which may be difficult to retrieve given the temporal delay between learning sessions. In fact, this is why distributed practice is related to desirable difficulties: It is assumed that the lags between practice sessions make the learning process more difficult, which in turn should improve long-term retention. Learners with low mathematical self-efficacy, however, could suffer from this additional difficulty and decide to stop to engage mentally or in practice with the topic (Zimmerman, 1995). That is, the effect of distributed practice and/or the amount of practice may be smaller for learners with low mathematical self-efficacy. A similar reasoning can be applied to performance avoidance goals (Dalbert & Radant, 2008; Elliot, 1999), because students who have problems coping with their mistakes may have relatively more problems with distributed practice, given that the distributed practice schedule initially increases the number of mistakes. Concerning work avoidance (Nicholls, Cobb, Wood, Yackel, & Patashnick, 1990) different scenarios are possible: On the one hand, massed practice requires students to work for a longer duration at a time, which might be disfavored by students with high work avoidance. On the other hand, in distributed practice students have to repeatedly bring themselves to start working, which might also be hard for students high in work avoidance. That is, work avoidance possibly could influence the effect of distributed practice and/or adherence to the practice schedule in different ways. Students with low concentration ability, however, might particularly benefit from distributed practice, as the distributed sessions are shorter than one massed session and hence requires the students to concentrate for a shorter duration at a time. Because there is currently only little prior research on the effects of (mathematical) self-efficacy, performance avoidance goals, work avoidance, and concentration difficulty on the efficacy of distributed practice, the interactions of these four characteristics with the distributed practice condition will be investigated in exploratory analyses, with a focus on the students' adherence to the practice schedule within their self-regulated learning[1].

**Research question and hypotheses**

The objectives of the present study were to investigate distributed practice in a real learning context including a relatively high degree of self-regulated learning. The main questions were whether distributed practice is used reliably by learners, and which learner characteristics promote (or hinder) its use. The sample consisted of high school students, and the material was relevant for their mathematics curriculum. We expected that distributed practice might not consistently be applied by the students in the context of their self-regulated learning (see also Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013

[1] Originally, another main purpose of the present study was to investigate the effect of the practice condition on the test performance and possible interactions with the mentioned characteristics. However, as the study suffered a severe dropout, the final sample was too small to examine interaction effects on the test performance. Therefore, the exploratory analyses are limited to the dependent variables of participation and adherence to the practice schedule (see Results).

for similar findings for adults). In addition, individual learner characteristics that might have affected the implementation of distributed practice were analyzed exploratory.

## Method

### Participants

In total, 158 students of eight courses from Grades 10 and 11 (first year of senior classes), attending three schools, were requested to participate in the current study[2]. These students were enrolled either in regular math courses or in intensive math courses, depending on their own choice. All schools were located around a medium-sized German city in neighborhoods with inhabitants of a medium socio-economic status. Participation was voluntary and could be terminated at any time. Only students who had written consent from their parents could participate. They were told that they would receive 10 Euro if they completed the study. Signing up for the study required providing an e-mail address, because the experimental part of the study took place online. Of the 87 students who signed up (40 female, 47 male; 58 of regular math courses, 29 of intensive math courses; $M_{age}$ = 16 years 5 months, age range: 15-17 years), 43 terminated their participation ahead of time, and only 44 students finished it completely (25 female, 19 male; $M_{age}$ = 16 years 6 months, age range: 15-17 years).

### Design

The independent variable was practice condition with two between-subjects levels: One group of students worked the exercises massed in one session and the other group worked the same exercises distributed across three sessions. Both conditions worked a total number of twelve practice exercises (three sets à four exercises). An expanding interval schedule was used for the practice sessions of the students in the distributed practice condition (see Procedure; Küpper-Tetzel, Kapler, & Wiseheart, 2014). The dependent variable was test performance, assessed two weeks after completing the last practice exercise.

Of the 87 students who initially signed up, 49 were assigned to the distributed practice condition (25 female, 24 male; $M_{age}$ = 16 years 5 months, age range: 15-17 years) and 38 were assigned to the massed practice condition (15 female, 23 male; $M_{age}$ = 16 years 6 months, age range: 15-17 years). A slightly larger proportion of students was assigned to distributed practice as we expected a larger dropout in this condition. To ensure

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

that the overall math performance level was roughly equal in both practice condition groups before the manipulation, students who signed up for the study were ranked by their most recent mathematics grade and then, within each class and grade level, randomly assigned to one of the two practice conditions. In order to minimize potential effects of class, the ratio of massed and distributed practicing students was similar in each class. As mentioned above, in total only 44 students completed the study (i.e., 17 distributed practicing students with a median math grade of 2.0 and 27 massed practicing students with a median math grade of 2.3[3]).

Additionally, a questionnaire assessing some motivational and cognitive characteristics of the students (see Table 1 for information on the scales and their reliability based on our data) was included. Finally, each time after students had finished an exercise, they were asked to rate how difficult they considered the respective exercise. These questions were included to examine if students of the distributed practice condition did in fact perceive the exercises to be more difficult than students of the massed practice condition and, thereby, if distributed practice proved to be a difficulty as perceived by the learners.

**Material**

In the course of the experiment, the students were introduced to basic statistics by student assistants with teaching experience, who were supervised by the authors. More specifically, the students were taught the definition of variables and their manifestations, the law of large numbers, the sum rule and the interpretation and creation of diagrams. The topic of basic statistics is generally part of the following school year, thus, no class had covered the topic in their current school year prior to the study. The lessons and practice material were prepared with the support of didactics experts with teaching experience in order to make the learning environment as realistic as possible. The complete material (lesson scripts, practice and test sets, and the scoring scheme) is provided in German online (https://osf.io/egt4j/?view_only=5fb49869de0345618a3c9171c77992fe).

Each practice set for the students consisted of four exercises and involved calculating absolute from relative frequencies, using the sum rule for probability calculation, naming variables and values, and preparing calculations for a diagram. The practice sets contained conceptually similar but not identical exercises, that is, solutions

[3] In Germany, grades range from 1 (very good) to 6 (inadequate).

could not be learned by heart. Each practice set could easily be finished in less than 30 minutes. An example of a practice set can be found in the Appendix.

Table 1

*Instruments Used to Assess Potential Moderators*

| Motivational characteristics | Employed instrument | Reliability |
|---|---|---|
| Mathematical self-efficacy | 7 items of a German *Academic Self-Efficacy Scale for School Children* (Jerusalem & Satow, 1999), adapted to mathematics<br><br>(Sample item: "In math, I can solve even the difficult problems if I try hard.") | α = 0.88 |
| Performance avoidance goals | 8-item German *SELLMO* (Spinath, Stiensmeier-Pelster, Schöne, & Dickhäuser, 2012), adapted to mathematics<br><br>(Sample item: "In math, my main concern is to avoid that the other students think that I am stupid.") | α = 0.86 |
| Work avoidance | 8-item German *SELLMO* (Spinath et al., 2012), adapted to mathematics<br><br>(Sample item: "In math, my main concern is not to have any difficult tests or work.") | α = 0.85 |
| Cognitive characteristic | | |
| Concentration | 6 items of the German *Learning Strategies in College – LIST* (Boerner, Seeber, Keller, & Beinborn, 2005; Wild & Schiefele, 1994)<br><br>(Sample item: "When I'm studying, I'm easily distracted.") | α = 0.92 |

**Procedure**

The students were asked to work through all of the practice and test exercises that followed the lecture at home; only the questionnaire and the lecture sessions at the beginning of the study were completed at school. Practicing at home resembles real-world learning settings because students usually have to do their homework outside the classroom in a self-regulated manner. In order to avoid students being particularly prepared or relying on help for the test at home, the test was announced as “further exercises”.

