Improving Learning from Texts:

Distributed Practice and Distributed Learning as Desirable Difficulty in Reading Single and Multiple Texts

Inaugural-Dissertation

Zur Erlangung der Doktorwürde der

Fakultät für Humanwissenschaften

der

Julius-Maximilians-Universität Würzburg

Vorgelegt von

Carla Elisabeth Greving

aus Kassel

Würzburg

2021

Erstgutachterin:	Professorin Dr. Mirjam Ebersbach
Zweitgutachter:	Professor Dr. Wolfgang Lenhard
Betreuer:	Professor Dr. Tobias Richter
Tag der Disputation:	18.03.2022

Table of Contents

Summary

Distributed practice is a well-known learning strategy whose beneficial effects on long-term learning are well proven by various experiments. In learning from texts, the benefits of distribution might even go beyond distributed practice, i.e. distribution of repeated materials. In realistic learning scenarios as for example school or university learning, the reader might read multiple texts that not repeat but complement each other. Therefore, distribution might also be implemented between multiple texts and benefit long-term learning in analogy to distributed practice. The assumption of beneficial effects of this distributed *learning* can be deduced from theories about text comprehension as the landscape model of reading (van den Broek et al., 1996) in combination with theories of desirable difficulties in general (R. A. Bjork & Bjork, 1992) and distributed practice in particular (Benjamin & Tullis, 2010). This dissertation aims to investigate (1) whether distributed learning benefits learning; (2) whether the amount of domain-specific prior knowledge moderates the effects of distribution, (3) whether distributed learning affects the learner's meta-cognitive judgments in analogy to distributed practice and (4) whether distributed practice is beneficial for seventh graders in learning from single text.

In Experiment 1, seventh graders read two complementary texts either massed or distributed by a lag of one week between the texts. Learning outcomes were measured immediately after reading the second text and one week later. Judgements of learning were assessed immediately after each text. Experiment 2 replicated the paradigm of Experiment 1 while shortening the lag between the texts in the distributed condition to 15 min. In both experiments, an interaction effect between learning condition (distributed vs. massed) and retention interval (immediate vs. delayed) was found. In the distributed condition, the participants showed no decrease in performance between the two tests, whereas participants in the massed condition did. However, no beneficial effects were found in the delayed test for the distributed condition but even detrimental effects for the distributed condition in the immediate test. In Experiment 1, participants in the distributed condition perceived learning as less difficult but predicted lower success than the participants in the massed condition.

Experiment 3 replicated the paradigm of Experiment 1 with university students in the laboratory. In the preregistered Experiment 4, an additional retention interval of two weeks was realized. In both experiments, the same interaction between learning condition and retention interval was found. In Experiment 3, the participants in the distributed condition again showed no decrease in performance between the two tests, whereas participants in the massed condition did. However, even at the longer retention interval in Experiment 4, no beneficial effects were found for the distributed condition. Domain-specific prior knowledge was positively associated with test performance in both experiments. In Experiment 4, the participants with low prior knowledge seemed to be impaired by distributed learning, whereas no difference was found for participants with medium or high prior knowledge.

In the preregistered Experiment 5, seventh graders read a single text twice. The rereading took place either massed or distributed with one week. Immediately after rereading, judgements of learning were assessed. Learning outcomes were assessed four min after second reading or one week later. Participants in the distributed condition predicted lower learning success than participants in the massed condition. An interaction effect between learning condition and retention interval was found, but no advantage for the distributed condition. Participants with low domain-specific prior knowledge showed lower performance in short-answer questions in the distributed condition than in the massed condition.

Overall, the results seem less encouraging regarding the effectiveness of distribution on learning from single and multiple texts. However, the experiments reported here can be perceived as first step in the realistic investigation of distribution in learning from texts.

Zusammenfassung

Verteiltes Üben ist eine bekannte Lernstrategie, deren positiver Effekt auf die langfristigen Behaltensleistung in vielen Experimenten gezeigt wurde. Beim Lernen mit Texten können die Vorteile der Verteilung von Lerninhalten sogar über verteiltes Üben, also die Verteilung von sich wiederholenden Materialien, hinausgehen. In realistischen Lernszenarien, wie zum Beispiel Lernen in der Schule oder Universität, werden multiple Texte gelesen, die einander nicht wiederholen, sondern ergänzen. Verteilung im Sinne des verteilten Übens könnte demensprechend auch zwischen diesen multiplen Texten eingesetzt werden und analog zum verteilten Üben langfristiges Behalten fördern. Annahmen über eine Wirksamkeit dieses sogenannten verteilten Lernens können von Theorien zum Textverständnis wie zum Beispiel dem Landscape Model of Reading (van den Broek et al., 1996) in Kombination mit Theorien zu wünschenswerten Erschwernissen im Allgemeinen (R. A. Bjork & Bjork, 1992) und verteiltem Üben im Besonderen (Benjamin & Tullis, 2010) abgeleitet werden.

In dieser Dissertation soll untersucht werden, (1) ob verteiltes Lernen Behalten fördert, (2) ob die Ausprägung des domänenspezifischen Vorwissens die Effekte des verteilten Lernens moderiert, (3) ob verteiltes Lernen die meta-kognitive Beurteilung des Lernprozesses des Lernenden beeinflusst und (4) ob verteiltes Üben für 7. Klässler*innen beim Lernen mit einzelnen Texten vorteilhaft ist.

Im ersten Experiment lasen 7. Klässler*innen zwei komplementäre Texte entweder massiert oder verteilt mit einem Abstand (lag) von einer Woche. Die Behaltensleistung wurde unmittelbar nach dem Lesen des zweiten Textes und eine Woche später erfasst. Die Beurteilung des Lernprozesses wurde direkt nach dem Lesen eines Textes erhoben. Experiment 2 replizierte das Paradigma von Experiment 1, wobei der Abstand zwischen den Texten in der verteilten Bedingung auf 15 Min reduziert wurde. In beiden Experimenten wurde eine Interaktion zwischen der Lernbedingung (massiert vs. verteilt) und dem Behaltensintervall (unmittelbar vs. später) gefunden. In der verteilten Bedingung zeigten die Teilnehmenden keine Verringerung

in der Leistung zwischen den beiden Tests, die Teilnehmenden in der massierten Bedingung jedoch schon. Trotzdem wurden keine Vorteile des verteilten Lernens beim späteren Test gefunden. Stattdessen zeigten sich unmittelbar nach dem Lernen Nachteile des verteilten Lernens. In Experiment 1 nahmen die Teilnehmenden der verteilten Bedingung das Lernen als weniger schwierig wahr, erwarteten jedoch geringeren Lernerfolg als Teilnehmende der massierten Bedingung.

Experiment 3 replizierte das Paradigma von Experiment 1 mit Studierenden in einem Laborsetting. In dem präregistrierten Experiment 4 wurde ein Behaltensintervall von zwei Wochen ergänzt. In beiden Experimenten fand sich eine ähnliche Interaktion zwischen Lernbedingung und Behaltensintervall. In Experiment 3 zeigten die Teilnehmenden in der verteilten Bedingung erneut keine Verringerung in der Leistung zwischen den zwei Tests, in der massierten Bedingung jedoch schon. Es fanden sich jedoch keine Vorteile des verteilten Lernens, auch nicht nach dem längeren Behaltensintervall in Experiment 4. Domänenspezifisches Vorwissen stand in beiden Experimenten in einem positiven Zusammenhang mit der Behaltensleistung. In Experiment 4 zeigten die Teilnehmenden mit geringem Vorwissen geringere Behaltensleistungen, wenn sie verteilt lernten, während keine derartigen Unterschiede für Teilnehmende mit mittlerem oder hohem Vorwissen gefunden wurden.

Im präregistrierten Experiment 5 lasen 7. Klässler*innen einen einzelnen Text zweimal. Das zweite Lesen erfolgte hier entweder unmittelbar (massiert) oder nach einer Woche (verteilt). Unmittelbar nach dem zweiten Lesen wurden die Beurteilung des Lernprozesses erhoben. Die Behaltensleistung wurde 4 Minuten oder eine Woche nach dem zweiten Lesen erfasst. Teilnehmende in der verteilten Bedingung erwarteten geringeren Lernerfolg als Teilnehmende in der massierten Bedingung. Es wurde ein Interaktionseffekt zwischen der Lernbedingung und dem Behaltensintervall gefunden, aber kein Vorteil für die verteilte Bedingung. In der verteilten Bedingung zeigten Teilnehmende mit geringem Vorwissen geringere Leistung bei Fragen im Kurzantwortformat als vergleichbare Teilnehmende in der massierten Bedingung.

Zusammenfassend wirken diese Ergebnisse wenig vielversprechend bezüglich der Wirksamkeit des verteilten Lernens und Übens mit einzelnen und multiplen Texten. Trotzdem sind die hier berichteten Experimente als ein erster Schritt zur Untersuchung des Verteilungseffektes beim Lernen mit Texten in realistischen Lernszenarien zu betrachten.

Chapter I

Introduction

Distribution as Desirable Difficulty in Learning from Texts: Theory, Evidence and Research Questions

Introduction

In 2010, I visited a university course named “learning and teaching with texts and pictures”. In this course, I learned about the distributed practice effect – but this was not the first time I was introduced to this concept. On the contrary, the idea of distributing learning evenly across time accompanied my whole school career. “Stop cramming!” was the advice, and it came with the assumption, that cramming all learning content into the last days (or last night) before a test is not beneficial for learning. Furthermore, it is associated with laziness, it is perceived as “the end-phase of a much longer, lousy studying strategy” (Young, 2018). As an avowed procrastinator, I was used to cramming and even enjoyed the short but intense learning phases. However, I somehow accepted that I should distribute my learning for better results, especially in the long run (Cepeda et al., 2006; Donovan & Radosevich, 1999). However, in starting my dissertation on learning from texts, I pondered the question whether distributed practice is also beneficial for text materials. Although some experiments showed beneficial effects for learning from single texts (Glover & Corkill, 1987; Krug et al., 1990; Rawson, 2012; Rawson & Kintsch, 2005), I found no evidence regarding learning from multiple texts. In school and university however, learning from multiple texts is the learning I was used to. In real-world learning scenarios, we seldom learn the very same facts (or word pairs) repeatedly and then are tested on these facts only. Instead, we have to learn several related concepts, a huge amount of learning material that is not repeated but complemented during the school year or semester. To my knowledge, distributing learning across time has seldom been applied to those learning scenarios, which raises the question whether distributing learning time is indeed beneficial for non-repeated but complementary materials.

In this dissertation, I will investigate the effects of distributing learning time in learning from multiple and single texts. Thereby, I will use the term distributed learning for learning with complementary (text) materials, while the term distributed practice still refers to the “classical” distributed practice of repeated materials.

The current work is grounded on several theoretical assumptions and theories, whose broad treatise lies beyond the scope of the current work. Nevertheless, these assumptions will be summarized shortly. First, learning and reading is perceived as a knowledge generating process, comparable to the assumption underlying the model of generative learning (Wittrock, 2010). Thus, the learner is not perceived as passive recipient of knowledge that has to be stored, but as an active "comprehender" and information processor, who generates his knowledge from information presented during the learning process (e.g. Mayer, 1996). Second, an associative network model of semantic memory is assumed, in which knowledge is stored interconnected (for an overview of models of semantic memory see McRae & Jones, 2013). Furthermore, this interconnected knowledge is retrievable in dependency of its current activation, following the spreading activation account (Anderson, 1983; Collins & Loftus, 1975).

On these theoretical grounds, I will present three models of text comprehension in Chapter I, and will have a look on the factors influencing text comprehension and learning from text. Subsequent to the introduction of models of text comprehension, the concept of desirable difficulties and the underlying cognitive mechanisms will be introduced before I will introduce distributed learning as a potential desirable difficulty in learning from multiple texts and present the research questions of this dissertation.

In Chapter II—IV, I will report five experiments investigating the research questions. In Chapter V, the findings of these experimental studies are summarized and their practical implications are discussed.

Reading and Comprehending Single and Multiple Texts

In the current work, distributed practice and distributed learning are investigated as learning strategies in learning from texts. To understand how distribution might affect learning from texts, it is necessary to have an understanding of *how* text materials are processed in general. Thus, in the following section, I will provide an overview of theoretical accounts explaining the processes underlying reading and learning from texts.

A text in this dissertation is defined according to Alexander & Jetton (2003), where a "text represents the inscription of ideas in linguistic form" (p. 201). This means that texts consist of written or spoken words that build a meaningful coherence. Texts consist of more than one unit (sentence), and the units are related to each other. In the last decades, several models of text comprehension are stated (for an historical overview see Pearson & Cervetti, 2015). One central—if not *the* central—model of text comprehension is the construction-integration model (Kintsch, 1988, 1992, 1994; for an overview of text comprehension models see McNamara & Magliano, 2009). Given its high impact on research and practice, I will first introduce this model and its assumptions about the processes during reading, understanding and learning from text. Considering the scope of this dissertation, I will forgo an introduction of how text propositions are built at all, but will focus on the processes that occur after a single text proposition has been built, thus the processes involved in comprehending texts, not sentences. Subsequently, I will introduce the landscape model of reading (van den Broek et al., 1996). This model is especially noteworthy in the context of the presented experiments, because of its specific assumptions about the fluctuating activation of prior knowledge during reading, which might be a central mechanism for implementing desirable difficulties in learning from texts. In the next step, I will introduce the document's model (Britt et al., 1999, 2013). In this model, assumptions about the mental representation of single texts are transferred to reading and comprehending multiple texts. As the current work focuses on reading of (multiple) textbook chapters, I will continue with an explanation how this model can be applied to reading multiple textbook chapters. Afterwards, I will introduce the distinction between comprehending text and learning from text, before I will present some key factors influencing comprehending text and learning from text.

The Construction-Integration Model

To learn from text, the reader first has to comprehend the text. The construction–integration model makes several assumptions about the processes involved in *comprehension*. In this model, comprehension can only be defined in terms of the situation as whole (Kintsch, 1992). Thus, the model does not explain the comprehension of isolated sentences or text propositions, but the comprehension of the whole text. In this model, it is stated that the reader first builds a text base, which is a linguistic representation of the text. However, for comprehension another representation is necessary: the situation model. In the situation model, the linguistic information of the text is lost, but the information and context given within the text is integrated into "some larger structure" (Kintsch, 1988, p. 163). To reach this larger structure, *inferences* from the text base are built. Thus, the reader builds connections between parts of the text to create an understanding of the text beyond the separated text passages, but also connects the information given within the text with his or her prior knowledge.

Furthermore, the construction – integration model is an interactive theory of comprehension. Following Kintsch (1992), comprehension always needs an interaction between the input, which might be a text, and the recipients goals and knowledge. This interaction is defined by *construction* and *integration* processes. The reader constructs a representation of the text (its word meaning, propositions and inferences) and integrates this representation with the knowledge he or she already has. Kintsch (1988, 1992) assumes that local, associative processes activate knowledge during reading. As this process is seen as associative, combining bottom-up (from the linguistic input) and top-down processes (from the knowledge of the reader), not only relevant knowledge, which fits in the context of the read text, is activated, but irrelevant knowledge as well. Thus, an additional process for the selection of the relevant information is needed to enable the reader to make the right inferences and build a coherent representation from this text base. This process is described as an *integration* process, thus, in this process, the pre-existing knowledge of the reader is integrated with text base to build a coherent representation of the whole.

To comprehend not only a single sentence, but also the whole text, propositions given within the text must also be related to each other. Due to the limited working memory capacity (for recent reseach see Oberauer, Farrell, Jarrold, & Lewandowsky, 2016), the amount of text propositions maintained simultaneously in working memory is restricted. Thus, the question arises how relations between all propositions given within in the text are created. In the construction–integration theory, it is assumed that the reader does not deactivate all propositions in working memory when she or he moves to the next sentence, but maintains some of the propositions (Kintsch, 1992). Those propositions are then reprocessed during further reading, which enables building a relation between sentences. However, maintaining a proposition in working memory is more useful for building a coherent representation of the whole text, if this proposition can serve as bridge between earlier and later sentences in the text. Furthermore, as only some propositions are maintained in working memory while others become deactivated, not all potential connections are built during reading. If this process is not successful in terms of a coherent situation model, the reader could also retrieve prior text propositions from memory. Nevertheless, this process is assumed to be resource demanding (Kintsch, 1992, but see also O'Brien et al., 1990). However, recent research has shown that the ability to automatically activate prior-read information also extends to information from other texts, if this information is needed to comprehend a subsequent text (Beker et al., 2016, 2019). Thus, the process of linking (far off) text propositions might rely more on reactivation processes than the maintaining processes described by Kintsch (1992).

To summarize this model in regard of understanding potential effects of didactical strategies of learning from multiple texts, the model establishes a good basis for the research question how learning from (multiple) texts can be supported by distributed practice. However, the mechanism proposed for the relation of different propositions (via mere associative processes or maintaining central propositions in working memory) might not be sufficient to explain comprehension of whole texts, in which also far off propositions has to be connected. A more recent model of text reading and comprehension is the landscape model of reading. This model builds up on the assumptions of the construction–integration model, but especially concentrates on the process of building coherence of the whole text.

The Landscape Model of Reading

The landscape model of reading (van den Broek et al., 1996) extends the construction–integration model by integrating research of the early 90s, in which the relation between comprehension processes and memory has been investigated.

One central concept in the landscape model of reading is coherence and the standards of coherence of the reading. It is assumed that the reader seeks coherence, thus, the "grammatical and semantical connectedness between sentences that form a text" (Bussmann, 1996, p. 198). As in the construction-integration model, this connectedness is reached by building *inferences*. Those inferences can be based solely' on the text material, thus, bridging two or more information of the text—bridging inferences—, or be based on the prior knowledge of the reader—elaborative inferences (van den Broek et al., 2015). To reach coherence, the reader might rely on several aspects of coherence, for example, van den Broek et al. (1996) focus on anaphoric clarity and causal explanation. It is assumed that anaphoric clarity and causal explanation can be reached by the processes within the construction-integration model. Although, van den Broek et al. (1996) discuss the three potential sources of coherence assumed by Kintsch (1988, 1992) they conclude that two of the sources—notably the mere associative process as well as the process in which the central propositions of the previous sentence are maintained in working memory during reading the subsequent text—might not be sufficient for building coherence of the whole text. The third source however—the reinstatement (or reactivation) of propositions of previous texts—might be especially important for reading comprehension. Van den Broek et al. (1996) consequently focus on this process.

They assume that the reader can use two further sources to reach coherence, if the coherence is not sufficiently build by the first two sources stated in the construction – integration model. First, the reader can reactivate information given earlier in the text. Second, the reader can use acquired background knowledge to fill conceptual gaps. Within this model, a two-stage model of knowledge activation is stated. The first stage is described as a quick memory search of related information. This is seen as cohort activation, a passive activation of prior knowledge and/or information given previously in the text. The second stage is described as a slower, more elaborated process. Here, the reader searches actively for the relevant knowledge (again, either represented as prior knowledge or as knowledge presented previously in the text).

Furthermore, it is assumed that the activation of knowledge and the reactivation of information of previous sentences is coherence-based: If activation on the first stage is sufficient to build a coherent representation (depending on the current standards of coherence of the reader), no further activation processes are initiated by the reader, but if this process fails to build a coherent representation, an active second stage memory search is undertaken.

Additionally, the landscape model of reading states that the activation of a concept fluctuates during reading a text. Thus, a concept is activated if the current sentence or proposition includes this concept, but is deactivated, if the current sentence does not include this concept. However, if a later sentence or proposition includes this concept again, the concept is reactivated. Additionally it is stated that the concepts, which are often reactivated during reading are more likely to be included of the memory representation of the text. Furthermore, concepts that are often activated simultaneously are more likely to be represented as interconnected nodes (van den Broek et al., 1996).

Taken together, the landscape model of reading gives a good insight in the online processes during text reading. It states several assumptions about the activation from memory and previously read text during reading and comprehending tests, which are also empirically supported (Linderholm et al., 2004; van den Broek et al., 1996). Especially the active processes of activating and reactivation knowledge might be interesting for implementing distributed practice in learning from text, as will be presented below.

The Document's Model

However, in the work presented here, potential effects of distribution will not only be discussed in the light of learning from single texts but also in regard to learning from multiple texts. Reading multiple texts instead of one single text provides additional demands on the reader. Instead of building only the situation model of the text, the reader has to build a model which integrates information from both texts and furthermore a representation of information about the different text sources. Thus, the reader of multiple texts builds two models: a situation model and an intertext model, which are summarized as document's model (Britt et al., 1999, 2013). Britt and colleagues (1999, 2013) assume that the document's model represents the content of the text (as the situation model for single texts) and further includes nodes about the source of information, thus, about the text the content is derived from. Those nodes then may include information about the documents characteristics (in the case of the current experiments, the textbook chapter), the author of the source (which might be less relevant or even unknown in case of text book chapters) and also an evaluation of the source (in case of text book chapters, the reader might evaluate the text as highly credible). This information is also linked to content information ("In the second textbook chapter, it is told that the bacterial cells are the simplest life forms on earth."). Furthermore, the document's model also includes links between the different texts, in which the relation between the content of the texts is represented. Thus, the reader may have represented a "opposes to" relation, if the texts present conflicting information, or a "complements" relation, if the content of one texts complements and expands the content of the other text(s). However, multiple texts that do not provide conflicting, but complementary contents seem to be a special case of multiple texts.

Complementary texts are multiple texts which are "convergent and require adding pieces of information together" (Primor & Katzir, 2018 p. 4). In the research of multiple texts, those texts have been investigated far less often than multiple texts that included a contradiction. Barzilai, Zohar and Mor-Hagani (2018) reported in their recent review, that only 4.9 % of the included experiments used complementary texts. This also comes with less theoretical framework for complementary texts, as most models – including the document's model - of multiple text processing focus on multiple texts that include discrepant or contradictory information (Strømsø, 2017). For example, in the research underlying the document's model, history texts are used as text materials. In history texts, it is likely that multiple texts deliver conflicting information, as one author may report other reasons for historical events than another author. In this case, the intertext model becomes more important to build a coherent representation of the text – the reader has to carefully evaluate the sources to dissolve this discrepancies. However, Strømsø (2017) assumes that in case of complementary, overlapping documents, the readers need to build an intertext model will be reduced. This makes especially sense in case of reading subsequent textbook chapters, in which the source remains the same, the author (if different at all) seems to be less relevant and the structure is often held constant between the chapters. Thus, in those cases of reading multiple textbook chapters, it seems more likely that new information will be, as Strømsø (2017) frames it, "seamlessly" adapted to previously processed information. Thus, if the reader is confronted with two textbook chapters (e.g. one about the plant cell and one about the bacterial cell), the knowledge representation of the content of the first text will be extended by the content of the second text. For example, if one content node from the first chapter would be "The plant cell has a cell nucleus", this node would be extended to "*Contrary to the bacterial cell*, the plant cell has a cell nucleus". Consequently, the further information from which text the content stems is not necessary, as the content can be easily summarized in just one node. Thus, it can be assumed that in textbook reading, the information about the source is ignored, and one coherent situation model including content from all read chapters is build. Therefore, in the special case of complementary texts, it could be assumed that these multiple texts are processed as if the reading of a single text would be interrupted (as for example by a pause) and the reader has no need to build a documents model but just one coherent situation model. In this dissertation, I will focus on learning from those complementary, multiple texts. However, this texts seem to have more in common with reading single than with reading multiple text and therefore, in the following, I will focus on presenting research regarding reading and learning from single texts.

Learning from Texts

After this overview of important models of text reading, I will focus on the question how readers learn from texts. First of all, understanding a text, thus, building a coherent situation model is seen as a necessary, although not sufficient condition for learning from text (Beker et al., 2017). However, what are the differences between comprehending and learning? Following Beker et al. (2017), comprehending texts refers to the processes during reading, whereas learning is a long-lasting change in long-term memory as consequence of text reading. Thus, the mental representation built during reading (e.g. in a situation or documents model) has to be retrievable at a later point of time. Consequently, to investigate learning from text, it is not sufficient to measure text comprehension immediately after reading, but also after some time has passed. This distinction is also similar to the distinction between performance and learning (Soderstrom & Bjork, 2015), which will be encountered in the section about desirable difficulties. Thus, the time of measurement seems to be important for the distinction between text comprehension and learning.

Furthermore, text representation should be distinguished from knowledge representation (Beker et al., 2017). Although the concepts of text representations and knowledge as networks are quite similar, Beker et al. (2017) argue that they differ in permanence and contextualization. While text representation can be seen as highly contextualized, knowledge representations are decontextualized. Consequently, to learn from a text, the text representation has to be integrated in a decontextualized knowledge representation. However, reading comprehension and learning from text are closely related, as comprehension is needed for learning from text. Consequently, factors influencing reading comprehension can be assumed to influence learning from texts as well. In the next section, I will introduce some important factors that influence reading comprehension, learning from text, or both. It should be noted, however that there are many more factors which influence reading comprehension, as for example motivation (e.g. Schaffner & Schiefele, 2013), or goals of reading (e.g. van den Broek et al., 2001), that will not be discussed here (for an overview of reader characteristics see van den Broek & Espin, 2012).

Effects of Prior Knowledge

In the presentation of models of text comprehension, it became clear that knowledge the reader has *before* reading is strongly related to comprehending text: to build a coherent situation model, the reader has to relate the text content to appropriate knowledge contents (Kintsch, 1988, 1992; van den Broek et al., 1996; McNamara & Magliano, 2009). For example, Kintsch (1988) stated that "the process of constructing a discourse representation relies heavily on knowledge" (p.164). This knowledge can be termed as *prior knowledge*. If this knowledge strongly relates to one domain (e.g. cell biology), it is referred to as *domain-specific* prior knowledge (Simonsmeier et al., 2021).

The amount of prior knowledge is positively associated with text comprehension after reading single (Ozuru et al., 2009; Priebe et al., 2012; Recht & Leslie, 1988; for the effects of the quality of prior knowledge see Kendeou & van den Broek, 2007) and multiple texts (Bigot & Rouet, 2007). Furthermore, a high amount of prior knowledge may even compensate for lower reading and cognitive abilities (O'Reilly & McNamara, 2007; Ozuru et al., 2009; Priebe et al., 2012; Recht & Leslie, 1988; Schneider et al., 1989; Voss & Silfies, 1996). Additionally, providing background information before reading a text also facilitates text recall (Rawson & Kintsch, 2002).

It can further be assumed that prior knowledge is positively associated with learning from texts. The amount of prior knowledge seems to influences readers inference building. More bridging inferences were found for readers with higher vocabulary knowledge (Dixon et al., 1988; Singer et al., 1992). Predictive inferences become less likely with lower prior knowledge (Calvo et al., 2003) and readers with lower prior knowledge generate less global inferences (Fincher-Kiefer, 1992). Given the assumption that those inferences are needed to decontextualize an information from the text, it can be derived that readers with high prior knowledge learn more from text than reader with low prior knowledge. Empirical evidence for this assumption can be derived from the experiments of McNamara and Kintsch (1996; Experiment 2) and Schneider et al. (1989), who measured text comprehension with some delay and also found a positive association with prior knowledge. Considering the time-dependent distinction between text comprehension and learning (Beker et al., 2016), it can be assumed that these results reflect learning from text despite mere text comprehension. Combined with the more general finding, that the amount of prior knowledge predicts learning outcomes (e.g. van Kesteren et al., 2014), it can be assumed that prior knowledge is associated with both, text comprehension and learning from texts.

Furthermore, the effects of text characteristics as for example cohesion on text comprehension even depend on prior knowledge (e.g. the reverse coherence effect, see McNamara, 2001; McNamara et al., 1996; McNamara & Kintsch, 1996; O'reilly & McNamara, 2007). Thus, similar to the expertise reversal effect (Kalyuga et al., 2003), the expertise, thus, the prior knowledge within the domain of the text, affects which text characteristics are beneficial for the reader.

Effects of Reading Ability and Working Memory

The ability to read and comprehend texts develops trough reading experience and depends on several learner characteristics, such as the ability to make inferences, comprehension monitoring, working memory ability, and word knowledge (Perfetti et al., 2005). In learning to read, readers first focus on decoding processes, thus, reconstruction the text base, before they reach the ability to build a situation model.

Evidently, reading ability and reading strategies are relevant for learning with written text materials. Readers struggling with basic reading processes, such as decoding words and grasping the meaning of sentences, need to invest cognitive resources in these processes, with the consequence that fewer resources are available for higher-level learning processes such as reading strategies (Cain et al., 2004). Furthermore, the processes undertaken during reading are assumed to take place in working memory (de Bruïne et al., 2021). Referring to Daneman and Carpenter (1980), working memory is a combination of a storage and a processing component, which sum is limited in capacity (but see Cowan, 2017, for a discussion of several definitions). Daneman and Carpenter (1980) developed the reading and listening span task, for which completion both components are needed, the storage and the processes working on the storage. With these measures, they investigated the interplay between reading and working memory and found a significant correlation between both. This is theoretically explained by the above-mentioned idea that the processes underlying reading take place in working memory and consequently, the individual capacity of working memory does restrict the ability of reading. However, Peng et al. (2018) showed in their meta-analysis that working memory is only moderately associated with reading comprehension (Peng et al., 2018) and the association is even reduced when other factors are included (such as intelligence, see Van Dyke et al., 2014).

Especially for skilled readers, reading seems to be effortless and therefor independent from working memory. This might be explained by the passive activation processes assumed in the two-stages model of knowledge activation in the landscape model of reading (van den Broek et al., 1996). When reading easy texts or texts that meet the knowledge of the reader, the reader has no need for the active memory search. Nevertheless, even in easy texts, the reader might stumble over discrepancies in the text. Such discrepancies or contradictions are used to investigate whether readers seek for global coherence. A well-known example is the story about Mary, which visits a restaurant with a friend. Mary is introduced either as fast food lover or as vegetarian. In the following text passage, Mary orders a cheeseburger. The participants showed longer reading times, if Mary, the vegetarian, orders a cheeseburger (Albrecht & O'Brien, 1993). This inconsistency effect indicates that the reader indeed stumbles over this discrepancy and needs more time to dissolve it. It can be assumed that this process depends on metacognitive monitoring. In a recent study by de Bruïne et al. (2021), they investigated if coherence monitoring depends on working memory capacity. They found that the inconsistency effect was reduced, if the reader has to engage in a dual-task. Therefore, the metacognitive coherence monitoring can be seen as effortful and dependent from working memory capacity. Consequently, it can be argued that the implementation of reading strategies might also be effortful. For less-skilled readers, who still are challenged by reading and understanding the text and are thus more dependent of their working memory capacity, this might indicate that reading strategies and metacognitive coherence monitoring may fail.

Effects of Standards of Coherence

Additional to prior knowledge and the (reading) abilities of the reader, the standards of coherence of a reader also influences how much she or he learns from a text (van den Broek et al., 2011).Within the landscape model of reading, it is assumed that reading includes passive processes (cohort activation) as well as more strategic or active processes. Consequently, the question arises when readers use which process. As descripted above, van den Broek et al. (1996) assume that the latter process is coherence-based, thus, it depends on the amount of coherence reached, whether the reader uses more strategic activation processes. Furthermore, van den Broek et al. (2011) assume that readers differ in the amount of coherence they seek for, thus, in their *standards of coherence*. The standards of coherence are then seen as benchmarks for strategic processes: if the standard is fulfilled with passive cohort activation, no more strategic activation processes are stimulated, if not, the reader engages in more strategic activation processes (van den Broek et al., 2011). However, what influences the standards of coherence of a reader? First of all, it is assumed that the standards of coherence are a function of the reader, the text and the reading situation (van den Broek et al., 2011). Thus, the very same reader has different standards of coherence depending on reading a novel or a textbook chapter or reading for entertainment or learning. Furthermore, van den Broek et al. (2011) state that the reader might not seek for maximal coherence, but for sufficient coherence. Additionally, the standards of coherence are influenced by many factors, as for example the reading goal, which seems to influence the inferences readers make during reading (Narvaez et al., 1999; van den Broek et al., 2001).

The standards of coherence are likely to influence learning from texts. To learn from a text, a coherent representation of the text content, embedded in the prior knowledge structure, is needed. However, this representation strongly depends on which benchmark of coherence the reader aims to reach – if the reader has low standards of coherence, it is likely that the representation of the text content might include less interconnections between the text content and prior knowledge. Conform to this assumption, van den Broek et al. (2001) found a significant benefit for students in the experimental group, which had higher standards of coherence provoked by the reading goal. However, they measured text recall quite immediately after reading, while Narvaez et al. (1999) did not find an effect of a learning related reading goal on learning measured one week after reading. Thus, it seems reasonable that the standards of coherence influence comprehension, but it is still an open research question if they influence learning from text.

Effects of Text Characteristics

Despite these characteristics of the reader, text characteristics might also determine the readers text comprehension and learning from text. For example, text might differ in the syntactic structure, the word frequency of the words used within text, text organization, usage of examples, et cetera and might render the text to be more or less *difficult* to read.

Easy and difficult texts might be labeled as “good” and “bad texts”. For example, the studies by McNamara and colleagues are guided by the question whether good texts are always better (McNamara, 2001; McNamara et al., 1996; McNamara & Kintsch, 1996; O’reilly & McNamara, 2007; Ozuru et al., 2009). “Good texts” are texts with high local and global cohesion, thus, texts which incorporate cues within the single sentences as well as between the sentences that guides the reader to connect the different propositions or ideas without the need to make bridging inferences (Crossley et al., 2016). Please note at this point that in the early work regarding the effects of "good texts", McNamara and colleagues (McNamara, 2001; McNamara et al., 1996; McNamara & Kintsch, 1996) used the term "coherence" instead of cohesion. However, as coherence is defined as the subjective making sense of the text (Wright, 2018), it is not a characteristic of the text but of the interplay between a text and reader's world knowledge and has to be distinguished from cohesion. A text can be cohesive without making sense (to the reader), thus, without being coherent, and vice versa (Wright, 2018). What McNamara and colleagues manipulated was the text cohesion, thus, the cues within the text for example with word repetitions or references that are more concrete.

The results regarding the effects of cohesive texts on text comprehension are somewhat mixed, as they seem to depend on the assessment of text comprehension, which was measured for example with free recall, or a sorting task which measure concept change (thus, the situation model), or questions, which could be text-based, problem solving or ask for bridging inferences. However, while the experiments of McNamara et al. (1996) and McNamara and Kintsch (1996) found better free recall for high-cohesive texts, Boscolo and Mason (2003) found no effect of cohesion on text recall, but better performance in bridging inference questions after reading high cohesion texts. Furthermore, the effects of cohesion do not only depend on prior knowledge, with the above mentioned finding that low-cohesive texts are beneficial for high-knowledge readers (non-significant "tendencies" in free recall, problem solving and bridging inference questions, Boscolo & Mason, 2003; problem-solving and bridging inference questions, McNamara et al., 1996; sorting task, McNamara & Kintsch, 1996). Furthermore, this interaction effect seems to depend on reading skill, as skilled readers seem to profit equally from low- and high-cohesive texts, whereas low-skilled readers with high knowledge profit from low-cohesion texts, but only for text-based questions (O'reilly & McNamara, 2007; Ozuru et al., 2009). Therefore, it can be summarized that text cohesion seems to affect text comprehension somehow, but to what extent depends on how text comprehension is measured and the reader's knowledge and reading ability.

Other text characteristics, as syntactic structure and word frequency (Feng et al., 2013) or presentation format, text organization and example content (McCrudden et al., 2004) were used to implement a higher difficulty in text and found no or small negative effects on text comprehension. Even Mills et al. (2015), who manipulated the text difficulty by varying sentence length, word frequency, verb cohesion, and syntactic complexity simultaneously, found no effects of text difficulty on text comprehension measured with multiple-choice retention and transfer questions.

More artificial manipulations of text difficulty were used by McDaniel et al. (2002). In one text condition, they scrambled the sentence order, in another text condition, they deleted letter from text. Sentence unscrambling increased free recall in reading expository texts, but not in fairy tales. I framed this manipulation of text difficulty undertaken by McDaniel et al. (2002) as artificial, because these manipulations do not refer to characteristics that vary naturally (as for example syntactic structure). Thus, McDaniel et al. (2002) did not investigate how natural differences between texts affect text comprehension, but a didactical strategy. Another didactical strategy, which was implemented in text reading, is perceptual disfluency. Adding perceptual disfluency means adding text features, which disturb the reading process, thus, lead to less fluent reading, as for example hard to read fonts. Disfluency is thought to lead to deeper processing and consequently to more learning (Yue et al., 2012). However, the effects of disfluency are mixed. Despite Yue et al. (2012) found beneficial effects of disfluency on text comprehension, this finding was seldom replicated (for a short review see Weissgerber & Reinhard, 2017). Weissgerber and Reinhard (2017) investigated disfluency with hard-to-read fonts and unscrambled sentences and found that an interaction between reading condition and time of test with greater benefits of reading hard-to-read fonts at the delayed test.

Furthermore, also the presentation format might affect text comprehension. For example, text comprehension was found to be lower in reading on computer screen than on paper (Hou et al., 2017; Mangen et al., 2013). McCrudden et al. (2004) found that sentence-by-sentence reading lead to lower text comprehension than whole-text presentation.

In a very different approach, Toyama (2019) investigated the contributions of reader, text and question to reading comprehension with an explanatory item response model with data of over 10 000 readers. She found that text features explained over half of the variance in item difficulty. Especially sentence length, word frequency and syntactic simplicity (and temporality) had the greatest effect on text comprehension, whereas the effects of cohesion were small, or not significant as for logical and referential cohesion. Regarding the assessment of text comprehension, the analysis showed that especially the abstractness of information requested by the questions increased item difficulty. Furthermore, and contrary to common considerations, literal recall questions were not easier, but more difficult than restructuring and bridging inference questions. On the reader's part, Toyama (2019) only included vocabulary knowledge and found that more knowledge is associated with lower question difficulty.

To summarize the effects of text characteristics on text comprehension, it can be stated, that it is quite difficult to reach a consistent picture. Some text characteristics seem to influence text comprehension, but the effects depend on reading ability and prior knowledge and also text comprehension measure. This might reflect how complex reading processes are, as they are a product of the interaction between reader and text, but also the assessment of text comprehension.

The evaluation of the effects of text characteristics becomes even more complicated when it comes to learning from text as defined as long-term change of memory measured by delayed tests. Learning from texts in dependency of text characteristics seems to be seldom investigated. From the above cited studies only Weissgerber and Reinhardt (2017) and McNamara and Kintsch (1996) implemented a delayed test. McNamara and Kintsch (1996, Experiment 2) found a greater increase between the immediate and delayed test in conceptual structure for low-cohesive texts, but a higher decrease in accuracy in question answering for this texts. Thus, these results also reflect the inconsistency found for reading comprehension. However, Wylie and McGuinness (2004) investigated the effects of 5 different text structures on text comprehension and learning. The participants were asked to recall the text content immediately after reading and two weeks later. Interestingly, despite one text structure that lead to the worst recall immediately after reading and two weeks later, the order of recall performance changed between immediate and delayed recall. Therefore, text structures, which lead to low performance immediately after reading lead to high recall two weeks later and vice versa. Weissgerber and Reinhardt (2017) similarly found that the time of test matters: While they did not find a disfluency effect for a hard-to-read font immediately after reading, they found beneficial effects of this font two weeks later.

Even if the effects of text characteristics on text comprehension and learning from texts are somewhat inconsistent and reflect the multi-faceted nature of text reading, it can be concluded that "good texts" in terms of "easier" texts are not necessarily better for learning from texts. Furthermore, the effects of text characteristics seem to depend of time of test.

Desirable Difficulties in Learning from Texts

This conclusion reminds of a group of learning strategies summarized under the term desirable difficulties. In the research of learning, one central goal might be to optimize the learning process in a way that the learning results become more long lasting. One account can be to facilitate the learning process, thus, enhance the *performance* during learning. However, those strategies do not necessarily lead to better *learning* in the long run. On the contrary, making the learning process easier might even lead to lower learning, and inducing more difficulties in learning might enhance learning in the long run (R. A. Bjork, 1994). Thus, some difficulties in the learning process are seen as *desirable*, as they make the learning process subjectively more difficult and may even lead to lower performance, but foster learning in the long-run (E. L. Bjork & Bjork, 2011; R. A. Bjork, 1994). Learning procedures that share such properties can also be summarized under the term *desirable difficulties* (Lipowsky et al., 2015).

Encoding and Retrieval

According to R. A. Bjork (1994), effective learning grounds on two aspects: encoding and retrieval. Regarding encoding, learners should be forced to multiple encoding, thus, they should encode the same information in relation to different concepts and ideas: If concept A is first encoded with concept B, next time it should be encoded with concept C. This might also be compared to building a situation model in text reading: The learner should not only encode the single information (respective sentence) isolated, but encode the information according to related concepts in a broader mental representation. Regarding retrieval, R. A. Bjork (1975, 1994) assumes that the retrieval from memory is itself a *memory modifier*, as he frames it. Retrieved items profit from the retrieval and become more retrievable at later occasions. Consequently, he assumes that practicing retrieval from memory during learning does benefit learning.

In their new theory of disuse, R. A. Bjork and E. L. Bjork (1992) distinguish between storage strength and retrieval strength. Storage strength refers to how good an item is learned, whereas retrieval strength refers retrievability of an item given a specific retrieval cue. It is further assumed that storage strength has no effect on the actual performance, which depends only on the retrieval strength. Storage strength is assumed to be accumulated in a monotonic function and an item that has reached a storage strength does not lose this storage strength over time. Therefore, storage strength is unlimited. In contrast, retrieval strength is assumed to decrease over time and with further learning. With more items added to memory, the retrieval strength of items already in memory decrease. Furthermore, R.A. Bjork and E.L. Bjork (1992) assume that the accumulation of storage strength depends on the retrieval strength: With high retrieval strength, the accumulation of storage strength is retarded. Additionally, they assume that studying and retrieval can increase storage and retrieval strength, but retrieval is more potent. Following R.A. Bjork (1975), retrieval can be made from short or from long-term memory, which is described as “depth of retrieval”. Retrieval from short-term memory is assumed to have little or no effect on later efforts to retrieve the item from long-term memory, whereas retrieval from long-term memory is assumed to facilitate later retrieval. Furthermore, in the retrieval-effort hypotheses, it is assumed that more difficult (successful) retrieval enhances learning (Pyc & Rawson, 2009). The greatest effect on subsequent retrieval strength should have a more difficult retrieval, thus, retrieval of items with low retrieval strength, of items that have a high storage strength.

The New Theory of Disuse and Learning from Texts

From this theoretical framework, three implications for learning from texts can be deduced. First, performance and learning has to be distinguished. As the performance during learning does depend on the retrieval strength, which is less stable and more situation dependent than storage strength, it is not a reliable measure of learning. Consequently, Bjork and Bjork (2011; but see also Soderstrom & Bjork, 2015), argue that performance and learning have to be distinguished. This distinction resembles the distinction between text comprehension and learning from texts. Second, it can be deducted that the goal of learning from texts should be high storage strength as well as high retrieval strength. Storage strength is necessary to integrate new information in memory at all, but this memorized information is not useful, if it cannot be retrieved later. Third, learning from texts can be assumed to be more beneficial, the lower the retrieval strength and the higher the storage strength is during reading.

Text Characteristics as Desirable Difficulties

Thus, as the text characteristics described above are also framed as more or less difficulty, it could be asked if they might also be desirable difficulties and benefit learning. However, difficulties in text reading might affect different aspects of reading. For example, disfluency is assumed to encourage the reader to engage in “deeper processing” and Weissgerber and Reinhardt postulate that disfluency might be a desirable difficulty. However, if the beneficial effects on learning only define desirable difficulties, the definition is circular. I would rather postulate that desirable difficulties must be defined as desirable by the processes they are assumed to evoke: “Desirable difficulties […] are desirable because they trigger encoding and retrieval processes that support learning, comprehension, and remembering” (R. A. Bjork & Bjork, 2020, p. 3). Therefore, to define a text characteristic as a “desirable” difficulty, the underlying processes should be taken into account. As retrieval is seen as memory modifier (R.A. Bjork, 1975, 1994), a text characteristic, which evoke retrieval processes might be a good candidate for operating as desirable difficulty.