**Prior testing and introductory lesson.** Prior to the experimental manipulation, the study started with a survey in school assessing students’ mathematical self-efficacy, self-rated difficulty to concentrate, performance avoidance goals and work avoidance. The questionnaire was programmed with LimeSurvey (Project Team / Schmitz, 2012) and answered individually by the students on tablets that were provided by the survey team. After the students had finished the questionnaire, they worked on a pretest on basic statistics and probability calculation, assessing whether the prior knowledge of students with regard to the study topic was comparable in all conditions. The survey and pretest were followed by three 45 minutes regular lecture sessions, in which the students were introduced to the topics specified above. The lecture sessions were spread over two or three days within one week, depending on the schedule of the respective class. At the end of the last lecture session, the students were told that, with their parents’ consent, they could voluntarily participate in an online study on the lecture topic if they provided us with an e-mail address, and that they would receive 10 Euro for the completion of the study.

**Practice and testing.** Between five and seven days after the last lesson, the students received their first practice exercises, provided via a personalized link that was sent by e-mail. The students were not allowed to keep the lesson material, that is, the material was not available for the practice exercises. However, after a practice exercise was completed, the correct solution was displayed on the screen. The test set was similar to the practice sets, but no correct solutions were provided after test exercises. It was not possible to go back to previous pages at any time. The practice and test sets were created and distributed with the research tool formr (Arslan & Tata, 2017). In the massed practice condition, the students received three practice sets on the first practice day. In the distributed practice condition, the students received the same three practice sets, but only one practice set on the first day of practice, the second practice set two days later, and the third practice set another five days later (i.e., expanding interval schedule). After the students had received each link, they had one and a half days to finish the exercises provided via that link. This relatively long period resembles classical homework settings and was provided to ensure that the students had enough time to actually work the exercises. However, checking the time spent on the exercises revealed that most students completed the exercises within one day. Only few students opened the link on one day and finished the exercises the next day, and even among these students, some may only have clicked on the link without actually starting with the exercises on the first day. That is, the probability that students of the massed practice condition distributed their exercises across one and a half days instead of completing them in one day is negligibly small. Moreover, this interval would still have been much shorter than the intervals between the practice sets in the distributed practice condition. Retention performance was tested two weeks after the last practice set was completed with exercises that were similar to the practice exercises.

**Scoring.** Each given answer (mostly numbers or single words) was either correct or wrong, that is, no partial points were granted. For each practice and test set, the maximum score was 15 points. Two raters scored the answers independently from each other according to a predefined scheme. Afterwards, the scores of both raters were compared and differences were discussed and most often resolved by these raters. To ensure that their reliability was not due to the discussion, a third rater rated the answers independently as well. The final ratings of the first two raters were nearly identical to the third (control) rater (more than 98% identical scores). Therefore, the combined final scores of the two first raters were analyzed.

## Results

The data as well as the processing and analysis scripts are provided online (https://osf.io/egt4j/?view_only=5fb49869de0345618a3c9171c77992fe). Because of the severe dropout in the course of the study (of 158 eligible students, only 44 finished the study), analyses concerning the effect of practice condition on retention performance turned out to be rather inappropriate: First, the remaining groups were rather small and not of equal size, and second, there seemed to be a selection bias concerning the dropouts, because the rate of completion was much higher in the massed practice condition (71%) than in the distributed practice condition (35%). We nevertheless report the analysis concerning the effect of practice condition for the sake of completeness, keeping in mind these limitations and that the results should be interpreted with caution. We used a Bayesian linear regression model to analyze the test performance, among other reasons because of the particularly small resulting sample size. One advantage of Bayesian modelling is that it provides a range of possible values for each estimated parameter and assigns probabilities to them, which facilitates interpretation especially when the results are not conclusive in classical statistical modelling (Kruschke, 2015). The linear regression model was estimated in R (R Core Team, 2016) using the package[4] brms (Bürkner, 2017). Test performance served as dependent variable and practice condition (distributed vs. massed practice) and performance in the first practice set (sum score) as independent variables. No priors were specified, that is, an improper flat distribution over the reals was used as prior distribution, which means that the results were highly data-driven and hardly influenced by the priors (Bürkner, 2017). The model was checked for proper chain conversion and autocorrelation, indicating no problems in this regard. The mean for the posterior distribution for the effect of distributed was about -1. That is, the students of the distributed practice condition were estimated to have a performance about 1 point (out of 15) lower than students of the massed practice condition (95% credible interval = -3.1 to 1.1). The evidence ratio of 0.13 confirms that a negative effect of distributed practice – contrary to our hypothesis – is more likely than no effect or a positive effect (which would be indicated by an evidence ratio of 1 or higher). According to Lee and Wagenmakers (2013), this is moderate evidence for a negative effect of distributed practice – but, again, these results have to be considered with caution.

As the main aim of the study was to investigate whether distributed practice works in self-regulated learning, subsequently exploratory analyses were conducted to examine which students completed the exercises in the context of their self-regulated learning. These analyses address two other questions that are important when implementing distributed practice in school, besides its general effect: (a) Which students are in general willing to invest additional effort into their mathematics learning by signing up for such a study, and, more specifically, (b) which students actually complete the distributed practice condition? On that account, conditional inference tree models were calculated (Hothorn, Hornik, & Zeileis, 2015). These models can be assigned to exploratory data mining, which has been frequently used in social and behavioral science (e.g. Salis, Kliem, & O'Leary, 2014; for an overview, see McArdle & Ritschard, 2013), and are useful for exploratory data analyses when there are no specific expectations regarding the relationship between a dependent variable and one or more independent variables. In addition, previous problems

[4] Further R-packages we used for data preparation and analysis were (in alphabetical order): BayesFactor (Morey & Rouder, 2015), partykit (Hothorn & Zeileis, 2015), psych (Revelle, 2016), rstan (Stan Development Team, 2018), and tidyverse (Wickham, 2017).

of overfitting and biases on the variable selection have been overcome in the current models (Hothorn, Hornik, & Zeileis, 2006; Strobl, Malley, & Tutz, 2009).

Conditional inference tree models seek to identify independent variables that can be used to split up the respective sample into groups that are maximally different with regard to the dependent variable (e.g., test performance score or participation in the study). This is accomplished by recursive binary partitioning: First, the model checks if the distribution of the dependent variable is unrelated to all independent variables. If this null hypothesis can be rejected, the model selects the independent variable that has the strongest relationship with the dependent variable (e.g., work avoidance). The sample then is split into two groups, based on the selected independent variable, in a way that minimizes the $p$-value. The resulting two groups (in the example, these could be participants with average or above-average work avoidance and participants with below-average work avoidance) show maximally different distributions of the dependent variable (i.e., low and high test performance). This process is then reiterated for each of the resulting subgroups until the null hypothesis in the first step cannot be rejected any longer.

Because participation in the current study was voluntary, a first conditional inference tree model was performed with the *enrollment in the study* as dependent variable (two levels: enrolled versus not enrolled in the study) in order to investigate which students were willing at all to enhance their math performance in the context of this study. As independent variables, the most recent math grade, the level of the attended math course (two levels: intensive math course versus regular math course), gender, and the cognitive and motivational characteristics listed above were included. The resulting conditional inference tree revealed that the sample could be divided into two groups, depending on whether their initial math grade was equal to/above average or below average (the median of the math grades was 2.7 on a scale from 1: very good, to 6: inadequate). A significantly larger proportion of students with better math grades signed up for the study, compared to students with lower math grades ($p < 0.001$, see also Figure 1). None of the other cognitive and motivational variables predicted students' enrollment in the study.