Given this approach to desirable difficulties, for example the manipulation of text cohesion might indeed be considered a desirable difficulty. Even if the findings of McNamara and colleagues are somewhat inconsistent, the derived effects correspond well to this idea (McNamara, 2001; McNamara et al., 1996; McNamara & Kintsch, 1996; O’reilly & McNamara, 2007; Ozuru et al., 2009). In low cohesive texts, the reader has to retrieve information from long-term memory to comprehend the text. Additionally, the retrieval strength can be assumed lower in less cohesive texts. The retrieval strength is lowered by less retrieval cues in the text. Reader with high prior knowledge are able to retrieve this knowledge during reading low cohesive text and to build a coherent representation of the text. Thus, it can be assumed that low cohesive texts could possibly act as a desirable difficulty: The reader is forced to retrieve information from long-term memory during reading and this retrieval is more difficult as less retrieval cues are presented. However, in the research of McNamara and colleagues text comprehension was mostly measured immediately after reading. As desirable difficulties might play out their advantages especially in the long run, more research is necessary to investigate whether lower cohesion operates as desirable difficulty.

Nevertheless, text characteristics in general and text cohesion in particular yield important disadvantages as didactical strategies. First and apparent, they depend to a great extent on prior knowledge and reading ability of the reader. The subjectively perceived text cohesion might differ between readers in dependency of prior knowledge: For some readers a reference might be very cohesive, for others not at all. For readers with low prior knowledge, the low cohesion text are not beneficial, because they are not able to retrieve the necessary information to build a coherent situation model. Therefore, when text cohesion is used as didactical strategy, the reader's ability must always be taken in account. However, given the complex process of reading, this might be the case for many didactical strategies that are instated in text reading. Second, and even more detrimental, it is complex to determine text cohesion, even if there are tools to assess text cohesion for English texts as Coh-Metrix (McNamara & Graesser, 2012) or TAACO (Crossley et al., 2016), and it is difficult to implement different levels of cohesion in teaching or self-regulated learning. Therefore, even if it might be theoretical fruitful to investigate text cohesion (or other text characteristics) as desirable difficulty, the practical utility is questionable.

Learning Strategies as Desirable Difficulties

Given these disadvantages, didactical strategies that can be implemented in text reading without the need to change the texts itself might be more fruitful to enhance learning from texts. Four learning strategies (applied didactical or in self-regulated learning) are well known and described as desirable difficulties: Interleaving, generation, tests as learning events and distributed practice (E. L. Bjork & Bjork, 2011; Lipowsky et al., 2015). In this dissertation, I will focus on the distributed practice effect. Nevertheless, I will shortly introduce the other three central desirable difficulties and their benefits for learning from texts.

The *generation effect* means the beneficial effect of generating a solution compared to be presented to that solution (E. L. Bjork & Bjork, 2011; Slamecka & Graf, 1978). Generation effects in general seem to be reliable (McCurdy et al., 2020). Furthermore, generation seems also beneficial when implemented in reading expository and narrative texts (Schindler & Richter, 2021). In learning from text, generation can be implemented by letter deletion or scrambled sentences (as for example in the experiment of McDaniel et al., 2002). However, one might even consider that inference generation itself can be seen as implemented generation. However, the adding of more generative elements in text reading is also effortful, as the text itself has to be changed to profit from the generation effect.

The beneficial effects of *testing as learning event* despite relearning are well investigated and reliable, also for prose (Rowland, 2014). It seems to be easy to implement testing in learning from text: after reading, the reader could be asked to answer some questions and the learning outcome could be increased. Testing was shown as more beneficial for learning with expository texts compared to rereading (Dirkx et al., 2014) and concept mapping (Karpicke & Blunt, 2011; Lechuga et al., 2015). However, in the last years the beneficial effects of testing for complex materials have been discussed (van Gog & Sweller, 2015), but it can be assumed that testing is also beneficial for complex materials as for example text materials (Karpicke & Aue, 2015).

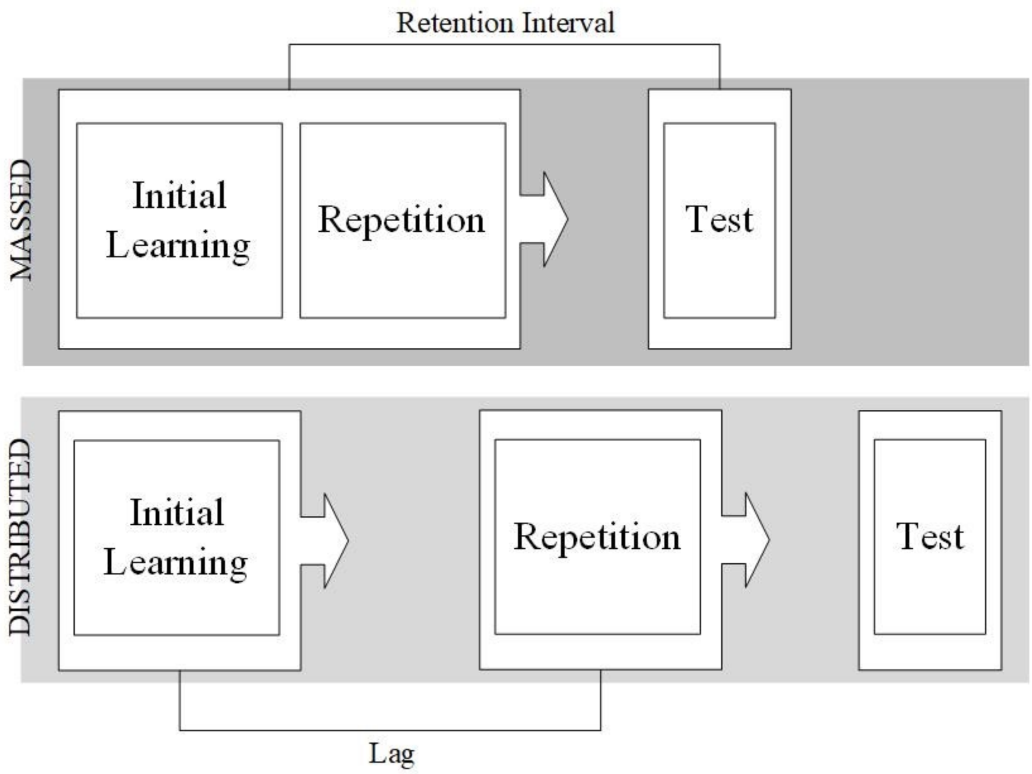
In case of *interleaving*, thus, interleaving separate topics in learning despite blocking the topics, the results regarding learning from texts are less encouraging. Brunmair and Richter (2019) found in their meta-analysis no significant effect of interleaving for expository texts, whereas they found positive effects especially for visual learning materials as for example paintings. Therefore, it can be questioned whether interleaving is beneficial in learning from text.

The Distributed Practice Effect

The distributed practice effect is quite a famous effect with a long research tradition and goes back to Ebbinghaus (1885). In distributed practice, learning occasions are not massed but distributed over time, thus, despite one long learning session, the learning is divided into at least two shorter learning sessions with some time in-between (e.g. E. L. Bjork & Bjork, 2011). In Figure 1.1, a typical study procedure of distributed practice is shown. In the massed condition, the first learning and the repeated learning occurs within one session, thus, immediately after initial learning, the repetition is presented. In the distributed condition, the repetition of learning materials occurs in a separated session. The time between initial learning and repetition is called *lag*, whereas the time between the last (single) learning session and the final test is called *retention interval*. Therefore, the lag in the massed condition is zero, whereas the lag in the distributed condition in Figure 1.1 is not. The retention interval for both experimental groups is identical.

Figure 1.1

Typical Study Procedure of Distributed Practice



It should be noted that the term distributed practice summarizes two different effects: the spacing and the lag effect (Cepeda et al., 2006; Küpper-Tetzel, 2014). Whereas the spacing effect refers to the "simple" effect that spaced repetition is always better than massed repetition, the lag effect refers to the differential effects of different lags. Regarding the lag effect, an interaction effect between retention interval and lag was found. A specific lag leads to best results at a specific retention interval. If this relationship has to be stripped-down, it can be summarizes as the longer the lag the longer the retention interval should be (Cepeda et al., 2008). Following Küpper-Tetzel (2014), I will also refer to both effects with the term distributed practice, however it should be noted that the effect that is investigated here is more of a spacing spacing than a lag effect, because lag and retention interval are not manipulated simultaneously. Nevertheless, in the interpretation of the effects of distributed practice the specific relationship between lag and retention interval has to be taken into account.

The beneficial effects of spacing can be seen as "one of the most general and robust effects from across the entire history of experimental research on learning and memory" (E. L. Bjork & Bjork, 2011, p. 59). And indeed, the beneficial effects of spacing are found to be robust for verbal recall tasks (Cepeda et al., 2006). Even for retention intervals of less than a minute and thus immediately after learning, retention of learned information was improved by 9% when learning materials were spaced (Cepeda et al., 2006). Although distributed practice has been more often investigated in the laboratory, it seems also to be beneficial in educational contexts and for younger learners (Carpenter et al., 2012; Küpper-Tetzel, 2014). Distributed practice effects have been shown with a broad span of materials and in real learning scenarios, as for example for mathematical practice in school with fourth and seventh graders (Barzagar Nazari & Ebersbach, 2019) or in inquiry science learning (Svihla et al., 2018).

Donovan and Radosevich (1999) found in their meta-analysis a mean effect size of d = 0.46 (95% CI [0.42-, 0.50]. In their analysis, not only verbal recall task but also voice recognition, word processing task and motoric tasks as typing and ladder climbing were included. The greatest effects were found for less cognitive demanding and complex but highly physical tasks as typing (d = 1.22, 95 % CI [0.88, 1.06]), whereas very weak effects of spacing were found for physical tasks that are complex and cognitive demanding as air traffic controller (d = 0.07, 95% CI [-0.05, 0.18]). Further analysis showed that the over-all complexity of task reduces the effectiveness of spacing. However, the complexity of tasks was classified by a rating of graduate and undergraduate students; therefore, it is not clear what makes a task more or less complex.

Despite the beneficial effects of distributed practice being such a robust finding, the meta-cognitive judgement of its effectiveness seems to be low. Therefore, people seems to prefer massed over distributed practice (for a short review see Son & Simon, 2012). To assess the meta-cognitive judgement of distributed and massed practice, participants are asked to estimate the proportion of items they will be able to recall in a test. For example, Kornell (2009) investigated distributed practice in learning with flashcards and asked for this meta-cognitive judgement immediately after learning. For items that were learned massed, the participants estimated higher recall success than for items that were learned distributed, although they actually recalled the distributed learned items better. This mistake in the meta-cognitive judgement of effectiveness might be explained by the discrepancy between performance and learning noted above. The performance in desirable difficulties is impaired by the introduced difficulty, but the learning is improved. The impaired performance might lead to a lower perceived fluency in learning. This fluency is then misinterpret as learning (Alter & Oppenheimer, 2009; R. A. Bjork et al., 2013). However, the evaluation on the basis of performance, thus, the fluency during learning, might affect the usage of distributed practice. If students perceived distributed practice as less efficient, they might choose to mass practice and then miss the well-proven advantages of distributed practice.

To summarize, distributed practice is a robust finding, that also benefits learning in real learning scenarios and younger learners (Carpenter et al., 2012; Küpper-Tetzel, 2014), despite its effectiveness is not always perceived during learning.

Distributing the Reading of Texts

To enhance learning from text, one should first ask how texts are used to learn – for example for an exam. Hartwig and Dunlosky (2012, but see also Gagnon & Cormier, 2018) asked students which learning strategies they use in exam preparation and found that 77% reread at least parts of textbook chapters. Consequently, distributed rereading might indeed be a potent strategy, as it is easy to implement and rereading is already used in learning for exams.

Actually, distributed practice effects have also been shown for learning from expository texts (Glover & Corkill, 1987; Krug et al., 1990; Rawson, 2012; Rawson & Kintsch, 2005; Verkoeijen et al., 2008). For example, in Experiment 1 of Rawson and Kintsch (2005), the participants read a text about carbon sequestration once or twice. Participants, who read the text twice, were randomly assigned to a massed or a distributed condition: Participants in the massed condition reread the text immediately after initial reading; participants in the distributed condition reread the text one week after first reading. Immediately or two days after rereading, the participants received a free recall test, in which they were asked everything they remember from the second section of the text and additionally answered 12 short answer-questions. In both tasks, they found that the effect of distributed rereading depends on time of test: Immediately after rereading, the massed condition outperformed the distributed condition, whereas tested two days later, the opposite was found and the distributed condition outperformed the massed condition.

Furthermore, in the delayed test, massed rereading was not beneficial compared to reading the text only once. In Experiment 2, they replicated this finding with different text materials and recall tasks. Therefore, distributed practice also seems to be beneficial for reading expository texts. However, despite Cepeda et al. (2006) found beneficial effects for spacing even for very short retention intervals, distributed rereading was only beneficial in the test two days later. Moreover, even detrimental effects were found in the immediate test. It could be discussed if this can be explained by the nature of repetitions in rereading. In experiments with shorter texts, benefits of distributed rereading were found even immediately after rereading (Glover & Corkill, 1987; Krug et al., 1990). Therefore, the length of the text might influence the interaction between time and text. This might be because of the nature of repetition in rereading texts. When relearning paired associates, as for example vocabulary learning, the lag in the massed condition is indeed near zero: The learner reads the word pair and starts rereading immediately after. However, in text reading, the lag in the massed condition is not zero. The information given within each sentence is distributed with a lag corresponding to the reading time for the entire text. Consequently, Rawson (2012) describes the "massed" condition contrary to Rawson and Kintsch (2005) as short-lag condition. However, even if I agree that the lag in massed rereading is different from the lag in massed presentation of paired associates, I would argue that the term "short –lag" is imprecise. The term refers to an unknown comparison (shorter than what?), whereas "massed" is clearly defined as the shortest possible lag, even if the exact time of the lag differs between learning materials. For example, this "minimum" lag depends on the length of the text: the longer the text, the longer the time between reading and rereading. This might explain the differences between the findings of Rawson and Kintsch (2005, 2012 respectively) on the one side and Glover and Corkill (1987) and Krug et al. (1990) at the other side, as different lags are beneficial at different retention intervals.

Additionally, the dependency of lag was explained by more integration of encoded information in the distributed (long-lag) condition. Drawing on the construction-integration model, it was assumed that the lag between rereading changes the baseline activation of nodes during reading, which affects the construction phase during reading and thus, the building of interconnections between nodes. However, Rawson (2012) did not explain how this change in baseline activation is affected by the lag between rereading and how the change in baseline activation affects the quantity of nodes. Nevertheless, Rawson (2012) found more within-sentence and between-sentence integration, indicating that indeed more integration took place. Rawson (2012) assumed that this interconnectedness is especially helpful in delayed recall. This might be explained by more retrieval cues, which makes the information more accessible. However, Rawson (2012) did not explicitly explain why this greater interconnectedness should only be beneficial at delayed test.

Furthermore, Rawson (2012) found that the performance of the massed condition declined across the retention interval, whereas the performance in the distributed condition did not. Therefore, it seems that distributed rereading affected forgetting.

Verkoeijen et al. (2008) also investigated distributed rereading and compared different lags. The similar texts as used by Rawson and Kintsch (2005) were reread massed, with a lag of 4 days (short-lag) or 3.5 weeks (long-lag). The test took place two days after rereading. Beneficial effects were found for the short-lag condition compared to massed rereading, but not for the long-lag condition. This might be explained by the lag effect: As noted above, the longer the lag, the longer is the optimal retention interval (Cepeda et al. 2008). In the experiment conducted by Verkoeijen et al. (2008), the same retention interval was used for the three different lags. It could be assumed that the long-lag condition might have outperformed the massed (and short-lag) condition after a longer retention interval.

To summarize, distributed practice seems to be also effective in rereading of texts, but the length of the text, the lag and the retention interval seem to influence the effectiveness.

However, Dunlosky et al. (2013) stated in their review that rereading in general seems to have low utility. Even distributed rereading seems to be less efficient or beneficial for learning than other learning techniques such as practice testing or elaborative interrogation. Furthermore, rereading is time-consuming, especially in real learning scenarios. For example, if a student reads a text for exam preparation in learning psychology, rereading the chapter about classical conditioning would take more time than take a test about the chapter. If the test (without feedback) is more beneficial than rereading, regardless of it being read distributed or massed, it is questionable if (distributed) rereading is a promotable learning strategy.

Additionally, in real-world learning settings, such as school learning, the possibilities of distributing text reading go beyond rereading. Thus, students may not reread textbook chapters, or take a test about it, but will read the subsequent chapter. This would be a learning from multiple, complementary texts scenario. Might such a realistic learning scenario profit from distributed practice?

Distributed Practice versus Distributed Learning

This learning scenario is not only relevant in learning from text, but also learning in general. When I learned about the plant cell today, should I learn about the bacterial cell tomorrow – or next week? However, research regarding distributed practice is mostly restricted to repeated learning. For example, Cepeda et al (2006) state “The distributed practice effect refers to an effect of interstudy interval (ISI) upon learning. ISI is the interval separating different study episodes of the *same materials*.” (p. 354). However, this restriction to repetition is less clear in other publications. For example, Schwartz et al. (2011) wrote a guide to improve learning efficiency and defined distributed practice as “learning that is spread out across relatively long periods of time rather than massed all at once” (p. 10). In this definition, the reader – who maybe is motivated to improve his or her learning efficiency – might consider distributed practice to be beneficial in learning in general, thus, also in learning about the bacterial cell *and* the plant cell.

Furthermore, I want to question whether the effects of distributed practice have to be restricted to repetitions. Thus, I would assume that distributing study time might also be effective for non-repeated but continued learning. I will use the term distributed *learning* to refer to distributing continued learning, whereas distributed *practice* refers to repeated learning.

In the following, I will give a short insight into theories explaining distributed practice. The focus here is on the question whether their assumptions are restricted to repetitions. Afterwards, I will present some research investigating distributed practice without (exact) repetitions, before I will resume what assumptions and research questions can be derived for distributed learning from texts.

Cognitive Mechanisms Underlying Distributed Practice and Learning

To give an insight in the underlying cognitive mechanisms, I will shortly introduce three major theories, which also have been combined in hybrid forms (e.g. Delaney, Verkoijen & Spirgel, 2010). Furthermore, I will focus on the role of repetition in these theories and reflect if those theories would also predicts benefits of distributed learning.

Deficient Processing Theories. *Deficient processing theories* assume that massed practice (but not distributed practice) leads to a deficient processing of the subsequent presentations of the to-be-learnt items. Several theories exist that assume different processes that are hindered by massed practice. One example of a deficient processing account of distributed learning is the (in) attention theory (Hintzman, 1974). Hintzman (1974) assumed that less attention is paid to items, which are recognized as familiar. If items are presented in a massed fashion, it is more likely that they are recognized and thus receive less attention than distributed items, which are not recognized as familiar. As the deficient processing should only occur if the item is recognized as familiar, the second learning occasion has to be at least to some amount a repetition of the first. Thus, beneficial effects of distributed learning would not be assumed.

However, although those accounts of distributed practice effect seem intuitively reasonable, they are challenged the findings regarding the lag effect. It seems to be difficult to explain why deficient processing should increase with increased lag, as the key mechanism should be more dichotomous (familiar or not) and furthermore, why a particular lag is beneficial for a particular retention interval (Cepeda et al. 2006). Thus, it seems to be unlikely that deficient processing is the only account for distributed practice effects.

Contextual-Variability Theory. Following the *contextual-variability theory*, distributed practice benefits from more contextual retrieval cues (Glenberg, 1979). The memory representation of an item does not only include the item itself, but also contextual references, such as the location, and learner-inherent properties such as the cognitive and affective stage.

It is assumed that this context of an item is more variable in distributed than in massed practice, and that variability increases with the time that has passed between several presentations of the item. Therefore, during the second presentation of an item, the memory representation of the first presentation is complemented by the contextual references of the second presentation. In the retrieval of this item, contextual information can function as retrieval cue and consequently, the increased contextual-variability in distributed practice is assumed to enhance retrieval at a later test.

In the contextual-variability theory, distributed learning can be assumed to have beneficial effects on learning. The key mechanism is the enrichment of the memory representation with additional contextual references. However, if an item reminds the learner of another, earlier presented item, this would also evoke retrieval of this first presented item (as demonstrated in the experiments by D'Agostini & DeRemer, 1973, and Madigan, 1969). Thus, the memory representation of the first content can still be complemented by the contextual references of the second content. Consequently, both contents would benefit from the additional contextual references.

The contextual-variability theory can explain the spacing as well as the lag effect, as more contextual variability is assumed for longer than for shorter lags, but the complex interaction between lag and retention interval might challenge it assumptions. In particular, it is difficult to explain why a certain lag should correspond to a certain retention interval when considering contextual variability as key mechanism (Cepeda et al., 2006).

Study-phase Retrieval Theory. The *study-phase retrieval theory* assumes that retrieval processes during the second presentation of an item account for the benefits of distributed practice (Thios & D'Agostino, 1976). The second presentation of the item stimulates the retrieval of the memory representation of its first presentation, which strengthens the memory trace of the item. In distributed practice, the first presentation needs to be retrieved from long-term memory, whereas in massed practice, the first presentation is still accessible from working memory. Furthermore, as the interstudy interval increases, the retrieval from memory becomes more difficult, which also might benefit learning in the long run (Bjork, 1975; Pyc & Rawson, 2009). As retrieval affects forgetting, the study-phase retrieval theory can explain that distributed practice especially benefits long-term learning. Furthermore, it can explain why interstudy intervals that are too long could hinder learning: With long retention intervals, retrieval might get too difficult and, therefore, impede benefits of distributed practice (comparable to unsuccessful retrieval in retrieval practice, Rowland, 2014). Furthermore, with longer interstudy intervals, the second presentation might even not be recognized as repetition, so that no retrieval attempt takes place (Johnston & Uhl, 1976).

In their meta-analysis, Janiszewski, Noel and Sawyer (2012) tested the hypotheses derived from these explanations of the spacing effect. Theories based on a retrieval explanation of distributed practice described the data more accurately than deficient processing and contextual variability theories.

Retrieval and Reminding. Accordingly, Benjamin and Tullis (2010) came to the conclusion that the effects of distributed practice are best explained by retrieval mechanisms. They further expanded the effects to *reminding.* Thus, beneficial effects of repetitions are explained by the reminding generated by similar items. This reminding facilitates retrieval of previously learned items, which enhances memory for these items. Thereby pure repetition leads to most reminding, whereas for example related words lead to less reminding. However, this reminding could also be sufficient to generate distributed practice effects. Despite the potential benefits, after a longer delay, reminding is also less likely due to forgetting of the earlier item. This applies also to repeated items, which are presented distributed instead of massed. However, with this reminding model, distributed practice effects can be generalized to not repeated, but related items, which are thought to generate reminding. Accordingly, Tullis et al. (2014) found a significant effect of reminding on the recall of associated but not repeated words. However, they did not find a lag effect, neither for repeated nor for reminding words.

This might be explained by the manipulation: they compared a short lag condition (16 intervening items) and a long lag condition (8 intervening items) with recall immediately after learning (even if they varied the retention interval in Experiment 1B ranging from five to 180 seconds, but found no differences between the retention intervals). The missing effects of distribution might be then explained by the lack of a massed condition (Cepeda et al., 2006).

Therefore, it can be assumed that the effects of distribution should not depend on repetition, but on reminding. However, research on distributed learning is rare. Braun and Rubin (1998) investigated the underlying mechanisms of the spacing effect. In Experiment 3, the participants were asked to learn lists of word pairs with massed and spaced presentation. The word within the pairs began with the same three letters, e. g. BURden and BURlap (example from Delaney et al., 2010). Therefore, the second word was not a repetition of the first, but the overlap in letters enables reminding. They found a spacing effect for the first as well as for the second word. Vlach and Sandhofer (2012; but see also Vlach, 2014) investigated spacing effects with more complex materials in school. The participating children learned in four sessions (5 min each) about food chains. The lessons were either presented massed (all sessions in immediate succession), clumped (two sessions one day, two session on the next day) or spaced (one session per day for four days). The materials illustrated the food chain in different biomes, for example arctic or grassland. Therefore, the *concept* of food chains was taught repeatedly but in different contexts. One week after the last session, children in the spaced condition outperformed children in the massed condition in the ability to make simple and complex generalizations. Smith and Rothkopf (1984) varied the presentation of an eight-hour statistics course. The course was composed of four videotaped sessions, which were presented either massed (all four within one day, with short breaks in-between) or distributed with a lag of one day. Five days later, the memory of information taught in the sessions was tested. Students in the distributed condition recalled 13% more information than students in the massed condition in free recall, 14% more in cued recall. As the lessons covered different topics (Lesson 1 and 2 descriptive statistics; Lesson 3 and 4 inferential statistics), it can be assumed that within the lessons, no direct repetition was undertaken. Instead, the learning was continued and the learners needed knowledge given in Lesson 1 to understand Lesson 2 and following. However, a similar study of Randler et al.(2008) came to different results. Randler et al. (2008) compared blocked and traditional teaching in biology class with seventh graders. The students received an educational unit about the water lily. The unit consisted of four session 45 min each. The blocked condition received all four session within one morning, whereas the traditional condition received the lessons distributed in a weekly schedule, thus, distributed with a lag of one week. Students in the traditional condition outperformed students in the blocked condition immediately after learning, whereas no differences were found seven weeks later.

To summarize the rare empirical data regarding distributed learning, these studies provide first evidence of beneficial effects of distributed learning, thus, without repetition but reminding. Therefore, it seems promisingly to go beyond the restrictive definition of Cepeda et al. (2006) and investigate distributed practice without repetition.

I would further argue that distributed learning might be especially valuable in learning from texts. First, distributed practice in learning from texts – especially with longer texts – is time-consuming and less efficient than testing or interrogative elaboration (Dunlosky et al., 2013). Second, reminding is an inherent process in text reading. Considering the theories of comprehending and learning from texts, the “reminding” within and between texts is inherent. Reminding might result in cohort activation, thus passive activation processes, but also in more elaborative, strategic and slow (re-)activation processes as in coherence-based retrieval (van den Broek et al., 1996). This reminding might have beneficial effects on long-term learning as “reminding is, in some sense, unintentional retrieval practice” (Tullis et al., 2014; p. 1537). Furthermore, reminding, contrary to repetition, might help the reader to decontextualize the information, as with the reactivation in a new context is added to the original context (Beker et al., 2017). Consequently, distributed learning might be especially beneficial in learning from multiple texts.

Distributed Learning with Multiple, Complementary Texts

As introduced above, reading multiple texts instead of one single text provides additional demands on the reader, whereas complementary texts are a special case. In complementary texts it can be assumed that new information will be "seamlessly" adapted to previously processed information (Strømsø, 2017).

However, the adaption of new information across complementary texts should depend on the activation of information from the texts read earlier. As noted above, this activation is triggered by reminding processes: In reading the second text, the reader is reminded of specific information given within the first text and activates this information. Comparable to activation of prior knowledge during reading, the activation of information from the earlier text might occur passively, either through memory-based processes, or actively, through active retrieval, thus, cohort activation or coherence-based retrieval according to the landscape model (van den Broek et al., 1996). The passive activation of information across multiple texts was demonstrated in experiments by Beker and colleagues (2016). Participants read texts with or without inconsistencies. In a previously read text, information was either provided that could explain and resolve the inconsistencies or no such information was provided. The reading time for target sentences after inconsistencies were longer than for target sentences after consistent information but only if no information that could be used to resolve the inconsistency was provided in the previous text. This pattern of results may be interpreted as an indication of routine activation of information from the previous text, which was apparently used for restoring consistency in the subsequent text (for similar results with children in Grades 4 and 6, see Beker et al., 2019).

Despite a lack of empirical evidence, it is assumed that the information read in another text (or document) is less accessible than information read in the very same text (Britt et al., 2013). Additionally, the information from the previous text might become even less accessible and thus more difficult to reactivate the more time has passed between reading the first and the second text. In terms of R. A. Bjork & Bjork (1992), in distributed reading of complementary texts, the retrieval strength is lower than in massed reading, thus, the accumulation of storage strength is accelerated. Furthermore, in distributed reading, information about earlier texts has to be retrieved from long-term memory instead of from short-term memory (or in terms of van den Broek et al. (1996), in coherence-based instead of cohort activation).

When applying these assumptions on distributed learning from complementary texts, it can be derived that time between complementary text may lead to similar effects as time between repetitions of the very same texts. When reading the subsequent text, the reader is reminded of the information given in the previous text and is forced to retrieve this information. Similar to distributed practice, a longer lag between the texts lowers the retrieval strength and thus makes the retrieval more difficult.

Research Questions

In the previous sections, I presented central theoretical frameworks regarding comprehending and learning from single and multiple texts as well as regarding desirable difficulties. I integrated these frameworks to answer the question how learning from text could be improved by desirable difficulties. I deducted that distributed learning, thus, distributed practice with continued but not repeated materials might benefit learning in general, if the materials evoke reminding. Furthermore I stated this might be especially applicable to and valuable in learning from multiple, complementary texts. In the following Chapters II—IV, I will present findings from five experiments that aimed to investigate these assumptions in both school and laboratory settings.

These experiments were guided by four central research questions:

- Is distributed learning with multiple, complementary texts beneficial for learning?
- Does the effects of distributed learning with multiple, complementary texts depend on the amount of domain-specific prior knowledge as in the reverse coherence effect?
- Does distributed learning affect the judgements of learning similar to distributed practice?
- Is distributed practice in reading single text beneficial for learning of school students?

In Chapter II, I will present two experiments conducted in a school environment with seventh graders. I specifically developed complementary text materials, which enable reminding without repeating information, in two domains, biology and physics to investigate distributed learning with complementary texts. In both domains, the second texts complements information given in the first text but does not repeat or contradict this information. In biology, the text covered the topic of living cells. Text 1described the plant cell, while Text 2 described the bacterial cell. Thus, the knowledge about cells developed by reading the first text is complemented by knowledge about the bacterial cell. As plant and bacterial cell share some features but also have some dissimilarities, the reader is reminded of the first text during reading the second. For example, in the first text, the function of ribosomes are explained, but only mentioned in the second text. In physics, the texts covered the law of conversation of energy (Text 1) and the first law of thermodynamics (Text 2). The complementary character of these texts is warranted by the fact that the first law of thermodynamics is a special case of the law of conversation of energy. Furthermore, information given in the first text, for example the definition of a closed system, is needed to understand the second text.

In addition to the beneficial effects of distributed learning with complementary texts, the experiments in Chapter II were also designed to investigate the effects of distributed learning on the meta-cognitive judgements of learning.

In Chapter III, I will present two experiments (one preregistered) conducted in the laboratory with university students. The design of the experiments resembles the design of the experiments presented in Chapter II to answer the research question whether distributed learning is beneficial in long-term learning from complementary texts for young adults in a more controlled setting. Additionally, the experiments were also designed to investigate the moderating effects of prior knowledge.

Furthermore, in Chapter IV, I will present an additional preregistered experiment in a school environment. Contrary to the previous experiments, distributed (repeated) practice with single texts instead of distributed learning was investigated. The research design here is a conceptual replication of Rawson and Kintsch (2005).The aim of this experiment was to investigate whether distributed practice with single texts is also beneficial for young learners. Furthermore, the moderating effect of domain-specific knowledge and the effect of distributed practice on the meta-cognitive judgments of learning was assessed.

In the final Chapter V, I will summarize the results of all five experiments and discuss their limitations but also their implications for theory and practice of distributed practice and learning with single and multiple texts.

Chapter II

Experiments 1 and 2

Beyond the Distributed Practice Effect: Is Distributed Learning Also Effective for Learning with Non-Repeated Text Materials?

A version of this chapter is published as:

Greving, C. E., & Richter, T. (2021). Beyond the distributed practice effect: Is distributed learning also effective for learning with non-repeated text materials? *Frontiers in Psychology*, *12*, 685245. https://doi.org/10.3389/fpsyg.2021.685245

Beyond the Distributed Practice Effect:
Is Distributed Learning Also Effective for Learning with Non-Repeated Text Materials?

Carla E. Greving & Tobias Richter

Abstract. Distributed learning is often recommended as a general learning strategy, but previous research has established its benefits mainly for learning with repeated materials. In two experiments, we investigated distributed learning with complementary text materials. Seventy-seven (Experiment 1) and 130 (Experiment 2) seventh graders read two texts, massed vs. distributed, by one week (Experiment 1) or 15 min (Experiment 2). Learning outcomes were measured immediately and one week later and metacognitive judgments of learning were assessed. In Experiment 1, distributed learning was perceived as more difficult than massed learning. In both experiments, massed learning led to better outcomes immediately after learning but learning outcomes were lower after one week. No such decrease occurred for distributed learning, yielding similar outcomes for massed and distributed learning after one week. In sum, no benefits of distributed learning vs. massed learning were found, but distributed learning might lower the decrease in learning outcomes over time.

Learning from texts is crucial for knowledge acquisition in school and in higher education. However, rereading texts is time-consuming and does not necessarily lead to successful learning (Dunlosky et al. 2013). Thus, exploring strategies that can improve learning from text is an important research focus. One central question is whether this goal can be accomplished better by making the comprehension process easier or by making it more difficult, which might engage the reader in deeper processing of the text (McNamara et al. 1996). The latter strategy is consistent with the desirable difficulties approach. Desirable difficulties are properties of learning procedures that make the learning process subjectively difficult, which may hamper learning in the short run but foster better long-term retention (E. L. Bjork & Bjork, 2011; R. A. Bjork, 1994; Lipowsky et al., 2015). Desirable difficulties might be involved in different learning procedures such as retrieval practice and distributed practice. Regarding the underlying cognitive mechanisms, several theories assume that the retrieval and activation of prior knowledge during learning is important for making a learning difficulty desirable (E. L. Bjork & Bjork, 2011; Delaney et al., 2010; Toppino & Gerbier, 2014).

In learning from text, a desirable difficulty can be introduced by distributing the rereading of text materials, thus, reading the same material for the second time, over time. Distributed rereading has been shown to enhance learning outcomes in the long run (in free recall $d = 0.56$, Rawson, 2012; in free recall $d = 0.73$, in text comprehension performance $d = 0.53$ Rawson & Kintsch, 2005). However, apart from self-regulated learning for an exam or test, the application of distributed rereading in learning situations is limited because it only focuses on repetition. In classroom learning, reviewing materials is uncommon (Dempster, 1989). For example, when reading a textbook in school, texts are rarely read repeatedly but instead are followed by reading new texts (e.g., an advanced chapter in a textbook) that are related to texts read earlier (e.g., an introductory chapter in a textbook). The extent that the distribution of learning time functions as a desirable difficulty when reading two or more texts that are complementary to each other remains an open question.

In the present research, we investigated whether distributed reading of two complementary texts (similar to subsequent chapters in a textbook) makes learning from these texts more difficult for school students but simultaneously improves long-term retention. We assumed that the information given in the first text must be retrieved when reading the second text to establish coherence, which is more difficult when the texts are distributed over time. This retrieval difficulty should lead to better text recall, especially in the long run. In the following sections, we discuss the relevant processes underlying learning with texts. We discuss the effects of prior knowledge and how information is (re)activated when reading. Based on a discussion of distributed practice, we propose the idea of distribution by time in learning with complementary texts.

Improving Learning from Text

Learning from text is based on comprehending the text, which involves the construction of a situation model (also termed mental model) of the text content. The situation model goes beyond a representation of the text itself (surface representation) or the information explicitly given in a text (propositional textbase, Kintsch, 1994). Deep understanding of texts and long-term learning might benefit especially from the active retrieval of information from long-term memory, which can be interpreted according to the principle of generative learning (Wittrock, 2010) or learning as active information processing (Mayer, 1996). In their landscape model of reading, van den Broek and colleagues (1996) termed this process as coherence-based retrieval. Contrary to the process of cohort activation, which is based on a passive spread-of-activation mechanism, coherence-based retrieval is a strategic, slow, and effortful process that aims at establishing coherence in accordance with readers' goals and standards of coherence.

Previous research has examined numerous possibilities to improve learning from text by increasing the difficulty of text processing and, arguably, by promoting active, coherence-based retrieval during comprehension. Several studies by McNamara and colleagues looked at the effects of lowered text cohesion on learning. These studies found that better (i.e., more cohesive) texts improve the comprehension of students with low prior knowledge, whereas the opposite was found for high-knowledge students (reverse coherence effect, McNamara et al., 1996; McNamara & Kintsch, 1996). Low-cohesion texts seem to foster the active retrieval of prior knowledge and knowledge-based inferences of high-knowledge students, which might account for their higher comprehension outcomes (McNamara, 2001). However, this effect also seems to depend on reading skill, as skilled readers also have been shown to profit from high-cohesion texts (Ozuru, et al., 2009). Therefore, especially less skilled readers need low cohesion texts to activate their prior knowledge.

Other text features have been varied to assess the extent that they stimulate active processing of information such as local and global coherence (Boscolo & Mason, 2003), syntactic structure of sentences (Feng et al., 2013), presentation format, text organization and example context (McCrudden et al., 2004), sentence order and letter deletion (McDaniel et al., 2002), and verb cohesion and syntactic simplicity (Mills et al., 2015), but these studies have yielded mixed results of increased processing difficulty, ranging from positive effects to negative effects on learning. Apart from obvious differences in the manipulation of text difficulty, one condition that might have contributed to the inconsistent results is that learning outcomes were measured immediately after reading in these studies. But desirable difficulties might play out their advantages in particular at longer intervals between learning and assessment of learning outcomes (Pashler et al., 2007; Rawson & Kintsch, 2005; Rawson, 2012; but see also Dunlosky et al., 2013).

Distribution of Text Reading as Special Case of Distributed Practice

Distributing text reading over time might be another largely unexplored possibility to promote active processing and especially to retrieve information when learning from texts. The effects of temporal spacing of materials have been studied extensively with regard to distributed practice in which repetitions of the same materials (or repetitive practice of similar materials) are distributed into several (shorter) learning sessions rather than one (longer) learning session (massed practice). For this type of learning, spacing usually has positive effects, especially for long-term retention, a phenomenon called the spacing effect (Cepeda et al., 2006). Longer interstudy intervals (i.e., the time between repetitions) are usually better for longer retention intervals, a phenomenon also known as the lag effect (Cepeda et al., 2006; 2008; 2009). In accordance with Cepeda et al. (2006) and Küpper-Tetzel (2014), we will use the term distributed practice to refer to both effects, thus, the spacing and the lag effect.

Positive effects of distributed practice have mainly been shown for simpler materials such as word pairs or learning the vocabulary of a foreign language (Cepeda et al., 2006; 2008), but some studies have also established distributed practice effects for more complex materials such as science concepts (Vlach & Sandhofer, 2012; Vlach, 2014) and expository texts (Rawson & Kintsch, 2005; Rawson, 2012; Verkoeijen et al., 2008; but see C. E. Greving & Richter, 2019, who did not find a benefit in seventh graders). Thus, distributed practice seems to be beneficial for learning with a broad range of materials.

Overall, the benefits of distributed practice are a robust empirical phenomenon (Cepeda et al., 2006; Carpenter et al., 2012). Donovan and Radosevich (1999) reported in their meta-analysis an overall mean weighted effect size (Cohen's *d*) of 0.46 (95% CI [0.42, 0.50]). In their review of learning techniques, Dunlosky et al. (2013) evaluated distributed practice as having high utility for learning. Despite these findings, students seem to underrate the effectiveness of distributed practice (see Son & Simon, 2012 for a review). For example, when learners were asked immediately after learning to estimate the proportion of items they would correctly recall in a posttest, their estimates were higher for items learned in a massed fashion compared to items learned in a distributed fashion (Kornell, 2009). One possible explanation for the negative effects of distributed practice on the meta-cognitive judgment of predicted learning success might be a lower experienced fluency during distributed practice (Alter & Oppenheimer, 2009; R. A. Bjork et al., 2013), which would be in line with the interpretation that distributed practice could induce a desirable difficulty.

Schwartz et al. (2011) defined distributed practice as “learning that is spread out across relatively long periods of time rather than massed all at once” (p. 10). This definition suggests that the benefits of spacing might not be restricted to learning materials that are repeated explicitly but also extend to learning with related but not repeated materials. Nevertheless, currently, the evidence for beneficial effects of distributing non-repeated learning materials over time is scarce. To prevent misunderstandings, we will use the term distributed learning to refer to the distribution of learning materials that are not repeated and the term distributed practice for learning that involves the distribution of repeated learning materials. Please note that the term distributed learning might also be understood as a superordinate term that subsumes both forms of learning with materials that are distributed overtime but that we use it in a more specific way here to designate distributed learning with non-repeated materials.

Especially, we are interested in distributed learning with multiple, complementary texts. At this point, it is useful to define the notion of text. We use a broad definition of text here according to which a "text represents the inscription of ideas in linguistic form" (Alexander & Jetton, 2003, p. 201). Thus, texts are made of written or spoken words, possibly accompanied by other modes of representation such as graphs, pictures, or animated pictures. We speak of a "text" when it can stand on its own, that is when the average reader can, in principle, establish a globally coherent, meaningful representation of the text content by reading the text (or listening to it) and drawing on their prior knowledge. Thus, textbook chapters usually qualify as texts in the sense of this definition, whereas, for example, paragraphs within a chapter would not be texts because they do not make sense when read on their own. That said, the example of textbook chapters shows that reading one text can be particularly helpful to understand another one. School learning often involves the reading of multiple texts that cover different aspects of the same topic such as subsequent chapters in a textbook. Those texts can be framed as complementary texts. Complementary texts are multiple texts that are “convergent and require adding pieces of information together” (Primor & Katzir 2018 , p. 4; see also Richter et al., 2020). As Carpenter et al. (2012) noted do textbooks typically not provide distributed repetition of concepts. Content provided earlier in the textbook often serves as background knowledge that is helpful for understanding later chapters. For example, the first lesson in a science class that covers the complex topic of cell biology might require students to view the plant cell under the microscope and then consult their textbook to read information about the different components of the cell (e.g., the functions of different organelles). In the next lesson, students might read the subsequent textbook chapter about the bacterial cell. In this chapter, they would learn about the structure of the bacterial cell and which organelles can be found in the bacterial cell. However, in this chapter, the functions of the organelles will not be explained again. Thus, the students must retrieve this information from memory to fully understand the chapter. The time between lessons could also vary. The second lesson might follow immediately after the first one (massed learning) or after some time has elapsed between the two lessons (distributed learning). This scenario leads to the question of whether the well-established benefits of the distribution of learning also occur for reading complementary but non-repeated text materials in the school learning environment.

An indication that distributed practice effects may occur even without repetition was provided by Braun and Rubin (1998, Experiment 3). In this experiment, the participants learned word lists with massed and distributed presentation of word pairs. The two words that formed a pair began with the same three letters. For example, if the first word was BURden, the second was BURlap (example from Delaney et al., 2010). Although the second word was not an exact repetition of the first word (i.e., only a partial overlap), a spacing effect occurred for the first and the second word. A study by Vlach and Sandhofer (2012) is another example of distributed learning without explicit repetition. Notably, these authors used more complex materials in a real-world educational setting. They investigated whether distributed practice can aid the generalization of science concepts in children. Elementary school students were taught the concept of food chains. The children received four lessons in massed (all sessions in immediate succession on one day), clumped (two sessions on one day, two sessions on the next day) or spaced (one session per day for four days) schedules. In each of the lessons, the food chain was illustrated within a different biome. Thus, the materials were not repeated exactly as in typical distributed-practice studies, but the concept of food chains was repeatedly embedded in different contexts. One week after the final learning session, the ability to make simple and complex generalizations of the concept to a new biome was assessed with two tasks. Children with a distributed learning schedule outperformed children with a massed learning schedule in both tasks (d = 0.89, respectively d = 1.91, calculated from $\eta 2$). Another study of distributed learning with non-repeated materials was conducted by Smith and Rothkopf (1984). In an eight-hour statistics course, parts of the course (four videotaped sessions) were presented in a distributed fashion with an interstudy interval of one day or massed within one day with only short breaks between the sessions. After five days, participants who received the distributed sessions outperformed participants who received the sessions in a massed fashion in free and cued recall with an increase of 13 % and 14%, respectively. In a more recent study, Randler and colleagues (2008) investigated blocked vs. traditional teaching in biology class. The traditional condition received four lessons of 45 min each in a weekly schedule, while in the blocked condition, the students received all lessons within one morning. Immediately after the last lesson, students in the traditional condition outperformed students in the blocked condition, whereas no difference was found seven weeks after the last lesson. In sum, these studies provide evidence of distributed learning effects with non-repeated materials. However, to our knowledge, no research exists that has examined distributed learning effects in learning with multiple, complementary texts.