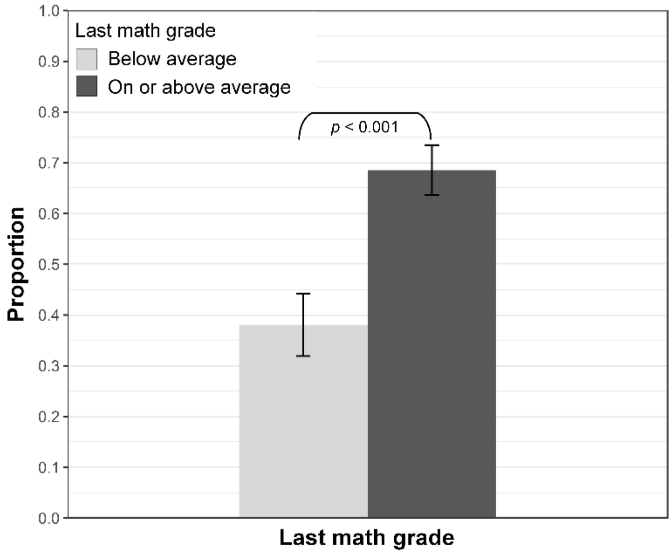



*Figure 1*. Proportion of students who enrolled in the study, separated by the level of their last math grade.

A second conditional inference tree model was performed with the *completion of the study* as dependent variable (two levels: study not completed versus study completed) because much less students completed the study than initially enrolled in the study. The independent variables were the same as in the first model, with two additional variables: condition (two levels, massed practice versus distributed practice) and performance in the first practice set (baseline performance, sum score). Both of these variables were irrelevant at the time of enrollment, because at that point the students were not yet assigned to a practice condition and had not yet completed a practice set (that is, these variables only apply to those who enrolled in the study). For this reason, these two variables were only included in the second model. The resulting conditional inference tree confirmed that the proportion of students who completed all practice sets and the test was significantly larger in the massed practice condition than in the distributed practice condition ($p = 0.007$). Additionally, within the distributed practice condition, a significantly larger proportion of female students completed the exercises compared to male students ($p = 0.001$). For the female students within the distributed practice condition, concentration difficulty had a significant impact on the completion of the study, with a significantly larger proportion of female students with lower concentration difficulty having completed all sets than female students with higher concentration difficulty ($p = 0.009$, see also Figure 2). No such differential effects were revealed in the massed practice condition or for male students.

None of the other cognitive and motivational variables predicted students' completion of the study.

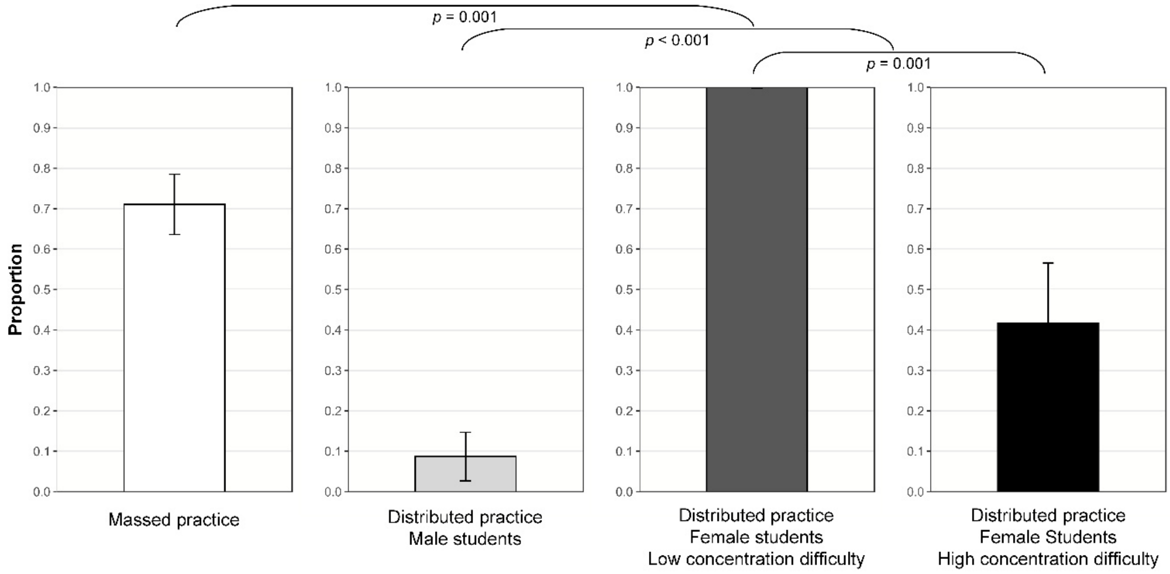



*Figure 2*. Proportion of students who finished the study, of those who initially enrolled, separated by the statistically relevant variables.

Finally, the mean perceived practice difficulty regarding the practice sets was analyzed using Bayesian t-tests, in order to determine whether distributed practice was experienced as being more difficult than massed practice. However, no evidence for differences between the conditions regarding the perceived difficulty was revealed for any of the practice sets or the test.

## Discussion

One of the main purposes of the present study was to investigate the effect of distributed practice on the mathematical performance of high school students using curriculum-relevant material. However, due to a severe dropout rate over the course of the study as a consequence of the study relying on self-regulated-learning, the effect of practice condition on final test performance has a low validity. In the sample that could be analyzed, however, the effect of practice condition unexpectedly indicated a negative effect of distributed practice as compared to massed practice. The main focus, then, was on exploratory analyses that were performed to identify factors that contributed to the students' participation in the study and completion of the study exercises. This is especially important with regard to self-regulated learning, as these factors can give insight into the question of whether students are willing to implement distributed practice in their own learning schedule and, more specifically, which students are willing to do so.

These exploratory analyses revealed some interesting results. First of all, the proportion of students who finished the study was significantly higher in the massed practice condition than in the distributed practice condition. Furthermore, within the distributed practice condition, additional differential effects were found: The proportion of students who finished their exercises was significantly higher among female students than among male students. In addition, within female students who practiced in a distributed manner, the proportion of students who finished the study was significantly higher for girls with low concentration difficulty than for girls with high concentration difficulty. None of these differential effects were found for the massed practice condition. That is, not only did the students complete their exercises more often in the massed practice condition, but for the distributed practicing students, personal characteristics had an additional influence on the completion of the exercises. Taken together, these results imply that distributed practice in self-regulated learning, contrary to massed practice, favors specific students in terms of their willingness to realize this strategy, while others are at a disadvantage. Finally, the perceived difficulty of the exercises was compared between the groups, but there was no difference regarding the difficulty judgments between the distributed and massed practicing students.

Despite the exploratory character of the present results, the differential effects on exercise completion – which were found only within the distributed practice condition – are relevant when implementing distributed practice in school learning. The observed differences between the massed and distributed practice conditions concerning the effects of individual characteristics on the completion of the exercises could be explained by different challenges posed by massed and distributed practice. In contrast to massed practice, with distributed practice the students have to actively decide to resume working on the exercises on multiple occasions, instead of being able to just continue working. That is, action has to be initiated more often in distributed practice than in massed practice, potentially resulting in a higher influence of personal characteristics related to study management on the completion of the exercises (Achtziger & Gollwitzer, 2009, 2007; Gollwitzer, Heckhausen, & Ratajczak, 1990). Ultimately, this higher challenge could then lead to fewer completed exercises in the distributed practice condition as shown by our results. The relevant individual factors observed in the present study were gender and concentration difficulty. What might stimulate girls in particular to follow a distributed practice schedule? One reason could be that girls presumably possess better self-discipline and self-regulation ability than boys, which is necessary to repeatedly initiate the distributed practice process in self-regulated learning (Duckworth & Seligman, 2006; Martin, 2011; Weis, Heikamp, & Trommsdorff, 2013). In this regard, perseverance and the willingness to invest mental effort might be other promising variables that could explain which students employ distributed practice on a self-regulated basis. In fact, females outperform males on these and other related motivational variables (Neigel, Behairy, & Szalma, 2017).

The finding that among females, lower concentration difficulty was associated with a greater success at completing distributed practice tasks contradicts the assumption that learners with poor concentration ability might profit from distribution. However, concentration ability is associated with a higher engagement in learning in general (Newmann, 1992; Skinner & Belmont, 1993) and might therefore support these females to complete the distributed practice sessions.

The analyses of the perceived difficulty of the practice and test sets provided no evidence for the fact that the perceived difficulty differed between massed and distributed practicing students. However, it should be noted that the students did not explicitly rate the difficulty of the practice strategy but only the difficulty of the exercises. The students were not able to directly compare the alternative practice strategies and, thus, could not rate the relative but only the absolute difficulty. Especially because the exercises generally were not perceived as particularly difficult, potential differences could have been minimized. Additionally, the lack of a meaningful difference could also be due to prior self-selection, because students who perceived the exercises as particularly difficult may have stopped working on them in the course of the study. That is, the question of whether distributed practice is in fact perceived as more difficult than massed practice should ideally be investigated in studies with a within-subjects design.

**Limitations**

First of all, the main results of the present study are rather exploratory and hence should be verified by further studies. The question of whether distributed practice generally improves performance in mathematics compared to massed practice, however, should be investigated in studies with less emphasis on self-regulated learning in order to maintain sufficient sample sizes and reduce potential selection bias. Additionally, though the topics were picked from the regular curriculum and hence were generally relevant for the students, their performance in our study did not influence their math grades and was not even shared with the teacher. This could have negatively impacted the motivation to participate in and complete the study. Ideally, in future studies on distributed practice in a self-regulated learning context, the personal relevance of the learned content should be increased compared to the current study – for example, by grading the performance. Finally, the students worked on the exercises at home, that is, the context and state in which the students participated in the study was barely controlled. However, this limitation should apply to both conditions equally and is no explanation for the differences between practice conditions.

**Conclusions and prospects**

One of the original questions of this study – whether distributed practice improves performance in mathematical learning in high school – can hardly be answered based on the present study. The moderate evidence for a negative effect of distributed practice should not be overemphasized due to a high likelihood of self-selection in the course of the study. As long as there is no further empirical confirmation of this unusual result, the general assumption that distributed practice improves performance in later tests compared to massed practice (Carpenter et al., 2012), even with coherent mathematical material (Rohrer & Taylor, 2006, 2007; Schutte et al., 2015), should be maintained. However, it should be seriously questioned whether this advantage holds if students ultimately complete less exercises under distributed practice conditions, as observed in the current study. The main finding here was that – in contrast to massed practice – distributed practice in semi-self-regulated learning (as the schedule was externally given and not chosen by the students themselves) seems to favor students with particular characteristics: in the current study, female students with lower concentration difficulty. Because self-regulated learning plays an important role especially in high school and at university, these differential effects concerning the application of distributed practice may be problematic if they result in performance improvements of a particular group of students while disfavoring others. Teachers may prefer strategies that improve the performance of all students equally. Therefore, it is vital to know whether and which learners are capable of successfully implementing distributed practice into their own learning schedule.

In future research it should be investigated if the implications of the exploratory results of this study can be replicated and whether and how students can be supported by implementing distributed practice effectively in their self-regulated learning. A potential measure to motivate students to keep on working even in the distributed practice schedule could be to inform them prior to the practice phase about the positive effects of distributed practice. At least in completely self-regulated learning, Ariel and Karpicke (2018) could show that informing students about the positive effect of distributed practice resulted in higher use of this practice strategy. To sum up, assuming that distributed practice – when implemented under external control – improves mathematical performance of learners, the question of whether this advantage emerges only for a subgroup of learners under self-regulated learning conditions is crucial and should be further investigated.

**Acknowledgments**

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

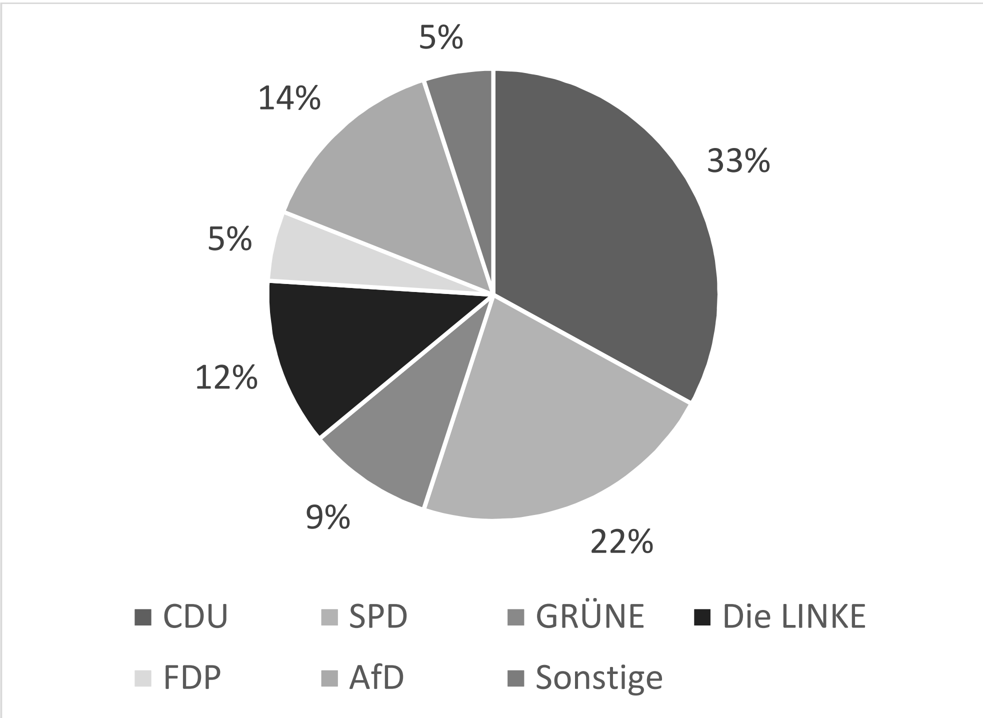
## Appendix. Exercise example

The complete material is provided in German online (https://osf.io/egt4j/?view_only=5fb49869de0345618a3c9171c77992fe). Each exercise set consisted of four exercises similar to the following.

**Exercise 1 – Political Parties**

The so called Sunday question (‘Sonntagsfrage’) is used to determine the political mood in Germany. Last sunday, 1300 citizens answered the following question: ‘Which political party would you vote for if there was an election next Sunday?‘

How many citizens would vote for the ‘CDU’ and ‘Die Grünen’? How many citizens would vote for the ‘AfD’?[5]




**Solution:**

CDU: 1300/100 = 13 → 13 * 33 = 429 voters

Die Grünen: 1300/100 = 13 → 13 * 9 = 117 voters

AfD: 1300/100 = 13 → 13 * 14 = 182 voters

[5] Note: In the experiment this chart was colored.

**Exercise 2 – Dices**

The six sides of a quadratic cube are labeled with numbers from 1 to 6. The following probabilities apply to the different sides:

P(1) = P(6) = 0.07

P(2) = P(5) = 0.13

P(3) = P(4) = 0.3

Calculate the probability for the following events if the dice is rolled once:

a) The dice rolls a 1 or a 6.

b) The dice rolls an even number.

c) The dice rolls a 3 or a 5.

**Solution:**

a) The dice rolls a 1 or a 6: P(1) + P(6) = 0.07 + 0.07 = 0.14

b) The dice rolls an even number: P(2) + P(4) + P(6) = 0.13 + 0.3 + 0.07 = 0.5

c) The dice rolls a 3 or a 5: P(3) + P(5) = 0.3 + 0.13 = 0.43

**Exercise 3 – Students I**

In the winter term 2015/2016 there were 244.322 students enrolled in different types of universities in Hessen. The distribution was as follows:

| | |
|---|---|
| University | 154274 |
| Theological University | 764 |
| University of Arts | 1737 |
| University of Applied Science | 83411 |
| University of Management | 4136 |

Please indicate the sample, name the characteristic (variable) and the possible values.