Rationale of the Present Experiments

We conducted two experiments to test the assumption that the temporal distribution of complementary multiple texts leads to better learning. Both experiments were conducted with students in Grade 7 (12-13 years old) in the school classroom. The experimental materials covered two different domains to gain tentative information about the generalizability of results. In Experiment 1, participants received the texts from both domains (within-subjects, Figure 2.1). In Experiment 2, each participant received only the texts from one domain, that is, topic was varied between-subjects.

Two pairs of expository texts from the natural sciences (biology and physics) were developed to match the typical contents and difficulty of texts that seventh graders read in their regular classes. The texts were coherent with each other in the sense that the second text built on concepts from the first text, resembling subsequent textbook chapters. Learners were randomly assigned to one of two learning conditions. They read the two texts per domain in a massed fashion or in a distributed fashion with a learning interval of one week (Experiment 1) or 15 min (Experiment 2) between the two texts. Immediately after learning, students judged four aspects of the learning process. They indicated the perceived difficulty of the reading task and predicted their learning success. Furthermore, they rated the perceived similarity between the texts and perceived learning coherence. Learning outcomes were assessed approximately 5 min after learning (immediate) and one week later (delayed).

The following hypotheses were derived from the theoretical considerations laid out in the previous sections:

Hypothesis 1 (main hypothesis): We expected the potential learning benefits of distributed learning for learning from both texts to depend on time of test. Immediately after learning (that is immediately after reading the second text), we expected no benefits of distributed over massed reading (Hypothesis 1a), whereas a learning benefit of distributed reading should emerge at a longer time interval of one week after learning (Hypothesis 1b).

Hypothesis 2: We expected domain-specific prior knowledge to be a positive predictor of learning outcomes. Text comprehension and learning from text are based on integrating new information with existing knowledge (Kintsch 1988). In line with this general notion, numerous empirical studies found domain-specific prior knowledge to be a strong predictor of text comprehension and learning from text (e.g. Ozuru et al., 2009; Schneideret al., 1989)

Hypothesis 3: We expected that distributed learning – as desirable difficulty – should lead to overall higher perceived difficulty (Hypothesis 3a) and lower expected learning success (Hypothesis 3b) compared to massed reading.

Hypothesis 4: Finally, we expected the perceived similarity (Hypothesis 4a), that is how similar participants judged the texts read at the two learning occasions, and learning coherence (Hypothesis 4b), that is as how strong they judged the relationships of the two text, to be lower in the distributed than in the massed condition. This hypothesis can be derived from the assumption that the passive retrieval of information is more difficult and therefore less likely to be successful in distributed learning.

In addition to testing these hypotheses, three learner characteristics, reading ability, working memory capacity, and (in Experiment 1) reading strategy knowledge were examined to control for pre-existing differences in these variables.

Experiment 1

Experiment 1 examined distributed learning with a one-week interval between reading the first and second text in the distributed condition, as opposed to no interval in the massed condition. The time interval was chosen because school lessons often follow a weekly schedule, which makes a one-week learning interval an ecologically valid interval with which to start. Moreover, learning intervals of one week have been used in previous studies on distributed rereading (e.g., Rawson and Kintsch, 2005).

Method

Design. Experiment 1 was based on a 2 x 2 x 2 design with the independent variables learning condition (massed vs. distributed learning), retention interval (immediate vs. one-week delay), and domain (biology vs. physics). Learning condition was varied between participants, retention interval and domain were varied within participants. Participants were randomly assigned to one of the two learning conditions within classes, thus, all experimental conditions were realized in each class. The assignment of the two different comprehension tests to one of the two retention intervals was counterbalanced between participants.

Participants. Ninety-seven seventh graders (52 boys, 45 girls) with a mean age of 12.32 years (*SD* = 0.47, age was not reported for nine students) from four classes of a German comprehensive school participated in the experiment. Parental permission was obtained for all participating students. Students without permission did not participate in the experiment. For those students, no data was recorded due to data protection regulations. Therefore, we have no information how many students did not receive their parents' permission to participate in the study. Students were randomly assigned to either the massed learning condition (n = 50) or distributed learning condition (n = 47). Students received sweets after each of the sessions and a magic cube puzzle after the last session as a reward for participation. Fourteen students missed one of the two learning sessions and were excluded from all analyses. Four students missed the first session in which the domain-specific prior knowledge test was assessed. Their data was excluded from analysis. The data of one student was excluded because of technical problems. In the end, 77 participants (massed learning condition: 38; distributed condition: 39) remained in the sample.

Power Analysis. This study was the first to investigate distributed reading with complementary text materials, which made it impossible to form expectations about effect sizes based on previous studies. The experiment was conducted in the classroom with a heterogeneous sample, which is likely to limit possible effects. Therefore, we based power calculations on the assumption of a small population effect (d = 0.3 or OR = 1.72, respectively), following Cohen’s conventions for effect sizes (Cohen, 1988). The power (1-β) for finding an interaction effect between learning condition and retention interval of this size, determined by simulation with the R package simr (Green & MacLeod, 2016), was high (1.000, 95% CI [.996, 1.000]) given the assumed Type-I error probability (α) of .05.

Text Materials. Two experimental texts were developed for each of the two domains. For the biology domain, the first text explained the plant cell and its components, and the second text explained the bacterial cell. For the physics domain, the first text explained the law of conversation of energy, and the second text explained the first law of thermodynamics. The length of the texts ranged from 504 to 633 words and the Flesch reading ease (German formula, Amstad, 1978) ranged from 46 to 60. The biology texts contained images illustrating the structure of the respective cell; this image was presented during the whole text. We added this image to enhance comprehension of the cell structure and to enhance the ecological validity of the text, as expository texts in biology usually contain images. The first physics text contained an illustrative image of an experiment by James Prescott Joules, which was explained in the text. See Appendix A for translations of the texts used in the experiment.

The texts were constructed as self-containing texts, comparable to two chapters in the same textbook. The texts were related to each other, but Text 2 was still comprehendible without reading the Text 1, provided that the relevant prior knowledge was available. Nevertheless, in Text 1, some basic information was provided, which was relevant for understanding Text 2 but not repeated in this text. For example, in the set of biology texts, the function of the ribosomes were explained in Text 1, but not in Text 2, even if in Text 2 it is mentioned that bacterial cell have ribosomes as well. In the set of the physics text, is was explained in Text 1 what the term closed system means, but in Text 1, the definition was not repeated, although the term is needed to understand the concept of internal energy. The topic of both the biology and the physics texts were chosen after consultation with teachers. The criteria were that the topics should be optional parts of the school curriculum that are not taught regularly in school. We also made sure that the topics were indeed not taught in the participating classes. Therefore, we considered the prior knowledge to be low enough to (1) be able to acquire new knowledge by reading the texts and (2) make Text 1 relevant for full comprehension of Text 2.

Assessment of Learning Outcomes (Text Comprehension). For each domain, two comprehension test forms (A and B) were constructed to assess learning outcomes. Each test form contained eight short-answer questions and seven multiple-choice questions (one correct response, three distractors). The two different types of questions were used because of their different requirements regarding memory processes (cued recall and recognition) and the accompanying differences in item difficulty. Each student received each test form, counterbalanced at either the immediate or the delayed test. The different test forms were constructed to ensure that different questions are posed at the two times of tests. The questions were constructed in pairs (except two questions in biology), thus, the questions differed in their wording (and/ or type) but referred to the same information.

For example, one short-answer question was, A bacteria cell does not have a cell nucleus. But where can you find the genome of the bacteria cell?, and one multiple-choice questions was, To which kind of organism does the bacteria cell belong?, with the response options (a) Prokaryotes, (b) Eukaryotes, (c) Plasmid, and (d) Organelle. The order of questions was randomized. All answers to the short-answer questions were scored as either incorrect (0) or correct (1) by two independent raters who were blind to the experimental conditions (Cohen's $\kappa = .91$). In the few instances of disagreement (0.5%), the score provided by one of the two raters (determined randomly) was used.

The questions could be answered based on information from both texts (12%), from Text 1 (45%) or from Text 2 (43%). The questions referring to Text 1 and 2 asked for information explicitly given in the respective text. The questions referring to information from both texts made a form of intertextual inference necessary, such as comparing plant and bacterial cell.

In a pilot study, 82 students from 3 classes of a comprehensive school read the two texts of one of the two domains (randomly assigned) in massed fashion. Afterwards, they answered questions (33 in biology, 28 in physics) and solved a cloze with 12 gaps. Following the feedback of the teacher of the classes as well as the item difficulties and conceptional reasons, we decided to remove the cloze from the test and revised the questions intensely. 11 questions were removed (7 biology questions, 2 physics questions). One biology question was divided into two parallel questions and 7 multiple-choice questions (3 biology questions, 4 physics questions) were additionally created paralleling tested short-answer questions. Furthermore, all questions were revised in wording, adding some background to the questions to increase the retrievability. For example, one question was "Why is the golgi apparatus called post office?". In revision, we added the following background information: "Plant cells have an organelle, which is called golgi apparatus. The golgi apparatus is also called post office. as introduction of the question and added the suffix Justify your answer.".

Nevertheless, the items were still difficult, with a mean item difficulty of .20 (*SD* = .16) in the short-answer questions and a mean item difficulty of .41 (*SD* = .15) for multiple-choice questions. Cronbach's α for the different learning outcomes tests ranged from .71 and .59 (physics form A and B) and .72 and .62 (biology form A and B).

Assessment of Domain-Specific Prior Knowledge. For the biology assessment, the participants answered two short-answer questions about basic terms that appeared in the text. Additionally, they received images of the cell structures of the bacterial and plant cell and were required to label the components of the cells. For the physics assessment, the participants answered five short-answer questions about basic terms that appeared in the text. The different amount of questions was chosen because we assumed that the labeling questions would take longer and produce more variance than the short-answer questions. The order of the domains and questions within domains were randomized. One third of the responses were scored by two independent raters (Cohen's κ = .75). The internal consistency was low (biology: Cronbach's α = .57 95% CI [.43, .65]; physics: Cronbach's α = .38, 95% CI [.17, .54]). However, as the questions were developed to cover the curriculum-orientated knowledge within the two domains, the questions differ relative broad in topic (e.g. one question about the (plant) cell, one question about the genetic makeup). For a curriculum-based knowledge test like this, internal consistency might not be the most informative way to estimate reliability (Schmitt, 1996). Moreover, prior knowledge was generally low in the present sample, which restricts the item variance and, hence, the inter-item correlations that the internal consistency is based on. Considering these circumstances, we decided to proceed with the prior knowledge measure despite the low internal consistencies.

Assessment of Further Learner Characteristics. To control for pre-experimental differences between the experimental groups, we assessed several learner characteristics. In a teacher questionnaire, we asked the teachers to provide the students' grades in biology and physics (ranging from 1 = "very good" to 6 = "unsatisfactory") along with other learner characteristics such as age. Knowledge about reading strategies was assessed with the Würzburger Lesestrategie Test (WLST; Würzburg Reading Strategy Test; Schlagmüller & Schneider, 2007; split-half reliability: $r = .90$, estimated in a sample of 4490 students in Grades 7-11). Reading ability was assessed with the subtest sentence verification of the German-speaking test of reading abilities ELVES (Richter & van Holt, 2005; Cronbach's $\alpha = .83$ assessed in the current sample) and working memory with a computerized version of a Reading Span Task (RSPAN; Oberauer et al., 2000; Cronbach's $\alpha = .86$, assessed in the current sample). These learner characteristics were included only to control for differences between learning conditions.

Metacognitive Judgments of the Learning Process. After each text, participants judged several aspects of the reading process on 5-point Likert scales. They made a prediction of their learning success (What do you think, how well will you remember the content of the text you just read?; response options ranging from 1: very bad to 5: very good) and rated the perceived reading difficulty (How difficult was it for you to read the text?; response options ranging from 1: very difficult to 5: not difficult at all).After reading the second text, they rated the perceived similarity of the texts (three items, e.g., The structure of the two texts was very similar, Cronbach's α = .97, response options ranging from 1: not true to 5: true) and the perceived learning coherence (Reading the second text helped me to understand what the first text was about, , response options ranging from 1: not true to 5: true).

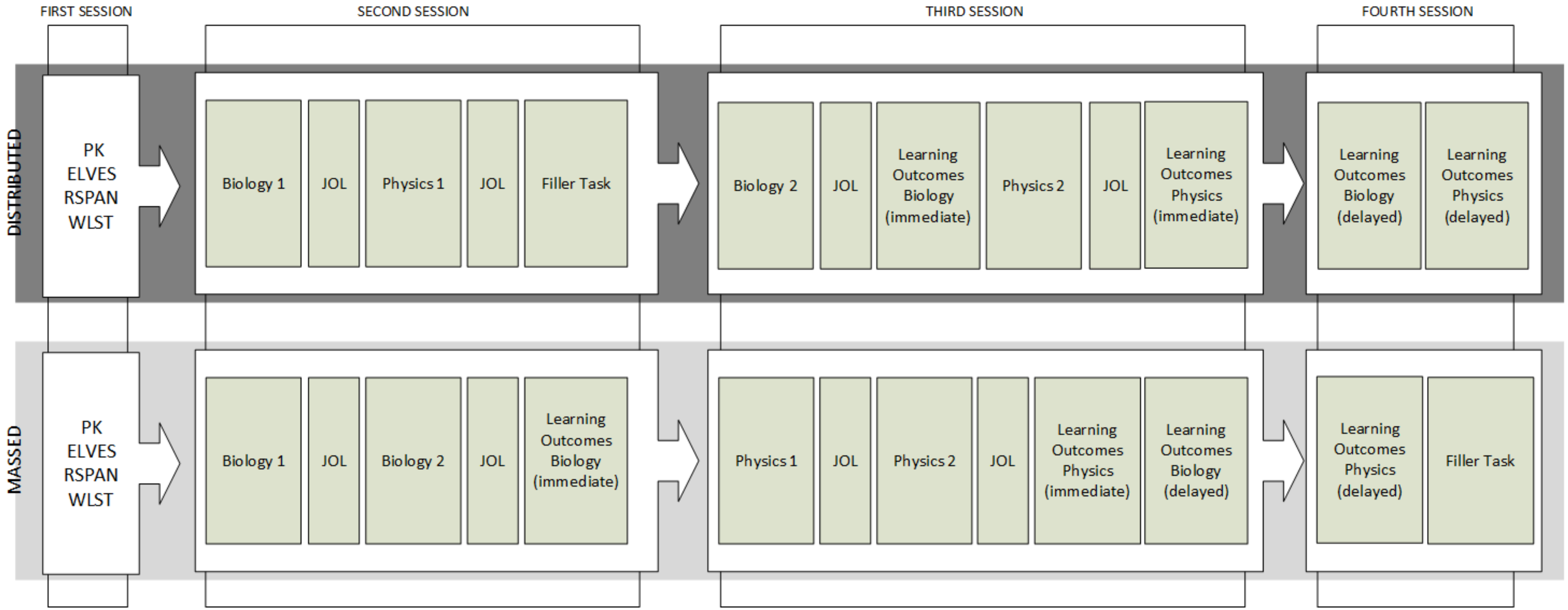
Procedure. All materials were presented on notebook computers (16.6" screen) and with the software Inquisit 3 (Version 3.0.6.0, 2011). The experiment was conducted in the classroom and consisted of four sessions (Figure 2.1). In the first session, students received examples for the different question types and an example of the reading task. Afterwards, the participants completed the four tasks related to learner characteristics: the assessment of domain specific prior knowledge, the WLST, the ELVES, and the RSPAN tests. In some classes, the RSPAN or the ELVES or both could not be conducted during the pretest session. In this case, the tests were conducted at the end of the final session.

The procedure at the remaining three sessions varied depending on the learning condition. In the second session, the participants were randomly assigned to the learning conditions. In the second and third session, the participants either read two texts in one domain (massed condition) or one text each of the two domains (distributed condition). Learning outcomes for each domain were assessed immediately after reading the second text in the domain and one week later.

All sessions started with a general instruction read aloud by the student research assistant who conducted the experiment. The instructions used in Experiment 1 are displayed in Appendix C. In the pretest, the instruction of each task was read aloud. For the following sessions all instructions were presented on screen. Participants were informed that they would read multiple texts. However, they did not know when they would read the texts and when the respective tests would take place. Thus, the participants were not aware about the assignment to different reading conditions. During the sessions, two instructors were present to help with technical problems and to ensure that all participants were working quietly. The participants read the experimental texts with the moving-window-method in a self-paced fashion. While reading, all sentences except the one that participants were currently reading were blurred. Thus, they could only read one sentence at a time. Participants were able to advance to the next sentences by pressing a key and to return to the previous sentences for rereading by pressing another key.

Figure 2.1

Procedure of Experiment 1 in the Biology Group



Note. Order of topics was counterbalanced between students (only one example is shown here). PK = Assessment of prior knowledge; ELVES = Assessment of reading ability; RSPAN = Assessment of working memory; WLST = Assessment of reading strategy knowledge; Biology 1 = first text biology; JOL = assessment of meta-cognitive judgments of the learning process; Physics 1 = first text Physics; Biology 2 = second text biology; Physics 2 = second text physics, Learning outcome physics = assessment of learning outcome for physics, Learning outcome biology = assessment of learning outcome for biology. Learning outcomes were measured immediate after learning (immediate) and one week delayed (delayed).

In two classes, the sessions could not be conducted as scheduled, which resulted in fewer days between Session 2 and Session 3 or between Session 3 and Session 4. Consequently, 10 participants in the distributed condition read the texts with a learning interval of three days instead of one week. For 10 other participants, the retention interval for the delayed test for at least one of the domains was only three days, and for 21 participants the delay was six days instead of one week. To examine the impact of these deviations from the experiment schedule, we ran all analyses regarding learning outcomes with and without the data of these participants. The effects remained unchanged. Therefore, we report the results for the full data set.

Data Analysis. We used generalized linear mixed effect models (GLMMs) for analyzing the effects of the independent variables on learning outcomes. The GLMM analyses were performed with the R packages lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), and lsmeans (Lenth, 2016). GLMM was used because of the multilevel structure of learning outcomes. A multilevel structure is typical for experiments where a sample of participants work on a sample of experimental items (Baayen et al., 2008). This was the case in our experiments, in which the participants answered 60 questions. A multilevel structure is also characteristic for classroom studies where students come from different classes. Such multiple levels create dependencies in a data set which are basically ignored by one-level analysis methods such as ANOVA or traditional multiple regression analysis. Consequently, using these methods can be misleading, among other things by underestimating standard errors and causing false-positive significance tests (for a discussion for continuous outcome variables see Baayen et al.,2008; Richter, 2006; for categorial outcome variables see Jaeger, 2008; Dixon, 2008).

We included class, participant and item as random effects (random intercepts) when the intra-class correlation of a dependent variable (a measure to quantify interdependencies in the data) exceeded .05. Thus, models with random intercepts but with no random slopes were estimated. Such models bear the risk of inflating Type-I error (Barr et al., 2013). However, including more random effect variance components also decreases power and easily overtaxes the information available in the data, leading to misspecified models that cannot be estimated (Matuschek et al., 2017). The fixed-effect structure of our models was already quite complex, making it impossible to estimate several random slopes and their covariances. Therefore, we estimated models with random intercepts only. For all models, the distribution of residuals was inspected visually for normality. For the interpretation of GLMM results, the predicted probabilities (back-transformed from the log odds) are reported. Type I error probability was set at .05 for all hypothesis tests. Directed hypotheses were tested with one-tailed tests.

Results

Differences Between Experimental Groups. There were no group differences between the two learning conditions with regard to working memory capacity, $t(66.06) = -1.48$, $p = .144$, knowledge about reading strategies, $t(54.28) = 0.13$, $p = .962$, or the teacher-reported grades in biology, $W = 832.5$, $p = .122$, and physics, $W = 820.5$, $p = .371$. However, the groups differed in their reading ability. The participants in the massed condition outperformed the participants in the distributed condition, $t(65.48) = -2.35$, $p = 0.022$. Therefore, we controlled for reading ability in all analyses. Table 2.1 provides the means and standard deviations of all dependent variables and learner characteristics observed in the two experimental groups. Correlations of all measured variables (including correlations within the learning conditions) are provided in Appendix B (Table B1).

Table 2.1

Means and Standard Deviations of Dependent Variables and Learner Characteristics in Experiment 1

	Biology				Physics			
	Massed		Distributed		Massed		Distributed	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Dependent variables								
Learning outcome (immediate)	5.95	3.19	4.37	2.47	4.76	2.92	2.93	2.93
Learning outcome (delayed)	4.59	2.78	4.07	2.02	3.63	2.05	3.14	1.73
Perceived difficulty	3.38	0.80	3.35	0.72	3.02	0.82	3.01	0.69
Self-predicted success	2.94	0.70	2.73	0.73	2.72	0.68	2.43	0.59
Perceived similarity	3.34	0.67	3.02	0.65	3.35	0.54	2.95	0.75
Perceived learning coherence	2.90	1.16	2.26	0.95	2.92	0.92	2.36	1.10
Domain-specific learner characteristics								
Domain-specific prior knowledge	2.78	2.42	2.93	2.32	3.56	1.92	5.19	2.26
Grades	2.56	0.71	2.48	0.89	2.28	0.70	2.29	0.85

Further learner characteristics	Massed		Distributed	
	M	*SD*	*M*	*SD*
Reading ability	17.16	5.53	14.09	4.28

Working memory capacity	.61	.13	.54	.18
Reading strategy knowledge	54.42	9.83	55.11	11.57

Note. All domain-specific variables are provided separately for the domains. Domain was varied within-subjects.

Learning outcomes. To test Hypotheses 1 and 2, we estimated a generalized mixed model with learning condition (contrast-coded: distributed = 1, massed = -1) and retention interval (contrast-coded: immediate = -1, delayed = 1) with their interaction and domain-specific prior knowledge (*z*-transformed) as predictors. Text comprehension performance served as the dependent variable. In addition, we included domain2 (contrast-coded: physics = 1, biology = -1) and reading ability as predictors for control purposes. Items and participants were included as random effects (random intercepts). All estimates are provided in Table 2.2.

Table 2.2

Parameter Estimates and Significance Tests for the Generalized Mixed Model for Learning Outcomes in Experiment 1

Fixed effects		
Predictors	Estimate	*SE*
(Intercept)	-1.41 ***	0.18
Learning condition	-0.11	0.08
Retention interval	-0.12 **	0.04
Prior knowledge (*z*-standardized)	0.15 *	0.06
Domain	-0.32 *	0.16
Reading ability (*z*-standardized)	0.23 **	0.08

Learning condition x Retention interval	0.08 ⁺	0.04
Random effects		
σ^2	3.29	
τ_{00}	0.32 $_{ID}$	
	1.43 $_{Item}$	
Goodness of fit		
Deviance	4157.23	

Note. Learning condition (contrast-coded: distributed = 1, massed = -1). Retention interval (contrast-coded: immediate = -1, delayed = 1). Domain (contrast-coded: biology = -1, physics = 1). Prior knowledge and reading ability were included z-standardized. Directional hypotheses were tested one-tailed.

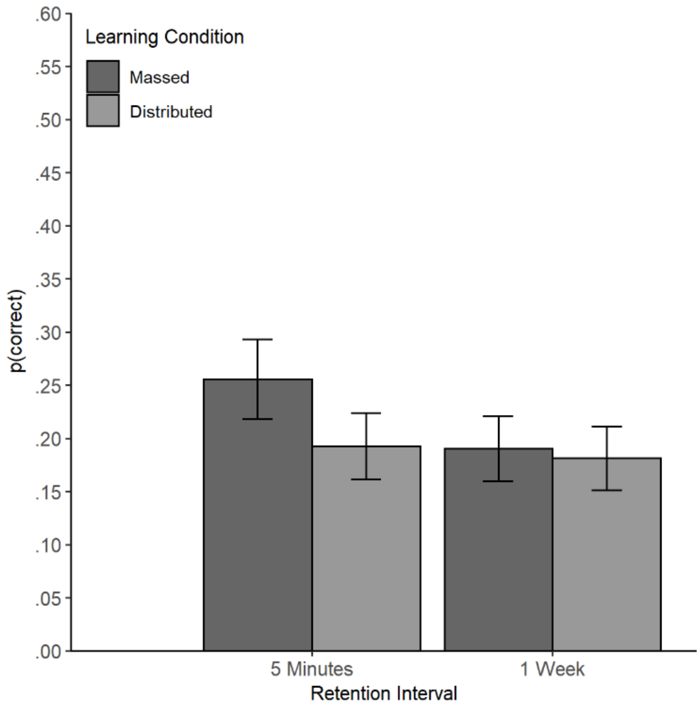
* $p < .05$ (two-tailed). ** $p < .01$ (two-tailed). *** $p < .001$ (two-tailed).

$^{+}p < .05$ (one-tailed). $^{++}$ $p < .01$ (one-tailed). $^{+++}$ $p < .001$ (one-tailed).

No main effect of the learning condition was found (β= -0.11, *SE* = 0.08, *z* = -1.36, *p* = .175). No overall difference in performance was found between the massed condition (probability = .22, *SE* = .03) and the distributed condition (probability = .19, *SE* = 0.03), *OR* = 1.24 (95% CI [0.91, 1.70]). A main effect of retention interval emerged (β = -0.12, *SE* = 0.04, *z* = -2.86, *p* = .004) with better comprehension performance in the test immediately after reading the text (probability = .22, *SE* = .03) compared to the test after one week (probability = .19, *SE* = 0.03), *OR* = 0.81 (95% CI [0.68, 0.95]). However, this main effect was qualified by an interaction of retention interval and learning condition (β = 0.08, *SE* = 0.04, *z* = 1.92, *p* = .027, one-tailed, Figure 2.2).

Figure 2.2

Interaction of Learning Condition and Retention Interval in Experiment 1



Note. Learning outcomes (text comprehension) by learning condition at the short and long retention interval in Experiment 1 (estimated based on Model 1, back-transformed probability of a correct answer). Error bars represent standard errors (+/- 1 *SE*).

Planned comparisons revealed that the learning outcome of participants in the massed condition decreased from the immediate to the delayed test (immediate test: probability = .26, *SE* = .04; delayed test: probability = .19, *SE* = .03), z = 3.23, p =.001, *OR* = 0.68 (95% CI [0.55, 0.85]). In contrast, the learning outcomes of participants in the distributed condition remained stable (immediate test: probability = .19, *SE* = .03; delayed test: probability = .18, *SE* = .03), z = 0.67, p = .504, *OR* = 0.93 (95% CI [0.74, 1.16]). At the shorter retention interval, the participants in the massed condition outperformed participants in the distributed condition (z = -2.06, p = .040), *OR* = 1.45 (95% CI [1.02, 2.05]). However, at the longer retention interval, no difference between the two learning conditions emerged (z = -0.35, p = .728), *OR* = 1.07 (95% CI [0.75, 1.52]).

Additionally, we found the predicted main effect of prior knowledge ($\beta = 0.15$, $SE = 0.06$, $z = 2.52$, $p = .012$), indicating that text comprehension performance was positively associated with prior knowledge. A difference of one standard deviation in prior knowledge corresponded to an odds ratio of 1.16 (95% CI [1.03, 1.31]).

In sum, we did not find the benefit of distributed reading predicted by Hypothesis 1. On the contrary, at the immediate test an advantage of the massed condition was found. This advantage disappeared at the delayed test, but it did not turn into an advantage for the distributed condition. In the distributed condition, we found no decrease between the short and the long retention interval. Hypothesis 2, stating that prior knowledge would benefit learning from the text materials was supported by the data.

Metacognitive Judgments of the Learning Process. We estimated two multivariate linear regression models with the metacognitive judgments as dependent variables and learning condition (contrast-coded: distributed = 1, massed = -1), text (contrast-coded: first text = -1, second text = 1) and their interaction plus the domain (contrast-coded: biology = -1, physics = 1) as predictors. We used the R package car for the hypothesis tests (Fox et al., 2013). The variables self-predicted success and perceived reading difficulty were recoded so that higher values correspondent to higher difficulty and lower predicted success, corresponding to Hypothesis 3. Means were estimated from the corresponding univariate linear regression models using the R package lsmeans (Lenth, 2016).

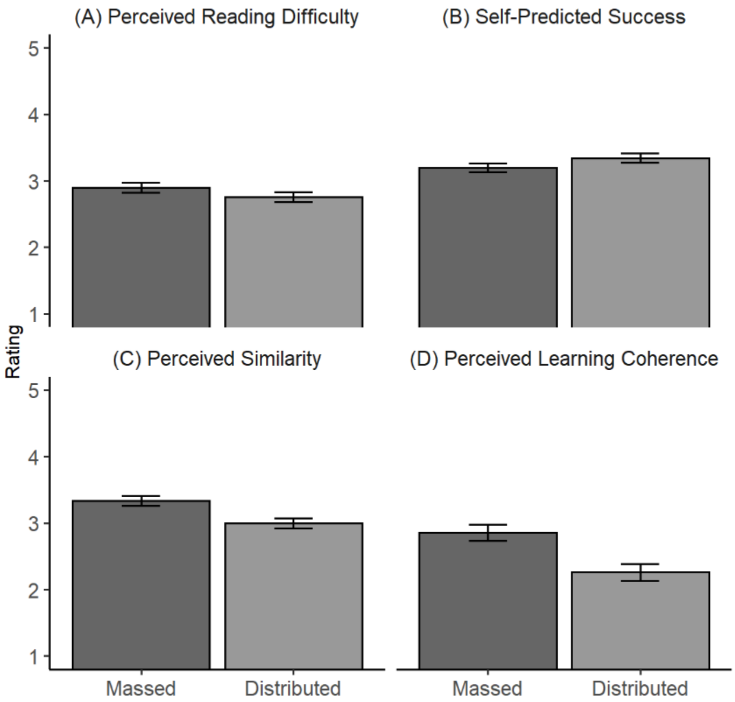
Meta-cognitive Judgments of Predicted Success and Reading Difficulty. The model revealed an effect of the learning condition, $F(2, 278) = 3.23$, $p = .041$. However, despite the higher difficulty and lower predicted success predicted by Hypothesis 3, in the distributed condition, participants in the distributed condition perceived reading as less difficult ($M = 2.75$, $SE = 0.07$), but predicted lower success ($M = 3.34$, $SE = 0.07$) than participants in the in the massed condition (perceived reading difficulty: $M = 2.90$, $SE = 0.07$; predicted success: $M = 3.20$, $SE = 0.07$, Figure 2.3). Please note that the univariate tests failed to reach significance (for both tests: $|t| > 1.32$ and $p > .065$, one-tailed) even though the multivariate analysis supports Hypothesis 3.

We found no interaction between learning condition and text, $F(2, 278) = 0.50$, $p = .606$). Thus, participants in the distributed condition perceived both texts as more difficult than participants in the massed condition.

Meta-cognitive Judgments of Perceived Similarity and Learning Coherence. The model revealed a significant effect of the learning condition $F(2, 137) = 9.13$, $p < .001$). In line with Hypothesis 4, participants in the distributed condition perceived less similarity and learning coherence between the texts (perceived similarity: $M = 3.00$, $SE = 0.08$; learning coherence: $M = 2.26$, $SE = 0.12$) than participants in the massed condition (perceived similarity: $M = 3.34$, $SE = 0.08$, $t(138) = -3.11$, $p = .002$; learning coherence: $M = 2.85$, $SE = 0.12$, $t(138) = -3.34$, $p = .001$; Figure 2.3).

Figure 2.3

Main Effect of Learning Condition on Meta-Cognitive Judgments of Learning in Experiment 1



Note. Estimated means of (A) perceived difficulty (recoded, 1: very easy; 5: very difficult), (B) predicted success (recoded, 1 = very good; 5 = very bad), (C) perceived similarity and (D) perceived learning coherence (1= not true; 5 = true) for the learning conditions, Experiment 1. Error bars represent standard errors (-/+ 1 *SE*).

Discussion

In Experiment 1, learning outcomes were not enhanced by distributed learning, not even at the longer retention interval. Nevertheless, learning outcomes decreased from the short- to the long-retention interval only in the massed but not in the distributed condition. Consequently, learning outcomes in the massed and the distributed condition were on par one week after the second text had been read. On the one hand, one interpretation of this pattern of effects is that distributed reading made the learning outcomes more stable. On the other hand, given we did not find a benefit of distributed reading and the performance in both conditions was very low in both groups, this pattern might also be the result of a bottom effect in the distributed condition. Thus, distributed learning might have been too difficult for participants to be desirable for long-term learning. Furthermore, it might be argued that distributed learning is confounded with the retention interval. Comprehension questions referred to Text 1, Text 2, or both texts. As participants read Text 1 one week earlier in the distributed condition, the short retention interval for questions that referred to this text was actually not five minutes but one week.

Consistent with previous studies, Experiment 1 showed that learning from text increased with higher levels of domain-specific prior knowledge (e.g., Schneider et al., 1989; Ozuru et al., 2009). The results regarding the meta-cognitive judgments of predicted success and reading difficulty were somewhat mixed, but seem to indicate that distributed learning changes the perceived difficulty in learning, with lower perceived difficulty during reading but, at the same time, lower predicted success of learning. Furthermore, the meta-cognitive judgments of perceived similarity and learning coherence suggest that distributed learning made coherence-building across texts subjectively more difficult for the learner, possibly by making the retrieval of information from the first text more effortful.

The main weakness of Experiment 1 is the potential confound between learning condition and retention interval. This confound is a design feature of distributed learning and introduces some ambiguity into the interpretation of results. One way to preserve the potential benefits of distributed reading and simultaneously eliminate the potential drawbacks is to use a shorter interval between the two texts. This possibility was explored in Experiment 2.

Experiment 2

Studies on the benefits of the retrieval practice effect have shown that retrieval success is essential for long-term retention. Thus, the low, albeit stable level of learning outcomes in the distributed condition might indicate that the interstudy interval of one week chosen in Experiment 1 was too long to enable successful retrieval of information from the first text. Therefore, we changed the lag between the two texts from one week to 15 min in Experiment

2. This relatively short lag might encourage active retrieval of the first text when reading the second text. The same hypotheses were tested as in Experiment 1. As in Experiment 1, we tested our hypotheses simultaneously with two different sets of texts from two domains.

Method

Design. Experiment 2 was based on a 2 x 2 x 2 design with the independent variables learning condition (massed vs. distributed learning), retention interval (immediate vs. one-week delay), and domain (biology vs. physics). Learning condition and domain were varied between participants, and retention interval was varied within participants. Participants were randomly assigned to one of the four resulting experimental groups. To minimize differences in learner characteristics between the experimental groups as in Experiment 1, we first formed homogeneous blocks of participants for each class matched according to prior knowledge and reading ability and then assigned participants from these groups randomly to the experimental conditions (randomized block design) within classes. All experimental conditions were realized in each class. The assignment of the two comprehension tests to the two levels of the factor retention interval was counterbalanced between participants.

Participants. Participants in Experiment 2 were 160 seventh graders (77 boys, 83 girls), with a mean age of 12.97 (*SD* = 0.44) from eight classes of different schools (Gymnasium and comprehensive schools). For all participating participants, parental permission was obtained. Participants without parental permission solved riddles instead of participating in the experiments (alternatively, they were allowed to visit parallel classes). As in Experiment 1, participants were randomly assigned to massed (*n* = 81) and the distributed condition (*n* = 79). Participants received sweets and a magic cube puzzle as a reward for participation.

Fourteen participants missed one of the two learning sessions and were excluded from analysis. Thirteen participants missed the domain-specific prior knowledge test, their data was excluded from analysis. The data of two participants was excluded because of technical problems, and one participant could not complete the experiment because of language issues.

In the end, the data of 130 participants were analyzed (68 in the massed and 62 in the distributed learning condition).

Power Analysis. As for Experiment 1, we estimated the power for a small interaction effect between learning condition and retention interval by simulation using the R package simr (Green & MacLeod, 2016). The power (1-β) for detecting a small effect (d = 0.3 or OR = 1.72) was high (1.000, 95% CI [.996, 1.000]) given the Type-I error probability (α) of .05.

Materials. The materials of Experiment 1 were used for Experiment 2 with slight changes. The low learning outcomes in Experiment 1 might be due to the fact that the texts were too difficult. Therefore, we revised the experimental texts by implementing more examples and illustrative metaphors. The Flesch reading ease changed only moderately (Biology 1: 57, Biology 2: 46, Physics 1: 43, Physics 2: 57).

The comprehension tests (two for each domain with multiple-choice and short-answer questions) remained unchanged (Cohen's κ = 84). . The mean item difficulty was .33 (SD = .20) for the short-answer questions and .52 (SD = .13) for the multiple-choice items.

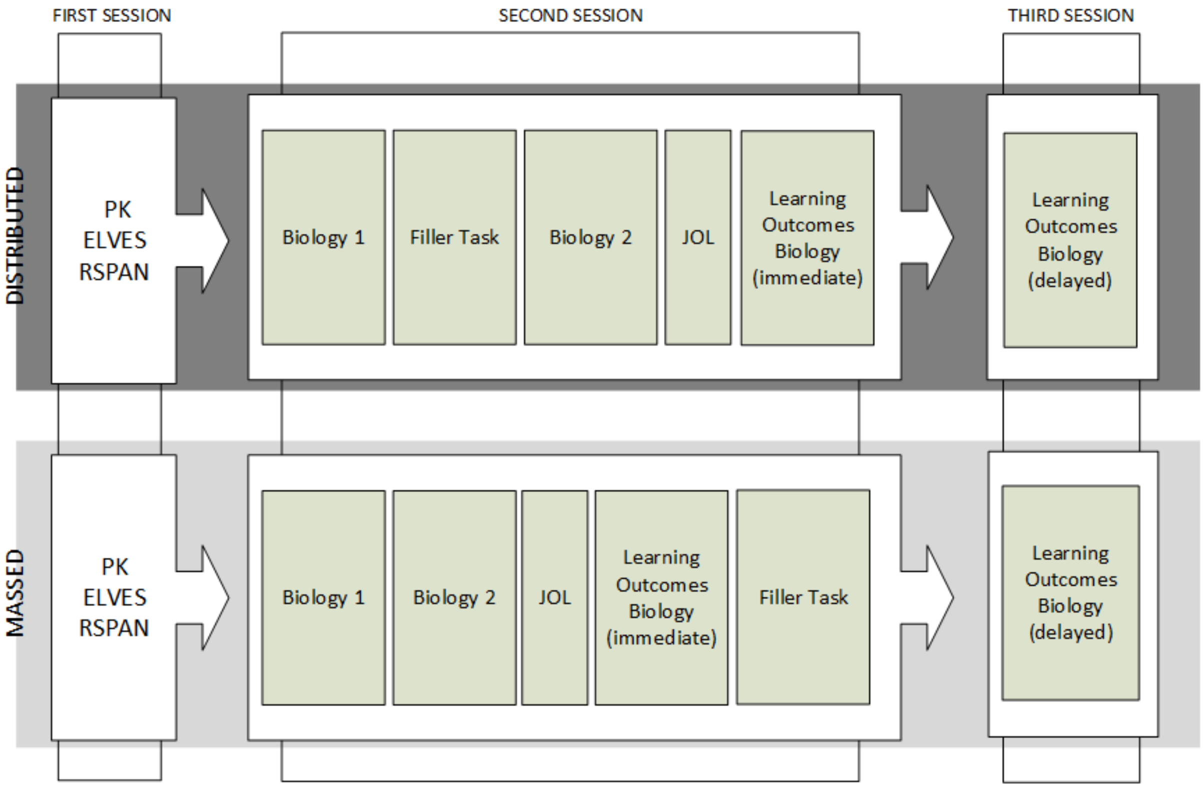
Reading ability (Cronbach's α = .75), knowledge about reading strategies, and working memory capacity (Cronbach's α = .85) were assessed with the same measures as in Experiment 1. In some classes, the RSPAN was aborted because of time issues. Therefore, for all participants, only the performance of the first 10 sequences was included in the analysis. In the prior knowledge test, one question from the physics part was dropped to equal the time spent on the physics test and the biology test (shortened physics test: Cohen's κ = .86).

Unlike Experiment 1, metacognitive judgments of reading difficulty and predicted learning success were provided only after reading the second text to create a massed condition without any disruption. The ratings of perceived similarity (Cronbach's α = .30) and learning coherence remained unchanged. The time allotted for the metacognitive judgments was set at 5 min, and filler questions were added at the end to ensure that all participants received the immediate test exactly 5 min after learning.

Procedure. The procedure of Experiment 2 matched that of Experiment 1 with slight changes (Figure 2.4). After the pretest in Session 1, the experimental groups were matched by prior knowledge and reading ability.

Figure 2.4

Procedure of Experiment 2 in the Biology Group



Note. The example was drawn from the biology group, the procedure was the same for the physics group with the respective texts for physics. PK = Assessment of prior knowledge; ELVES = Assessment of reading ability; RSPAN = Assessment of working memory; Biology 1 = first text biology [first text phyiscs]; JOL = assessment of meta-cognitive judgments of the learning process; Biology 2 = second text biology [second text physics]; Learning outcomes biology = assessment of learning outcome for biology [assessment of learning outcome for physics]. Learning outcomes were measured immediate after learning (immediate) and one week delayed (delayed).

In Session 2, participants read the two texts. In the distributed condition, reading was interrupted by a 15-min filler task between the first and second text, whereas in the massed condition, participants read the second text immediately after the first text without an intervening task. After reading the second text, the participants completed the assessment of the meta-cognitive judgments of the learning process, followed by the immediate assessment of learning outcomes.

The delayed assessment of learning outcomes was administered in Session 3, which was originally planned for one week after Session 1. However, this session was rescheduled in two classes. Thus, the length of the retention interval varied between 7-9 days.

Results

Differences Between Experimental Groups. No differences were found between the learning conditions in the teacher reported grades of the participants in biology ($W = 438$, $p = .410$), but slight differences were found in the physics grades ($W = 506.5$, $p = .047$; massed: $Min = 1$, $Q1 = 1$, $Mdn = 2$, $Q3 = 2$, $Max = 3$; distributed: $Min = 1$, $Q1 = 2$, $Mdn = 2$, $Q3 = 3$, $Max = 4$). Also, no differences in working memory capacity and reading ability were found between the learning conditions (working memory: $F(1,116) = 0.07$, $p = .788$; reading ability: $F(1,126) = 0.11$, $p = .738$). Additionally, no differences were found between the groups receiving the biology or the physics texts (working memory $F(1,116) = 1.61$, $p = .207$; reading ability: $F(1,126) = 0.77$, $p = .381$), indicating that the matching procedure was effective.

Table 2.3 provides the means and standard deviations of all dependent variables and learner characteristics observed in the four experimental groups. A correlation matrix (including the correlations within learning conditions/topics) of all dependent variables is provided in Appendix B (Table B2).

Table 2.3

Mean and Standard Deviation of Dependent Variables and Learner Characteristics in Experiment 2

	Biology				Physics			
	Massed		Distributed		Massed		Distributed	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Dependent variables								
Learning outcome (immediate)	6.41	3.28	6.37	3.85	6.50	2.77	5.12	2.81
Learning outcome (delayed)	5.67	2.97	6.21	3.79	5.41	3.04	5.04	2.41
Perceived difficulty	3.74	0.75	3.33	0.99	3.38	0.85	3.38	0.91
Self-predicted success	2.91	0.57	2.53	0.97	2.79	0.54	2.81	0.78
Perceived similarity	3.39	0.63	3.57	0.68	3.24	0.66	3.24	0.42

Perceived learning coherence	2.76	0.89	2.57	1.10	2.82	1.00	2.56	1.22
Learner characteristics								
Domain-specific prior knowledge	1.97	1.85	2.93	2.32	6.24	2.10	5.19	2.26
Grades	2.31	0.89	2.48	0.89	1.82	0.77	2.29	0.85
Reading ability	15.22	3.76	16.10	6.06	17.12	5.99	15.66	4.27
Working memory capacity	.66	.18	.71	.19	.76	.13	.69	.17

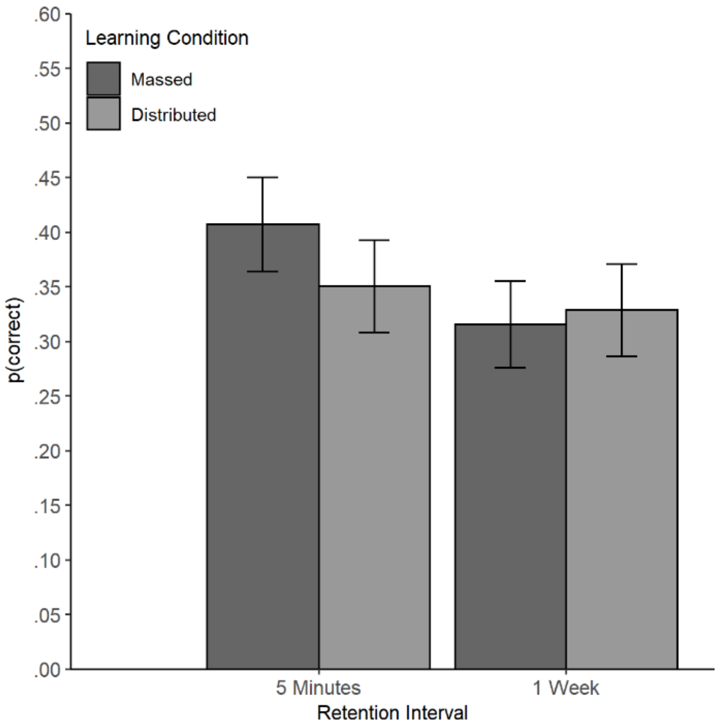
Note. All variables are reported separately for the domains as domain was varied between-subjects.