**Solution:**

Sample: all students = 244322

Characteristic (variable): type of university

Values: University, Theological University, University of Arts, University of Applied Science, University of Management

**Exercise 4 – Students II**

In the winter term 2015/2016 there were 244.322 students enrolled in different types of universities in Hessen. The distribution was as follows:

| | |
|---|---|
| University | 154274 |
| Theological University | 764 |
| University of Arts | 1737 |
| University of Applied Science | 83411 |
| University of Management | 4136 |

Calculate the missing relative frequencies and the angles for a pie chart:

University: ___

Theological University: 0.003

University of Arts: 0.007

University of Applied Science: ___

University of Management: ___

University: ___

Theological University: 1.08°

University of Arts: 2.52°

University of Applied Science: ___

University of Management: ___

**Solution:**

Relative frequencies:

University: 154274 / 244322 = 0.63

Theological University: 0.003

University of Arts: 0.007

University of Applied Science: 83411 / 244322 = 0.34

University of Management: 4136 / 244322 = 0.02

Angles:

University: 0.63 * 360 = 226.8°

Theological University: 1.08°

University of Arts: 2.52°

University of Applied Science: 0.34 * 360 = 122.4°

University of Management: 0.02 * 360 = 7.2°

**References**

Achtziger, A., & Gollwitzer, P. M. (2007). Rubicon model of action phases. *Encyclopedia of Social Psychology*, 2, 769–770.

Achtziger, A., & Gollwitzer, P. M. (2009). Rubikonmodell der Handlungsphasen. In V. Brandstätter & J. H. Otto (Eds.), *Handbuch der Allgemeinen Psychologie: Motivation und Emotion* (pp. 150–156). Göttingen: Hogrefe.

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. doi: 10.3102/0034654316689306

Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied*, 24(1), 43–56. doi: 10.1037/xap0000133

Arslan, R. C., & Tata, C. S. (2017). formr.org survey software (Version 0.16.13).

Barzagar Nazari, K., & Ebersbach, M. (2018a). *Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom*. Manuscript submitted for publication.

Barzagar Nazari, K., & Ebersbach, M. (2018b). *Distributed practice in mathematics: Recommendable especially for students on a medium performance level?* Manuscript submitted for publication.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35(2), 201–210. doi: 10.3758/BF03193441

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.

Boerner, S., Seeber, G., Keller, H., & Beinborn, P. (2005). Lernstrategien und Lernerfolg im Studium: Zur Validierung des LIST bei berufstätigen Studierenden. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37(1), 17–26. doi: 10.1026/0049-8637.37.1.17

Brunmair, M., & Richter, T. (2018). *Similarity matters: A meta-analysis of interleaved learning and its moderators*. Manuscript submitted for publication.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. doi: 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. doi: 10.1007/s10648-012-9205-z

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice. *Experimental Psychology*, 56(4), 236–246. doi: 10.1027/1618-3169.56.4.236

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. doi: 10.1037/0033-2909.132.3.354

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*, 1–19. doi: 10.1007/s10648-017-9426-2

Dalbert, C., & Radant, M. (2008). Psychologie der Schülerpersönlichkeit. In M. K. W. Schweer (Ed.), *Lehrer-Schüler-Interaktion* (pp. 127–154). Wiesbaden: VS Verlag für Sozialwissenschaften.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, 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* (Vol. 53, pp. 63–147). London: Academic Press.

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades and achievement test scores. *Journal of Educational Psychology*, 98(1), 30.

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. doi: 10.1177/1529100612453266

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169–189. doi: 10.1207/s15326985ep3403_3

Gollwitzer, P. M., Heckhausen, H., & Ratajczak, H. (1990). From weighing to willing: Approaching a change decision through pre- or postdecisional mentation. *Organizational Behavior and Human Decision Processes*, 45(1), 41–65. doi: 10.1016/0749-5978(90)90004-S

Hirsch, C. R., Kapoor, S. F., & Laing, R. A. (1982). Alternative models for mathematics assignments. *International Journal of Mathematical Education in Science and Technology*, 13(3), 243–252. doi: 10.1080/0020739820130301

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674. doi: 10.1198/106186006X133933

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 partytioning in R. *Journal of Machine Learning Research*, 16, 3905–3909.

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

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. doi: 10.1126/science.1152408

Kruschke, J. (2015). *Doing Bayesian Data Analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). London: Academic Press.

Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition*, 42(5), 729–741. doi: 10.3758/s13421-014-0394-1

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course.* Cambridge: Cambridge University Press.

Martin, A. J. (2011). School motivation of boys and girls: Differences of degree, differences of kind, or both? *Australian Journal of Psychology*, 56(3), 133–146. doi: 10.1080/00049530412331283363

McArdle, J. J., & Ritschard, G. (2013). *Contemporary issues in exploratory data mining in the behavioral sciences*. New York: Routledge.

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language*, 27(5), 521–536. doi: 10.1016/0749-596X(88)90023-X

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (R package version 0.9.12-2). Retrieved from https://CRAN.R-project.org/package=BayesFactor

Neigel, A. R., Behairy, S., & Szalma, J. L. (2017). Need for cognition and motivation differentially contribute to student performance. *Journal of Cognitive Education and Psychology*, 16(2), 144–156. doi: 10.1891/1945-8959.16.2.144

Newmann, F. M. (1992). *Student engagement and achievement in American secondary schools*. New York: Teachers College Press.

Nicholls, J. G., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing students' theories of success in mathematics: Individual and classroom differences. *Journal for Research in Mathematics Education*, 21(2), 109–122. doi: 10.2307/749138

Project Team / Schmitz, C. (2012). LimeSurvey: An open source survey tool. Hamburg: LimeSurvey Project. Retrieved from http://www.limesurvey.org

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

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, 40(1), 4–17.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, 24(3), 355–367. doi: 10.1007/s10648-012-9201-3

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. doi: 10.1002/acp.1266

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6), 481–498. doi: 10.1007/s11251-007-9015-8

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. doi: 10.1037/a0037559

Salis, K. L., Kliem, S., & O'Leary, K. D. (2014). Conditional inference trees: A method for predicting intimate partner violence. *Journal of Marital and Family Therapy*, 40(4), 430–441. doi: 10.1111/jmft.12089

Saxon, J. (1982). Incremental development: A breakthrough in mathematics. *The Phi Delta Kappan*, 63(7), 482–484.

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. doi: 10.1016/j.jsp.2014.12.003

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85(4), 571–581. doi: 10.1037/0022-0663.85.4.571

Spinath, B., Stiensmeier-Pelster, J., Schöne, C., & Dickhäuser, O. (2012). *Skalen zur Erfassung der Lern- und Leistungsmotivation: SELLMO*. Göttingen: Hogrefe.

Stan Development Team. (2018). RStan: The R interface to Stan (R package version 2.17.3). Retrieved from http://mc-stan.org/

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323–348.

Toppino, T. C., Kasserman, J. E., & Mracek, W. A. (1991). The effect of spacing repetitions on the recognition memory of young children and adults. *Journal of Experimental Child Psychology*, 51(1), 123–138. doi: 10.1016/0022-0965(91)90079-8

Weis, M., Heikamp, T., & Trommsdorff, G. (2013). Gender differences in school achievement: The role of self-regulation. *Frontiers in Psychology*, 4. doi: 10.3389/fpsyg.2013.00442

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

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

Yazdani, M. A., & Zebrowski, E. (2006). Spaced reinforcement: An effective approach to enhance the achievement in plane geometry. *Journal of Mathematical Sciences and Mathematics Education*, 1, 37–43.

Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-Efficacy in changing societies* (pp. 202–231). New York: Cambridge University Press.

# CHAPTER V: Summary and general discussion

## 1. Summary and discussion

Distributed practice is a promising learning strategy that, compared to massed practice, is supposed to enhance processing during learning and improve retrieval skills – the two important features of successful learning strategies (Bjork, 1994). A rich empirical body exists on the positive effect of distributed practice on the retention of incoherent verbal material. When isolated facts or word(-pair)s have to be memorized, it is usually of advantage to spread the repetitions over several separated learning occasions compared to learning the same duration or with the same number of repetitions without interruption (Cepeda et al., 2006). However, less is known about the effect of distributed practice within mathematical learning. Though in some studies the effect of distributed practice on mathematical performance was investigated, different knowledge types were addressed (conceptual and/or procedural knowledge; Rittle-Johnson & Schneider, 2015) and the practiced content varied widely regarding the complexity (targeting different levels of semantical understanding; Tietze, 1988). Additionally, it was only rarely investigated if and how the effect of distributed practice emerges in real learning contexts, because only in few studies distributed practice was implemented in schools or universities. The few existing studies often did not include pure massed practice conditions or combined distributed practice with other learning strategies, impeding reliable statements on the effect of distributed practice compared to massed practice specifically (e.g., Chen et al., 2017; Schutte et al., 2015). In the present dissertation I addressed some of these research gaps by following the question if distributed practice can be used to improve the mathematical performance of school students. Three studies were conducted that are summarized shortly.

The research objective that was pursued with the first study, *Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom*, was to investigate whether distributed practice generally improves the performance of school students learning mathematics. Two experiments were conducted, one in Grade 3 and one in Grade 7. The third graders practiced a semi-formal multiplication method and the results provided clear evidence for a positive effect of distributed practice compared to massed practice one week after the last practice session. This positive effect, however, did not endure five weeks later, six weeks after the last practice. The discussed reasons were that the students probably practiced related content working on other mathematical topics, which may have diminished the differences between the two practice conditions in Grade 3. In Grade 7, the students practiced basic stochastics and although the evidence was less strong, the positive effect of distributed practice was more stable and similar performance differences between distributed and massed practicing students were observed one and six weeks after the last practice. However, the hypothesis that the positive effect of distributed practice would be even more pronounced in the later test compared to the first test could not be confirmed in both grade levels.

In the second study, *Distributed practice in mathematics: Recommendable especially for students on a medium performance level?*, the results of the first study should be substantiated. Additionally, it was investigated if the inclusion of feedback in Grade 7 results in an overall better performance level and/or a larger effect of distributed practice. Therefore, an experiment very similar to the second experiment of the first study was conducted, where seventh graders learned and practiced basic stochastics. This time, however, the students received a sample solution after each practice exercise. Additionally, there was a small design change because an expanding interval schedule was applied for the students of the distributed practice condition (two and five days between the practice sessions instead of practice on three consecutive days). In the first test, which was conducted two weeks after the last practice, surprisingly no evidence for an effect of practice condition was revealed. It was not until the second test, six weeks after the last practice, that strong evidence for a positive effect of distributed practice could be observed. That is, this time the positive effect of distributed practice was, as originally expected, especially pronounced in the long run. A possible explanation for the missing effect in the first test may be that the provision of sample solutions improved the performance for students of both conditions, which may have reduced the effect of the practice condition intermediately. Finally, an important exploratory result of the second study was that the positive effect of distributed practice was particularly large in the medium performance range. In the lower and higher performance ranges, by contrast, the performance between first practice and test remained rather stable. Because this relationship was only observed exploratory, additional studies are needed to figure out if this was a random result or if it can be confirmed that the positive effect of distributed practice on mathematical performance especially emerges for medium performers.

Picking up the dimensions that were introduced in the beginning of this dissertation, the exercises created for the first two studies targeted both conceptual and procedural knowledge. The exercises in Grade 3 primarily targeted procedural knowledge, because the students repeatedly had to perform the several steps of the semi-formal multiplication method (splitting up the multiplicands, multiplication of the splitted numbers and addition of the results). In Grade 7, rather a mix of conceptual and procedural knowledge was required (e.g., being able to calculate simple probabilities or draw tree diagrams, but also correctly interpreting tree diagrams). Compared to most other studies on distributed mathematical practice, the degree of semantical understanding involved in both studies was relatively high: The content was thoroughly taught and explained prior to the practice periods, and it was made sure that the respective procedures and solutions could be understood. In most cases, simply memorizing the correct procedure or solution would have been sufficient to correctly solve the tasks. However, like in real-world mathematics education, the students were provided with enough background information to be able to really grasp the underlying concepts and hence improve semantical understanding.

Encouragingly, even in this setting the collected evidence of the first two studies generally supported the results known from classical laboratory studies on the effect of distributed practice in that in most cases, students who followed a distributed practice schedule outperformed students practicing in a massed manner in intermediate and/or long-term follow-up tests. Most importantly, in no case a negative effect of distributed practice was observed. That is, the assumption stated in the Introduction, that even in the case of more complex, mathematical material distributed practice does improve performance compared to massed practice was supported.

An interesting exploratory result from the second study, however, indicated that in the case of more complex content the positive effect of distributed practice might especially emerge in medium performance ranges. This result was similar to results of two abovementioned studies on mixed mathematical practice (interleaved and distributed practice; Hirsch et al., 1982; Saxon, 1982). The proposed explanation was that potentially, both students of low and high ability are rather *unaffected by practice at all*, because independent from practice they are not able to understand the topic or at least follow the provided procedure (low performers) or they grasp the underlying concepts very quickly (high performers). Using the familiar dimensions again, regarding the high performers this could mean that if the new concept can quickly be connected to existing conceptual and/or procedural knowledge in long-term memory (i.e., a high level of semantical understanding is reached virtually immediately), the knowledge is stored in a very strong and accessible memory trace that is available for a relatively long time, regardless of practice or the practice procedure. On the other hand, if the ability is too low (maybe because the student did not even pay attention to the lecture) and the student is unable to integrate the newly learned concepts into long-term memory at all, even when repeatedly provided practice exercises and sample solutions, practice and hence the practice condition itself do not matter, too. However, for students whose performance *could* be influenced by practice exercises, it could be shown that distributed practice exceeds massed practice, at least in the long term.

A crucial question that was not answered by the studies and that should be subject to future research, is which features of memory exactly are improved by distributed practice. The presented studies provide evidence for the fact that distributed practice can be used to improve mathematical performance even when more complex material is involved. This is a promising result that can be built upon in future research. There, it should be systematically investigated whether only memory (the strength of a specific memory trace) or the degree of the organizational integration (level of semantical understanding) is improved by distributed practice. Taking up the model provided by Delaney et al. (2010), it could be theorized, for example, that distributed mathematical practice does not only improve the item content information but also associative information, which could help the students to really grasp the logic behind the correct solution. One way to analyze this may be to include transfer tasks in the tests, because a higher level of semantical understanding should result in a higher transfer performance than the more isolated memorization of specific procedures and solutions.

Based on the promising results of the first two studies, in the third study, *Distributed practice: Rarely realized in self-regulated mathematical learning*, distributed practice was investigated as a learning strategy in a more self-regulated context. That is, the procedure differed from the other studies in that the students were assigned one of two practice conditions, but had to realize the practice sessions themselves, at home. This procedure resulted in a large drop-out, particularly in the distributed practice condition. Regarding the final performance, which was tested two weeks after the last practice, this time a negative effect of distributed practice was observed. However, this result should not be overemphasized, because a self-selection bias is very likely. The reasons for the large drop-out were the focus of the further analyses. Exploratory analyses revealed differential effects only in the distributed practice condition: Mainly female participants with little concentration difficulties finished their exercises in the distributed practice condition. Male participants generally and female participants with higher concentration difficulties finished their exercises less often. No such differential effects were observed in the massed practice condition. These effects should be further investigated in future studies as they implicate that massed and distributed practice may impose different requirements on the learner, with distributed practice stressing more personal characteristics than massed practice.