Learning Outcomes. To test Hypotheses 1 and 2, we estimated a generalized mixed model with learning condition (contrast-coded: distributed = 1, massed = -1) and retention interval (contrast-coded: immediate = -1, delayed = 1) with their interaction and domain-specific prior knowledge (z-transformed) as predictors. Text comprehension performance served as the dependent variable. In addition, we included domain (contrast-coded: physics = 1, biology = -1) as predictor for control purposes. Items and participants were included as random effects (random intercepts). The parameter estimates are provided in Table 2.4.

Paralleling the results of Experiment 1, there was no main effect of learning condition (β = -0.05, *SE* = 0.09, *z* = -0.54, *p* = .588). Participants in the massed condition (probability = .36, *SE* = .04) performed equally well as participants in the distributed condition (probability = .34, *SE* = 0.04), *OR* = 1.10 (95% CI [0.78, 1.55]). The model revealed a significant main effect of retention interval (β = -0.13, *SE* = 0.04, *z* = -3.19, *p* = .001), with better performance at the short retention interval (probability = .38, *SE* = 0.04) compared to the long retention interval (probability = .32, *SE* = 0.04), *OR* = 0.77 (95% CI [0.66, 0.91]). Again, this main effect of retention interval was qualified by an interaction with the learning condition (β = 0.08, *SE* = 0.04, *z* = 1.95, *p* = .025, one-tailed, Figure 2.5), as predicted in Hypothesis 1. Planned comparisons revealed that the performance of participants in the massed condition decreased between the two retention tests (immediate test: probability = .41, *SE* = 0.04; delayed test: probability = .32, *SE* = 0.04), *z* = 3.79, *p* < .001, *OR* = 0.66 (95% CI [0.53, 0.82]). In contrast, performance in the distributed condition did not decrease significantly (immediate test: probability = .35, *SE* = 0.04; delayed test: probability = .33, *SE* = 0.04), *z* = 0.84, *p* = .399, *OR* = 0.90 (95% CI [0.71, 1.14]. Contrary to Experiment 1, we found no benefit of massed learning in the test immediately after reading (*z* = -1.35, *p* = .179), *OR* = 1.29 (95% CI [0.89, 1.86]). However, even though the participants in the distributed condition showed no decrease, they did not perform better than the participants in the massed condition at the delayed test (*z* = 0.32, *p* = .746), *OR* = 0.94 (95% CI [0.64, 1.38]).

Figure 2.5

Interaction between Learning Condition and Retention Interval in Experiment 2



Note. Learning outcomes (text comprehension) by learning condition at the short and long retention interval in Experiment 2 (estimates based on Model 1, back-transformed probability (p) of a correct answer). Error bars represent standard errors (+/- 1 *SE*).

Consistent with the findings in Experiment 1 and Hypothesis 2, we found a significant positive effect of prior knowledge ($\beta = 0.61$, $SE = 0.11$, $z = 5.65$, $p < .001$). A difference of one standard deviation in prior knowledge corresponded to an odds ratio of 1.84 (95% CI [1.49, 2.27]).

In sum, we did not find the benefit of distributed reading for lasting learning that was predicted in Hypothesis 1. However, we found no detrimental effects of distributed reading at the immediate but still no decrease in the distributed condition between the immediate and the delayed test. Hypothesis 2 stating that domain-specific prior knowledge would benefit learning from the texts was again supported by the data.

Table 2.4

Parameter Estimates and Significance Tests for the Generalized Mixed Model for Learning Outcomes in Experiment 2

Fixed effects		
Predictors	Estimate	*SE*
(Intercept)	-0.65 ***	0.16
Learning condition	-0.05	0.09
Retention interval	-0.13 **	0.04
Prior knowledge (*z*-standardized)	0.61 ***	0.11
Domain	-0.46 **	0.17
Learning condition x Retention interval	0.08 +	0.04
Random effects		
σ^2	3.29	
τ_{00}	0.32 $_{\text{ID}}$	
	1.43 $_{\text{Item}}$	
Goodness of fit		
Deviance	4157.23	

Note. Learning condition (contrast-coded: distributed = 1, massed = -1). Retention interval (contrast-coded: immediate = -1, delayed = 1). Domain (contrast-coded: biology = -1, physics = 1). Prior knowledge was included z-standardized. Directional hypotheses were tested one-tailed.

* $p < .05$ (two-tailed).** $p < .01$ (two-tailed).*** $p < .001$ (two-tailed).

+$p < .05$ (one-tailed).++ $p < .01$ (one-tailed).+++ $p < .001$ (one-tailed).

Metacognitive Judgments of the Learning Process. We estimated two multivariate linear models with the metacognitive judgments as dependent variables and learning condition (contrast-coded: distributed = 1, massed = -1), *z*-standardized prior knowledge and their interaction, and domain (contrast-coded: biology = -1, physics = 1) as predictors. In both models, no significant main effects of learning condition were found (predicted success and reading difficulty: $F(2,126) = 0.93$, $p = .397$; perceived similarity and learning coherence: $F(2,126) = 1.01$, $p = .367$).

Discussion

Experiment 2 replicated the main findings of Experiment 1 but with a much shorter interstudy interval (15 min instead of one week). Distributed learning did not lead to better learning results, not even at the delayed test. Nevertheless, students in the distributed condition showed no decrease in learning outcome from the immediate to the delayed test, whereas performance of students in the massed condition did. This result is especially noteworthy because both experimental groups performed equally well at the test immediately after learning. Thus, contrary to Experiment 1, the fact that the learning outcomes in the distributed condition did not decrease from the immediate to the delayed test cannot be attributed to disadvantages due to a longer retention interval or a floor effect at the immediate test.

Contrary to the findings of Experiment 1, distributed learning was not associated with higher perceived difficulty and predicted success. Furthermore, no differences were found in perceived learning coherence and similarity between the learning conditions. Apparently, the 15 min interstudy interval in the distributed learning condition was not sufficiently to make learning subjectively more difficult. Nevertheless, distributed learning made a difference for the learning outcomes by slowing down the decrease in learning from the immediate test to the delayed test after one week.

General discussion

The present experiments addressed the question of whether distributed learning is beneficial for learning with multiple, complementary texts, especially in the long-term. The experiments were conducted with seventh graders who read two expository texts from the two domains of physics and biology. Experiment 1 implemented a long interstudy interval (one week), whereas Experiment 2 implemented a short interstudy interval (15 min). Both experiments showed a highly similar pattern of learning outcomes. Learning outcomes decreased from the immediate to the delayed test in massed but not in distributed reading. Nevertheless, participants who had read the texts in a distributed fashion performed no better than participants in the massed condition at the delayed test. Moreover, the results for metacognitive judgments show that participants in Experiment 1 perceived the text in distributed reading as less difficult and predicted that they would be less successful in recalling information from this text, and they perceived the two texts as less coherent and less similar compared to massed reading. This could indicate that the participants recognized the lower learning coherence and anticipated the detrimental effects on learning in the immediate test, but did not use these metacognitive judgments to engage more in reading, what might have been reflected in higher perceived difficulty and better learning outcomes. In contrast, no such effects occurred in Experiment 2 with the short interstudy interval of 15 min.

Inhibitors of Advantages of Distributed Learning

Overall, we found no support for our assumption that distributed reading benefits long-term retention. Given that distributed practice is effective even without repetition, some features of our experiments, or of distributed reading in general, might have reduced the benefits of distribution. First of all, long-term retention was measured one week after learning, which is a relatively short retention interval according to prior research that suggests retention intervals of four weeks or longer (Rohrer, 2015). As mentioned above, desirable difficulties might play out their advantages in particular at longer intervals between learning and assessment of learning outcomes (Pashler et al., 2007, Rawson & Kintsch, 2005, Rawson, 2012, see also Dunlosky et al., 2013). This might be also the case for distributed reading.

Another feature that might have reduced the benefits of distribution, is that we varied the retention interval within subjects. However, immediate test without feedback facilitates long-term learning more than a delayed retention test, provided that retrieval is successful (Karpicke & Roediger, 2010; S. Greving & Richter, 2018). If such a testing effect occurred in the present experiments, students in the massed condition might have been advantaged, especially in Experiment 1, because of their better recall success at the immediate test. This advantage might have led to an underestimation of the benefits of distributed reading. In future research, a between-subjects variation of the retention interval should be considered.

Additionally, distributed reading lacks (in)direct feedback. Research in retrieval practice has shown that feedback is crucial for retrieval practice effects with short-answer questions and after an incorrect response (Kang et al., 2007; Pashler et al., 2005). In distributed practice, the repetition of learning material provides not only an additional learning occasion but also serves as a kind of feedback for the retrieval of the first learning session. This indirect feedback might be essential to the benefits of distributed learning. For example, in the study by Vlach and Sandhofer (2012), the concept of food chains was repeated in each biome. Thus, even if the children failed to remember parts of the food chain of the biome explained in the last session, they had the opportunity to update the knowledge and then use the repeated presentation of the concepts as feedback for their retrieval and overall comprehension. However, in distributed reading, no such indirect feedback is provided.

An important distinction of our study is that we investigated distributed learning in a field setting with younger learners. The distributed practice effect is a robust finding with adult learners in laboratory settings, but few studies have investigated the extent that these findings generalize to younger learners and to learning in real-world educational settings (Küpper-Tetzel, 2014). However, in school contexts, even distributed practice might fail to benefit learning, for example due to a noisier learning environment (Goossens et al., 2016). A recent study that investigated the effects of rereading schedules in the same age groups arrived at conclusions very similar to those of the present study (C. E. Greving & Richter, 2019). In this study, seventh graders read a text about a bacterial cell twice. Rereading was implemented either in a massed fashion, with no interruption between reading and rereading, or in a distributed fashion, with a lag of one week between reading and rereading. Students rereading the texts in a distributed fashion predicted their recall to be lower, and indeed they showed lower recall and text comprehension performance, but only immediately after reading. One week later, no difference was found between the massed and distributed condition. Moreover, a lower decrease between tests occurred in the distributed condition. The performance remained stable in the distributed condition, but the performance between the immediate and delayed post-test decreased in the massed condition. However, distributed rereading appears to have no advantage over massed rereading even after a retention interval of one week. Given the positive results for distributed rereading with adults (e.g., Rawson & Kintsch, 2005), distributed rereading might be less effective for younger learners than for adult learners, possibly because of the lower comprehension skills of younger learners. A similar relationship might hold for distributed reading, as investigated in the present experiments. Thus, research with adult learners seems to be necessary to evaluate the effectiveness of distributed reading.

Limitations and Suggestions for Further Research

The results of the two experiments are consistent but need to be interpreted with a number of limitations in mind. One limitation of the research reported here is that we were able to examine the effects of only two lags, which differed considerably in length. In Experiment 1, we tried to implement an “authentic” lag of one week, which is quite typical for school environments with a weekly schedule. However, this design decision caused a potential confound of distributed vs. massed reading with the time between reading the first text and the retention test in the distributed condition. We implemented a lag that diminishes this confounding variable in Experiment 2 by choosing a lag that was in line with earlier research on rereading with very short lags (Glover & Corkill, 1987). However, the school setting put constraints on the length of this lag, and 15 min was the maximal lag we could implement within one session. The end result could have been that the long lag implemented in Experiment 1 was too long, whereas the short lag implemented in Experiment 2 was too short to produce beneficial effects of distributed reading. Further research should explore a wider range of lags between 15 minutes and one week.

A second limitation is that the retention interval of one week used in our experiments might have been too short to produce an advantage of distributed reading. A study with university students that was otherwise very similar to the present experiments used a two-week retention interval between the final learning session and the assessment of learning outcomes, with comparable results and still no evidence for an advantage of distributed over massed reading at the long retention interval (C. E. Greving & Richter, 2020). However, apart from the fact that university students differ in many ways from secondary school students, a systematic manipulation of retention intervals, including also intervals of several weeks, would be needed.

Finally, although all efforts were made to maximize the ecological validity of the study, for example by selecting a topic that might appear and typical expository texts, the learning situation still deviated in several respects from normal school learning. In particular, the individual results in the learning outcome measure did not count for students' grades. Thus, the study combined a highly demanding learning task with a low-stakes assessment, which is unusual in school learning and may have caused the students not to take the learning task seriously enough. The low accuracy in the learning outcomes measure suggests such an interpretation.

Conclusion

Despite these limitations, some cautious implications can be drawn. The distribution of learning episodes is often recommended as general learning strategy (Schwartz et al., 2011, Putnam et al., 2016). Refining this general recommendation, our research suggests that teachers who wish to use distribution of learning as didactic strategy should restrict the usage to materials which include at least some kind of repetition.

This research was the first to investigate distributed learning with multiple complementary texts. We found evidence that distributed learning might change memory, indicated by lower decrease between immediate and delayed test, but contrary to our expectations, it did not lead to better learning outcomes at a retention interval of one week. These results have practical import. Given well-established benefits of distributed practice, students are often advised to space their learning activities rather than massing them in one learning session. However, this advice is not always restricted to repetitive learning. As noted above, Schwartz et al. (2011, p. 10) defined distributed practice as “learning that is spread out across relatively long periods of time rather than massed all at once.” Similarly, in their study guide for college students, Putnam et al. (2016) give students the recommendation to engage in distributed learning because “by spacing your studying you will learn the material in less time than if you tried to cram all of your studying into the night before the test” (p. 655). But is distributed practice a good strategy even for non-repeated materials? To date, the research on distributed practice cannot provide a clear answer to this question. It seems that more research on distributed learning with complementary learning materials is needed before it can be recommended as an effective learning strategy to students and teachers. The research reported here can be seen as a first step in this direction.

Chapter III

Experiments 3 and 4[1]

Learning from Complementary Texts:
Is Distributed Reading a Beneficial Learning Strategy?

A version of this chapter is submitted as:

Greving, C. E., & Richter, T. (2020). *Learning from complementary texts: Is distributed reading a beneficial learning strategy?* Manuscript submitted for publication.

[1] Please note that within Chapter III, I will refer to Experiment 3 as Experiment 1 and Experiment 4 as Experiment 2, following the numeration in the submitted manuscript.

Learning from Complementary Texts:
Is Distributed Reading a Beneficial Learning Strategy?

Carla E. Greving & Tobias Richter

Abstract. In two experiments, we investigated whether distributed reading of complementary texts enhances long-term learning in undergraduate students. In Experiment 1, 97 participants read two complementary texts either in a massed fashion or distributed by 1 week. Learning outcomes were assessed with retention intervals of 5 minutes and 1 week after learning. We found an interaction between learning condition and retention interval. After 5 minutes, participants in the massed condition outperformed the distributed condition, whereas after 1 week, no differences were found. Learning outcomes were not moderated by prior knowledge. Experiment 2 replicated Experiment 1 with an additional retention interval 2 weeks after learning and a sample size of 119 participants. As in Experiment 1, learning outcomes depended on time of test. However, even 2 weeks after learning, no benefit of distributed reading was found. Furthermore, prior knowledge moderated the effect of distribution with detrimental effects of distribution for students with low prior knowledge. Thus, we showed that the effects of distributed reading depend on time of test. Distributed reading seems to be detrimental immediately after learning and for students with low prior knowledge.

What can students do to enhance their learning? One general advice is to distribute learning rather than cramming time available for learning into one session. This advice is ostensibly based on experimental evidence. For example, in a journal for K-12 teachers, Agarwal and Roediger (2018) state that "[s]*paced practice* boosts learning by spreading lessons and retrieval opportunities out over time so that new knowledge and skills are not crammed in all at once" (p. 9, but see also Putnam et al., 2016). Despite these general recommendations, the definition of—and most evidence for beneficial effects of—distributed practice is restricted to *repetitions* of learning material, e.g. repeating the same word pairs in vocabulary learning (e.g. Bloom & Shuell, 1981; for a review see Cepeda et al., 2006).

C. E. Greving and Richter (2021) investigated distributed reading of complementary texts in seventh graders. Distributed reading did not lead to better learning outcomes, not even at a delayed test one week after learning. However, whereas participants in the massed condition showed a significant decrease from the immediate to the delayed test, the learning outcomes in the distributed condition remained stable, leading to equal learning outcomes one week after learning. The experiment by C. E. Greving and Richter (2021) was conducted in the classroom with relatively young learners. In the present experiments, we investigated distributed reading of complementary texts in university students in a laboratory setting. Apart from these differences, Experiment 1 was a replication of the study by C. E. Greving and Richter (2021). Experiment 2 additionally used a longer interval between learning and test (i. e. retention interval). In the following, we will provide a brief overview of research on distributed practice and the underlying cognitive mechanisms. We will focus on the role of repetition in distributed practice and introduce distributed learning with complementary texts as a potential extension of distributed practice.

Distributed practice as memory booster

In learning of repeated materials, distributing the learning time over two or more learning sessions seems to yield beneficial effects for long-term retention. The positive effects of this learning strategy is referred to as distributed practice. The beneficial effects of distributed practice are well supported by many laboratory studies (Cepeda et al., 2006; Donovan & Radosevich, 1999).

The three major theories explaining the effects of distributed practice can be summarized as deficient processing theory (Hintzmann, 1974), contextual-variability theory (Glenberg, 1979) and study-phase retrieval theory (Thios & D'Agostino, 1976). The available research suggests that especially retrieval processes play a crucial role in distributed practice (Cepeda et al., 2006; Gerbier & Toppino, 2015). In their meta-analysis, Janiszewski et al. (2003) tested several hypotheses derived from different explanations of the distributed practice effect. Theories based on a retrieval explanation of distributed practice described the data more accurately than deficient processing and contextual variability theories.

The *study-phase retrieval theory* assumes that retrieval processes during the second presentation of an item account for the benefits of distributed practice (Thios & D'Agostino, 1976). The second presentation of the item stimulates the retrieval of the memory representation of its first presentation, which strengthens the memory trace of the item. In distributed practice, the first presentation needs to be retrieved from long-term memory, whereas in massed practice, the first presentation is still accessible from working memory. This retrieval strengthens the memory for the item. Furthermore, as the time between learning occasions (i.e., the *interstudy interval*) increases, the retrieval from memory becomes more difficult, which also might benefit learning in the long run (Bjork, 1975; Pyc & Rawson, 2009).

To retrieve an item from memory, it is not necessary to repeat this item but rather present a new item that elicits retrieval of the previous one. Thus, if an item reminds the learner of another, earlier presented item, this would also evoke retrieval of this first presented item (D'Agostino & DeRemer, 1973; Madigan, 1969, Benjamin & Tullis, 2010). Consequently, if study-phase retrieval is (one of) the crucial mechanism(s) of distributed practice, the beneficial effects of distributing learning materials should not be restricted to repeated materials. It is still an open question whether learning without repetition but with *coherent* learning materials promotes long-term learning analogous to distributed practice. We will refer to this kind of learning as *distributed learning*. We define coherent materials as learning materials in which no explicit repetition is incorporated, but the materials enable reminding of previously learned materials.

Evidence on distributed learning

Evidence on distributed learning is relatively scarce. Braun and Rubin (1998, Experiment 3) created a word list in which the first three letters of a word were repeated but not the word itself –for example, SHAllow and SHArp (example from Delaney et al., 2010). Participants' task was to indicate whether the current word was presented before, whether the first three letters of the current word were presented before, or whether the word was not presented before at all. In an immediate free recall and recognition test, Braun and Rubin found a spacing effect for both, the word presented first (SHAllow) and the word presented second (SHArp). Thus, the experiment by Braun and Rubin indicates that even distributed learning of non-repeated learning materials might be beneficial for learning, provided that the materials and learning tasks necessitate the retrieval of information presented earlier.

Furthermore, beneficial effects of distributed practice have also been shown for words presented twice but paired with different information at each presentation (D'Agostino & DeRemer, 1973), as for example different words (first presentation: speed – ENGINE, second presentation: valve – ENGINE, Madigan, 1969) or in different sentences (first presentation: The child swatted the BUMBLE BEE, second presentation: The cow ate the BUMBLE BEE).

In a more applied setting, Smith and Rothkopf (1984) distributed the presentation of an eight-hour statistics course. The statistics course consisted of four different lessons and was either presented on one day (massed) or distributed over four days with one lesson per day (distributed). Five days after the final lesson, the students who had visited the distributed course outperformed the students who had visited the massed course. Similar results were found for student’s conceptual understanding (Budé et al., 2011). However, both experiments were conducted in quasi-experimental settings, opening the door for alternative explanations.

In a field study, Theobald et al. (2018) investigated the learning patterns of students during a web-based course using log-file analysis. Students who distributed their study time to a greater extent performed better in the final exam than students who crammed most of the study time within the last week(s) before the final exam. Unfortunately, it was not measured which materials the students decided to study. Thus, it remains unclear to what extent the distribution involved repetitions and to what extent it was distributed learning with complementary materials.

Summarizing the theoretical assumptions and empirical results, it can be assumed that distributed learning might benefit retention even if no repetition takes place, but subsequent learning content serves as a *reminder* of the previously learned content. Therefore, learning with materials containing reminding cues that allow the retrieval of information learned earlier should also benefit from spacing.