To sum it up, in the third study – mainly due to an unexpected high drop-out – not the effect of distributed practice was in the focus but rather the implementation in a more self-regulated context. The negative effect of distributed practice compared to massed practice therefore should not be overemphasized – the study design was only restrictedly appropriate to investigate the effect itself. However, the results of this study illustrate problems that hitherto were only marginally brought up when it comes to research of distributed practice: Are students (equally) able to actually realize the strategy within their self-regulated learning? Until more research is available on this subject, distributed practice ideally should be externally guided.

## 2. Conclusion

Generally, at least when integrated in the classroom schedule, distributed practice seems to improve the mathematical performance of school students. Therefore, repeated practice with mathematical exercises should ideally be distributed, based on the results of the presented studies over several days. Shorter intervals, accomplished by distributing practice over a school day, in some circumstances may also be practical for the school routine and potentially, even these shorter interstudy intervals already result in positive effects on the mathematical performance (see Schutte et al. 2015), though they have not been compared to a pure massed practice condition in similar settings yet. Additionally, it is likely that related forms of learning (e.g., physics or other mathematical topics besides semi-formal multiplication and stochastics) similarly benefit from distributed practice. Nonetheless, research on the effect of distributed practice would profit from more studies with a greater variety of material to prove this hypothesis.

It is not yet clarified which dimensions of mathematical learning are particularly improved by distributed practice – rather isolated memory of conceptual knowledge and procedural steps and/or the degree of semantical understanding. Based on the conducted studies no conclusion was possible, because the effects on more isolated forms of memory and on semantical understanding cannot be differentiated. The inclusion of proper transfer performance tasks could change that, because a higher performance in transfer tasks could be attributed to semantical understanding.

Regarding the actual implementation of distributed practice into educational practice, several important aspects have to be considered by both teachers and researchers: Usually, the majority of mathematical practice is part of the learning that happens outside the classroom. Because class time in school is scarce, it is mostly used for input and short homework reviews or the discussion of questions. Repetition and practice, on the other hand, are largely part of homework and individual study time. Unfortunately, by both researchers and teachers it should be expected that students rarely implement distributed practice into their practice schedule autonomously and without prior training (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). In addition, they only unreliably realize an externally provided practice schedule, as in Study 3 of this dissertation. Due to the limited amount of time in school, it is unlikely that teachers are able to include much more practice into their school routine. However, they could not give all exercises on one topic at once and rely on the students to distribute their practice themselves, but instead give out the exercises across several days or even weeks. Here, it would be crucial to make the exercises mandatory to avoid incomplete coverage of the practice material, such as in Study 3. With time, the students would probably get used to shorter, but more frequent practice assignments. Additionally, they should be instructed early on about the positive effects of distributed practice (and other learning strategies, for that matter) and how they can be implemented in their self-regulated learning in order to increase the probability that finally they can make use of it themselves (Dunlosky et al., 2013).

To conclude, distributed practice remains a promising practice strategy for complex topics like mathematics. Future studies should especially focus on (a) the question which features of memory exactly are improved by distributed mathematical practice compared to massed practice and (b) pursue indications from the third study by investigating whether distributed practice compared to massed practice is differentially used by learners within their self-regulated learning. Additionally, it should be investigated how students can be supported by the implementation of distributed practice, so that eventually as many students as possible can profit from this strategy.

# References

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science*, *13*(1), 27–33. doi: 10.1111/1467-9280.00405

Anderson, J. R. (2013). *Kognitive Psychologie*. Berlin, Heidelberg: Springer VS.

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. *Journal of Consumer Research*, *32*(2), 266–276. doi: 10.1086/432236

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). New York: Academic Press. doi: 10.1016/S0079-7421(08)60422-3

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The Foundations of Remembering: Essays in Honor of Henry L. Roediger, III*. New York: Psychology Press.

Baroody, A. J., Feil, Y., & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, *38*(2), 115–131. doi: 10.2307/30034952

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. doi: 10.3758/BF03193441

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*(4), 635–650. doi: 10.1017/S0142716410000172

Bjork, R. A. (1975). Retrieval as 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., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes* *to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Budé, L., Imbos, T., Wiel, M. W. van de, & Berger, M. P. (2010). The effect of distributed practice on students' conceptual understanding of statistics. *Higher Education*, *62*(1), 69–79. doi: 10.1007/s10734-010-9366-y

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. doi: 10.1007/s10648-012-9205-z

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice. *Experimental Psychology*, *56*(4), 236–246. doi: 10.1027/1618-3169.56.4.236

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. doi: 10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning. A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102. doi: 10.1111/j.1467-9280.2008.02209.x

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*, 1–19. doi: 10.1007/s10648-017-9426-2

Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly*, *33*(4), 655–680. doi: 10.2307/3587881

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.

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. doi: 10.1016/S0022-5371(72)80001-X

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum.

Delaney, P. F., Godbole, N. R., Holden, L. R., & Chang, Y. (2018). Working memory capacity and the spacing effect in cued recall. *Memory*, *26*(6), 784–797. doi: 10.1080/09658211.2017.1408841

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, 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* (Vol. 53, pp. 63–147). London: Academic Press.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, *1*(4), 309–330. doi: 10.1007/BF01320097

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795–805. doi: 10.1037/0021-9010.84.5.795

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. doi: 10.1177/1529100612453266

Ebbinghaus, H. (1913). *Memory. A contribution to experimental psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College, Columbia University. Retrieved from http://archive.org/details/memorycontributi00ebbiuoft

Gay, L. R. (1973). Temporal position of reviews and its effect on the retention of mathematical rules. *Journal of Educational Psychology*, *64*(2), 171–182. doi: 10.1037/h0034595

Gerbier, E., & Koenig, O. (2012). Influence of multiple-day temporal distribution of repetitions on memory: A comparison of uniform, expanding, and contracting schedules. *The Quarterly Journal of Experimental Psychology*, *65*(3), 514–525. doi: 10.1080/17470218.2011.600806

Gerbier, E., Toppino, T. C., & Koenig, O. (2014). Optimising retention through multiple study opportunities over days: The benefit of an expanding schedule of repetitions. *Memory*, *23*(6), 943–954.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112. doi: 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. doi: 10.1002/acp.3245

Grote, M. G. (1995). The effect of massed versus spaced practice on retention and problem-solving in high school physics. *School Science and Mathematics*, *95*(2), 97–101. doi: 10.1111/j.1949-8594.1995.tb15736.x

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 77–99). Oxford, England: Lawrence Erlbaum.

Hirsch, C. R., Kapoor, S. F., & Laing, R. A. (1982). Alternative models for mathematics assignments. *International Journal of Mathematical Education in Science and Technology*, *13*(3), 243–252. doi: 10.1080/0020739820130301

Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. S. (2015). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, *28*(4), 853–873. doi: 10.1007/s10648-015-9349-8

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. doi: 10.1016/j.learninstruc.2014.11.001

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

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science*, *42*(3), 373–388. doi: 10.1007/s11251-013-9285-2

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, *59*(4), 277–287. doi: 10.1080/02701367.1988.10609373

Lefrancois, G. R. (2015). *Psychologie des Lernens* (5th ed.). Berlin, Heidelberg: Springer-Verlag.

Maddox, G. B., Balota, D. A., Coane, J. H., & Duchek, J. M. (2011). The role of forgetting rate in producing a benefit of expanded over equal spaced retrieval in young and older adults. *Psychology and Aging*, *26*(3), 661–670. doi: 10.1037/a0022942

Malmberg, K. M., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 322–336. doi: 10.1037/0278-7393.31.2.322

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, *39*(4), 352–370. doi: 10.1037/h0069819

Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System*, *42*, 412–428. doi: 10.1016/j.system.2014.01.014

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. doi: 10.1006/cogp.1999.0734

Moulton, C.-A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: What kind of practice makes perfect?: A randomized, controlled trial. *Annals of Surgery*, *244*(3), 400. doi: 10.1097/01.sla.0000234808.85789.6a

Norman, D. A., & Shallice, T. (1986). Attention to action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–18). Boston, MA: Springer. doi: 10.1007/978-1-4757-0629-1_1

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning: Journal of Practical Research & Applications*, *4*(1), 11–18.

Rickard, T. C., Lau, J. S.-H., & Pashler, H. (2008). Spacing and the transition from calculation to retrieval. *Psychonomic Bulletin & Review*, *15*(3), 656–661. doi: 10.3758/PBR.15.3.656

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford Handbook of Numerical Cognition*. Oxford, England: Oxford University Press.

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*(4), 587–597. doi: 10.1007/s10648-015-9302-x

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. doi: 10.1016/j.tics.2010.09.003

Roediger III, H. L., & Karpicke, J. D. (2011). Intricacies of spaced retrieval: A resolution. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 23–47). New York, NY: Psychology Press. doi: 10.4324/9780203842539-7

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*(3), 355–367. doi: 10.1007/s10648-012-9201-3

Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Educational Psychology Review*. doi: 10.1007/s10648-015-9332-4

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. doi: 10.1002/acp.1266

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*(6), 481–498. doi: 10.1007/s11251-007-9015-8

Saxon, J. (1982). Incremental development: A breakthrough in mathematics. *The Phi Delta Kappan*, *63*(7), 482–484.

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. doi: 10.1016/j.jsp.2014.12.003

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology*, *19*(1), 107–122. doi: 10.1002/acp.1066

Simone, P. M., Bell, M. C., & Cepeda, N. J. (2013). Diminished but not forgotten: Effects of aging on magnitude of spacing effect benefits. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *68*(5), 674–680. doi: 10.1093/geronb/gbs096

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*(5), 763–767. doi: 10.1002/acp.1747

Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 255–262. doi: 10.1037/a0017892

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. doi: 10.1016/0022-5371(76)90047-5

Tietze, U.-P. (1988). Schülerfehler und Lernschwierigkeiten in Algebra und Arithmetik — Theoriebildung und empirische Ergebnisse aus einer Untersuchung. *Journal für Mathematik-Didaktik*, *9*(2–3), 163–204. doi: 10.1007/BF03339290

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 60, pp. 113–189). Waltham, MA: Academic Press.

Tulving, E. (1974). Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *American Scientist*, *62*(1), 74–82.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 796–800.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*(1), 235–269. doi: 10.1146/annurev.psych.55.090902.141555

Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(6), 545–559. doi: 10.1037/0278-7393.8.6.545

# Universität Kassel, Fachbereich Humanwissenschaften
# Erklärung zur Dissertation im Promotionsfach Psychologie

Erklärung über den Eigenanteil an den zur Veröffentlichung vorgesehenen wissenschaftlichen Schriften innerhalb der Dissertationsschrift.

**I. Allgemeine Angaben**

Barzagar Nazari, Katharina

Institut für Psychologie, Universität Kassel

Thema der Dissertation: *Investigating distributed practice as a strategy for school students learning mathematics*

**II. Nummerierte Aufstellung der eingereichten Schriften**

1. Barzagar Nazari, K., & Ebersbach, M. (2018). Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom. *Applied Cognitive Psychology*. doi: 10.1002/acp.3485
2. Barzagar Nazari, K., & Ebersbach, M. (2018). *Distributed practice in mathematics: Recommendable especially for students on a medium performance level?* Manuscript submitted for publication.
3. Barzagar Nazari, K., & Ebersbach, M. (2018). Distributed practice: Rarely realized in self-regulated mathematical learning. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.02170

**III. Darlegung des eigenen Anteils an diesen Schriften:**

Zu 1.:

- Konzeption: in Teilen
- Literaturrecherche: mehrheitlich
- Methodenentwicklung: mehrheitlich
- Entwicklung des Versuchsdesigns: mehrheitlich
- Datenerhebung: in Teilen
- Datenauswertung: vollständig

- Ergebnisdiskussion: überwiegend
- Erstellen des Manuskripts: überwiegend
- Überarbeitung nach peer-review: überwiegend

Zu 2.:

- Konzeption: in Teilen
- Literaturrecherche: mehrheitlich
- Methodenentwicklung: überwiegend
- Entwicklung des Versuchsdesigns: überwiegend
- Datenerhebung: in Teilen
- Datenauswertung: vollständig
- Ergebnisdiskussion: überwiegend
- Erstellen des Manuskripts: überwiegend
- Überarbeitung nach peer-review: überwiegend

Zu 3.:

- Konzeption: mehrheitlich
- Literaturrecherche: mehrheitlich
- Methodenentwicklung: überwiegend
- Entwicklung des Versuchsdesigns: vollständig
- Datenerhebung: in Teilen
- Datenauswertung: vollständig
- Ergebnisdiskussion: überwiegend
- Erstellen des Manuskripts: überwiegend
- Überarbeitung nach peer-review: mehrheitlich

## IV. Anschrift der Mitautorin

Prof. Dr. Mirjam Ebersbach: mirjam.ebersbach@uni-kassel.de

**Unterschrift der Antragstellerin**

………………………… …………………………

Ort, Datum Unterschrift

**Ich bestätige die von Frau Katharina Barzagar Nazari unter Punkt III abgegebene Erklärung:**

Prof. Dr. Mirjam Ebersbach: …………………………

Unterschrift

# Eidesstattliche Versicherung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig, ohne unerlaubte Hilfe Dritter angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden.

………………………… ………………………………

Ort, Datum Unterschrift

## Danksagung

An erster Stelle danke ich Mirjam für das Vertrauen, das sie mir entgegengebracht hat, für die Zusammenarbeit, die durch stets wertschätzende und konstruktive Rückmeldungen geprägt war, und dafür, dass sie auch diejenigen meiner Interessen gefördert und unterstützt hat, die über mein unmittelbares Dissertationsthema hinausgehen. Ich habe es in großen Teilen Mirjam zu verdanken, dass meine Zeit an der Universität Kassel so spannend, vielseitig und bereichernd war!

Außerdem danke ich all meinen Kolleginnen und Kollegen für das freundliche Miteinander, das ich erlebt habe, und für viele erholsame Mittags- und Nachmittagspausen sowie den LOEWE-Hilfskräften und -Praktikantinnen für ihre tatkräftige Mitarbeit in den vergangenen Jahren. Insbesondere möchte ich meinen Kolleginnen Carla und Irina für den Austausch und die gemeinsame Zeit im Systembau danken (und Carla nochmal extra für das wahnsinnig gewissenhafte Korrektur lesen von Teilen dieser Arbeit!).

Ich möchte mich natürlich auch bei allen Schulen, Lehrerinnen und Lehrern sowie Schülerinnen und Schülern bedanken, die sich bereiterklärt haben, an den Studien teilzunehmen und damit die Arbeit für den LOEWE-Forschungsschwerpunkt überhaupt erst ermöglicht haben.

Nicht zuletzt gilt meine tiefe Dankbarkeit Erik, meiner Familie sowie all den anderen wunderbaren Menschen, die mich durch mein Leben begleiten und meine Freizeit mit mir verbringen. Sie alle haben dazu beigetragen, dass ich auch während der Tiefs und Hochs, die die Arbeit an dieser Dissertation mit sich gebracht hat, die ganz wesentlichen Dinge des Lebens nicht aus den Augen verloren habe.