Distributed learning with complementary texts

In research on text reading, such reminding processes are well known. In the landscape model of reading (van den Broek et al., 1996), a fluctuation pattern of activation is used to model the likelihood of reactivation. The reader strives not only for local coherence, i.e., coherence between nearby sentences, but also for global coherence, i.e., establishing a coherent representation of the information communicated in the whole text. To reach global coherence, sentences that are semantically or causally related to previous sentences re-activate the information from these sentences. These processes may occur passively as a spread-of-activation mechanism or through active retrieval of information. As Long and Chong (2001) summarize, the amount of activation depends on the relatedness of concepts, on how much the information has been elaborated in the text, and on the distance between information in the text (Albrecht & Myers, 1995, 1998; Albrecht & O'Brien, 1993).

In reading complementary texts, such as different chapters of a textbook, readers need to retrieve information from the previous text to be able to integrate and validate information across texts. In textbooks, earlier chapters usually provide background knowledge that is helpful if not necessary to understand later chapters. If (short) texts are read in direct succession, experiments by Beker et al. (2016) suggest that information from the previous texts can be activated through a passive resonance mechanism. Thus, readers seem to reactivate information from previous text automatically to reach a global coherent text representation. If the conditions for passive activation are not met, readers may still engage in the active retrieval of information, a process that requires cognitive effort and is likely to depend on readers standards of coherence (van den Broek et al., 2011).

To sum up, retrieval processes play a crucial role in comprehending complementary texts. Previously read texts may provide the reader with relevant background information, which might be activated automatically or through active, strategic retrieval. In analogy to distributed practice, a longer time between texts make this retrieval from memory more difficult, which could benefit learning.

C. E. Greving and Richter (2021) investigated distributed reading of complementary text materials in two experiments. Seventh graders read two texts either massed (with a very short interruption, Experiment 1, or no interruption, Experiment 2) or distributed by one week (Experiment 1) or 15 minutes (Experiment 2). Learning outcomes were measured immediately and one week later. In both experiments, distributed learning did not benefit learning. However, participants who read the texts distributed showed no decrease between the immediate and delayed test, while participants who read the texts massed showed a large decrease. Thus, distributed reading might have affected memory, but this effect did not translate into benefits in long-term retention.

The experiments of C. E. Greving and Richter (2021) have been conducted in the classroom with seventh-grade students. C. E. Greving and Richter (2019) investigated distributed *re*reading with seventh graders and find no benefits of distribution. Considering these similar results, it seems possible that younger learners profit less from distribution of text reading (involving repetition or not). Although previous research suggests that distributed *practice* (involving repetition) might also be effective for younger learners and in educational settings (Barzagar Nazari & Ebersbach, 2019; Bloom & Shuell, 1981; Gluckman et al., 2014; Grote, 1995; Sobel et al., 2011; Vlach, 2014; Vlach & Sandhofer, 2012; for a review and discussion see Küpper-Tetzel, 2014), it is questionable whether this also hold for distributed (re)reading.

To learn from texts, readers do not only need basic skills (as decoding or vocabulary), but also higher order comprehension skills (Rapp et al., 2007). It has been found that younger readers exhibit lower comprehension skills in general (Perfetti et al., 2005) and especially lower text-integration skills (Barth et al., 2015). For readers with low reading skills, it is especially important that the prior knowledge that must be integrated with incoming text information is easily accessible (Smith & O'Brien, 2016). However, the goal of distribution is lower accessibility of prior knowledge. Thus, lower reading skills of seventh graders in the experiments by C. E. Greving and Richter (2021) might have reduced potential effects of distributed learning.

Additionally, the young readers might even have lacked the necessary prior knowledge to comprehend and learn from the texts. To build a coherent situation model of the texts the reader has to retrieve appropriate prior knowledge (Kintsch, 1988, 1992; van den Broek et al., 1996). When the reader lacks this prior knowledge, the reader might not be able to make sense of the text, even if the text is objectively cohesive. In distributed learning, cohesion is already reduced, because the related texts are read separately, as opposed to massed reading where related texts are read in quick succession. Consequently, the amount of prior knowledge might moderate the effects of distributed learning, in analogy to the reverse cohesion effect (McNamara et al., 1996). As the participants in the experiments of C. E. Greving and Richter (2021) showed low performance even in the immediate test, it could be assumed that the participants indeed did not have the necessary prior knowledge to build a coherent situation model and thus profit from distributed learning.

Finally, as an additional detriment, the classroom setting might have introduced more noise and distraction, a factor that is discussed to reduce the effects of distributed practice (Goossens et al., 2016) and, hence, might also affect distributed learning from complementary texts. This idea is also supported by research showing that coherence monitoring, i.e., the process to detect incoherence, is negatively affected by a dual task (de Bruïne et al., 2021).

Thus, it seems reasonable that college students in a laboratory setting could benefit from distributed reading even though no such beneficial effects of distributed reading were found in school students. Furthermore, the amount of prior knowledge might moderate the effects of distribution.

Rationale of the Present Experiments

In the current work, we investigated whether distributed reading of complementary texts might be a beneficial didactical method for undergraduates in a laboratory setting. In two experiments, participants read four texts, two from the domain of biology and two from the domain of physics. For both domains, two complementary texts were developed. The second text was coherent with the first text and complemented but did not repeat information from the first text.

Participants were randomly assigned to one of two learning schedules. They either read the texts in a massed learning schedule or in a distributed learning schedule. In the massed condition, the participants read the second text closely after reading the first text. In the distributed condition, the students read the second text one week after reading the first text. Metacognitive judgments of learning were assessed after each text. In Experiment 1, learning outcomes were measured five minutes and one week after reading the second text. Experiment 2 was similar to Experiment 1 but the delayed test of learning outcomes was two weeks after reading the second text.

Following the notion of distributed reading as expansion of distributed practice, we examined the general assumption that distributed reading might be a useful didactical strategy for university students. We expected that distributed reading is beneficial for learning, but that its effect depends on the retention interval. As massed reading might even show benefits immediately after learning (C. E. Greving & Richter, 2021), we expected a disordinal interaction effect between learning condition and retention interval with a benefit of massed reading immediately after reading and a benefit of distributed reading one week later (Hypothesis 1).

Considering that prior knowledge is ascribed a pivotal role in all modern theories of text comprehension (McNamara & Magliano, 2009) and has been found to be a strong predictor of text comprehension (e.g., Schneider et al., 1989), we expected a positive relationship between the amount of domain-specific prior knowledge and learning outcomes (Hypothesis 2).

In addition, we further examined the exploratory research question whether differential effects of distributed practice for different levels of domain-specific prior knowledge would arise.

Experiment 1

In Experiment 1, we investigated distributed reading of complementary texts with an interstudy interval of one week and a retention interval of five minutes and one week.

Method

Participants. Ninety-seven undergraduates of the University of Kassel (63 women, 32 men, 2 missing information) with a mean age of 23.91 years (*SD* = 4.49) participated in the experiment for course credit or a monetary reward (23.30 €). Participants were randomly assigned to either the massed learning condition (*n* = 49), or distributed learning condition (*n* = 48). Six participants did not show up for one of the later sessions. Five participants of the distributed condition rescheduled a session, which changed the interstudy interval. The data of these participants was excluded from data analysis, resulting in a sample size of 86 participants (distributed condition: *n* =41; massed condition: *n* = 45).

Power Analysis. As C. E. Greving and Richter (2021) did not find a benefit of distributed reading, we based the power calculations on the assumption of a small population effect (d = 0.30 or OR = 1.72, respectively), following Cohen's conventions for effect sizes (Cohen, 1988). The power (1-β) for finding an interaction effect between learning condition and retention interval of this size, determined by simulation with the R package simr (Green & MacLeod, 2016), was high (1.000, 95% CI [.996, 1.000]) given the assumed Type-I error probability (α) of .05.

Text Materials. The texts materials were taken from previous studies on distributed reading and have originally been developed for secondary school students (C. E. Greving and Richter, 2019, 2021). Despite the fact that the present study was conducted with undergraduate students, the texts remained unchanged to keep the experiments as comparable as possible. For biology, the first text explained the plant cell and the second text explained the bacterial cell. For physics, the first text explained the law of conversation of energy while the second text explained the first law of thermodynamics. The length of the texts ranged from 504 to 633 words and the Flesch reading ease (German formula, Amstad, 1978) ranged from 46 to 60, which corresponds to a medium difficulty. The biology texts contained images illustrating the structure of the respective cell; these images were presented alongside the whole text. The image was added to enhance comprehension of the cell structure and to enhance the ecological validity of the text, as expository texts in biology usually contain images. The first physics text contained an illustrative image of an experiment by James Prescott Joules, which was explained in the text and presented next to the respective paragraph. See Appendix A for translations of the texts used in the experiment. Participants were told that the materials had been developed as school materials for younger learners and thus might seem easy, but that they still should take the learning serious in order to learn and remember as much as possible.

Assessment of Learning Outcomes. For each domain, two different test forms (Form A and B) with eight short-answer questions and seven multiple-choice questions were constructed. In multiple-choice-questions, participants were asked to select the right answer out of four. For example, one short-answer question was, *A bacteria cell does not have a cell nucleus. But where can you find the genome of the bacteria cell?* One multiple-choice questions was, *To which kind of organism does the bacteria cell belong?*, with the response options (a) *Prokaryots*, (b) *Eukaryots*, (c) *Plasmid*, and (d) *Organelle*. The questions could be answered based on information from both texts (12%), from Text 1 (45%) or from Text 2 (43%). The questions referring to Text 1 and 2 asked for information explicitly given in the respective text. The questions referring to information from both texts made intertextual inference necessary, such as comparing plant and bacterial cell. All answers to the short-answer questions were scored as either incorrect (0) or correct (1) by a rater who was blind to the experimental conditions. Thirty percent of the answers were double coded by an additional, independent rater (Cohen's κ = .75). In the instances of disagreement (13%), the rating provided by the rater who rated all answers was used. Questions asking for more than one piece of information (e.g. differences between bacterial and plant cell) were divided in sub questions to enable dichotomous coding. The mean item difficulty was .61 (*SD* = .26).

Assessment of Domain-Specific Prior Knowledge. For the biology assessment, participants answered five short-answer questions adapted from a prior knowledge test for medical students (von Dülmen et al., 2006; e.g. *How many base triplets are there?*), two basic short-answer questions about cells (e.g., *What is an animal or plant cell? Write down everything you now about cells.*) and two short-answer questions concerning biological everyday phenomena (e.g., *Why does salad gets mushy when dressing is added?*). Additionally, they received images of the cell structures of the bacterial and plant cell and were required to label the components of the cells. For the physics assessment, participants answered five short-answer questions adapted from von Dülmen et al. (2006; e.g. *Name two flow types of liquids.*), five basic short-answer questions about the topics of the text materials (e.g., *What does the term 'conversation' mean?*), and two short-answer questions concerning physical everyday phenomena (e.g., *Why does the sky appear to be blue?*). The order of domains and questions within domains were randomized. The answers were scored by two independent raters (Cohen's κ = .95).

Assessment of Control Variables. To control for equal abilities in the experimental groups, we assessed several learner characteristics. We asked the participants whether they had visited a basic course, advanced course or no course in biology and physics during their last two school years. In addition, we requested them to provide their Abitur examination grade (ranging from 1 = "very good" to 6 = "unsatisfactory", according to the German grading system). In addition, we assessed knowledge about reading strategies with an adapted version of the Würzburger Lesestrategie Test (WLST; Würzburg Reading Strategy Test; Schlagmüller & Schneider, 2007; split-half reliability: $r = .90$, estimated in a sample of 4490 students in Grades 7-11). Reading ability was assessed with the subtest sentence verification of the German-speaking test of reading abilities ELVES (Richter & van Holt, 2005; Cronbach's $\alpha = .91$, computed in the current sample) and working memory with a computerized version of a Reading Span Task (RSPAN; Oberauer et al., 2000; Cronbach's $\alpha = .82$, computed in the current sample).

Additional Variables. In addition to the descripted variables, meta-cognitive judgements and reading times were assessed. As the current paper focuses on the learning outcomes, these variables are not reported here.

Procedure. All materials were presented on notebook computers with a 16.6“ screen. The experiment was implemented with the program Inquisit 3 (Version 3.0.6.0, 2011).

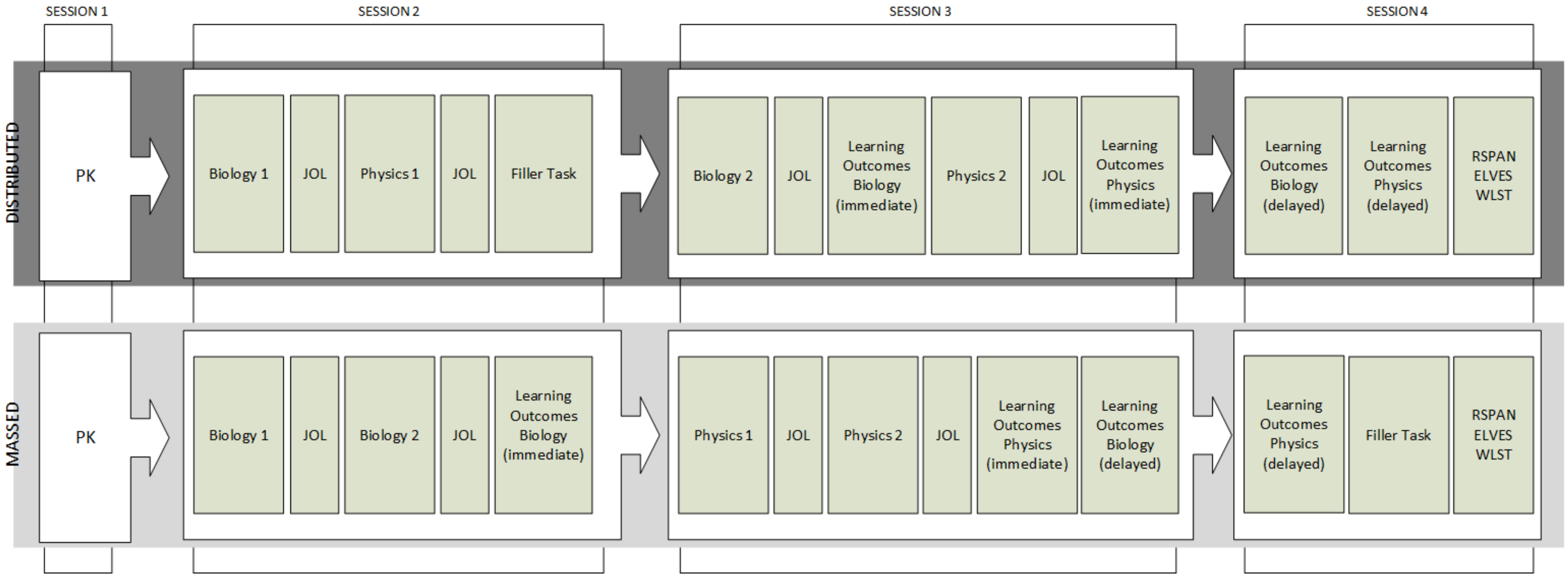
Experiment 1 consisted of four sessions. In Session 1, participants completed the assessment of domain specific prior knowledge (this assessment was combined with an unrelated experiment for practical purposes). Session 2 and Session 3 included the learning phases (for detailed information about the procedure, see Figure 3.1). In Session 4, the participants answered the final test of learning outcomes and completed the ELVES, RSPAN and WLST tests.

Design. Experiment 1 was based on a mixed 2x2x2 design with the independent variables learning condition (massed vs. distributed by one week, between-subjects), retention interval (five minutes vs. one week, within-subjects), and domain (biology vs. physics, within-subjects). The participants were randomly assigned to the experimental groups.

Data Analysis. We used generalized linear mixed effect models (GLMMs) with a logit-link. The models were estimated with the R packages lme4 (Bates et al., 2015), lmerTest and lsmeans (Lenth, 2016) in the R environment. GLMM were chosen to account for the multilevel structure of the data (persons and items), which, in this case, calls for a model with crossed random effects (Baayen et al., 2008). Participant and item were included as random effects (random intercepts) if the ICC of the dependent variable exceeded .05. For the interpretation of GLMM results, the predicted probabilities (back-transformed from the log odds) are reported. For all hypothesis tests, Type-I error probability was set to .05. Model assumptions were inspected visually with the package sjPlot (Lüdecke, 2018).

Figure 3.1

Procedure of Experiment 1



Note. Order of topics was counterbalanced between students (only one example is shown here). PK = Assessment of prior knowledge; ELVES = Assessment of reading ability; RSPAN = Assessment of working memory; WLST = Assessment of reading strategy knowledge; Biology 1 = first text biology; JOL = assessment of metacognitive judgments of the learning process; Physics 1 = first text Physics; Biology 2 = second text biology; Physics 2 = second text physics, Learning outcome physics = assessment of learning outcome for physics, Learning outcome biology = assessment of learning outcome for biology. Learning outcomes were measured immediate after learning (immediate) and one week delayed (delayed).

Results

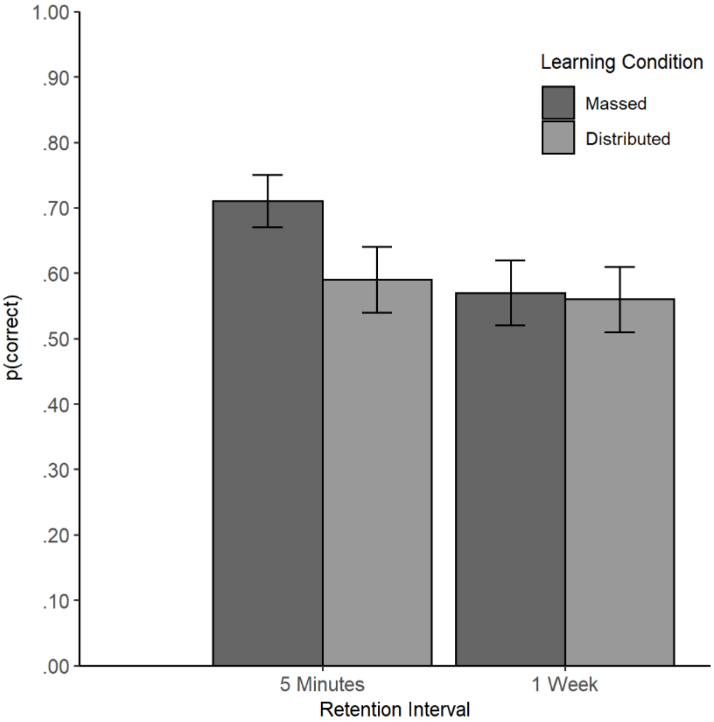
Group differences. The learning conditions did not differ with regard to the Abitur examination grades (t (77.90) = -0.03, p = .975), prior attendance of intense courses (biology: $\chi^2(2, N = 84) = 0.92$, p =.632, physics: $\chi^2(2, N = 84) = 0.49$, p =.784), working memory capacity (t(81.64) = -0.23, p = .815), reading ability (t(74.61) = -0.23, p = .817), knowledge of reading strategies (t(82.40) = -0.67, p = .514), or prior knowledge (biology: t(78.56) = -0.42, p = .675, physics: t(83.68) = -1.77, p = .080).

Learning Outcome. Two nested generalized linear mixed models were estimated. In the first step, learning condition (contrast coded: distributed = 1, massed = -1) and retention interval (contrast coded: immediate = -1, delayed = 1), the interaction of both and domain (contrast coded: biology=-1, physics=1) were included as predictors and the score as dependent variable. In the second step, we included the variable prior knowledge (z-transformed) and its interaction with learning condition and retention interval as additional predictors. In the following, we will report only the central results, for all estimates see Table 3.1.

Model 1. As predicted by Hypothesis 1, the model revealed a significant interaction between the learning condition and the retention interval (β = 0.12, SE = 0.03, 95% CI [0.06, 0.18], z = 3.91, p < .001). Planned comparisons revealed that this interaction was driven by a benefit of the massed condition after five minutes (M = .71, SD = .05, 95% CI [.63, .78]) compared to the distributed condition (M = .59, SE = .04, 95% CI [.50, .68]), z = -2.97, p = .003, OR = 1.73, 95% CI [1.21, 2.49] (Figure 3.2). After one week, participants in the distributed condition (M = .56, SE = 0.05, 95% CI [.46, .65]) performed as well as participants in the massed condition (M = .57, SE = 0.05), 95% CI [.49, .66], z = -0.46, p = .648, OR = 1.09, 95% CI [0.76, 1.56]. Thus, participants in the massed condition showed a significant decrease between the tests (z = -7.17, p < .001, OR = 0.55, 95% CI [0.47, 0.65]), whereas participants in the distributed condition did not (z = -1.59, p = .113, OR = 0.87, 95% CI [0.74, 1.03]).

Figure 3.2

Interaction of Learning Condition and Retention Interval in Experiment 1.



Note. Learning outcomes (text comprehension) by learning condition at the short and long retention interval in Experiment 1 (estimated based on Model 1, back-transformed probability of a correct answer,). Error bars represent standard errors ($+/-$ 1 SE).

Model 2. The second model paralleled the results of the first model. Additionally, we found the expected main effect of domain-specific prior knowledge predicted in Hypothesis 2 (β = 0.42, SE = 0.05, 95% CI [.33, .51], z = 9.28, p < .001). Higher prior knowledge was associated with higher scores in the short-answer questions, a difference of one standard deviation in prior knowledge corresponded to an odds ratio of 1.53 (95% CI [1.40, 1.67]). However, we did not find an interaction between learning condition and domain-specific prior knowledge (β = 0.05, SE = 0.04, 95% CI [-.04, 0.14], z =1.09, p = .276). To summarize, in discordance to Hypothesis 1, we did not find the expected beneficial effects of distributed reading for the long-term performance in short-answer questions. On the contrary, we found a disadvantage of the distributed condition at the test immediately after reading. However, this disadvantage disappears at the longer retention interval. As predicted in Hypothesis 2, we found a positive relationship between prior knowledge and success in the short-answer questions. Regarding our explorative research question, we have to conclude that the effects of distributed learning seems to be independent from the domain-specific prior knowledge.

Table 3.1

Parameter Estimates and Significance Tests for the General Mixed Model for Learning Outcomes in Experiment 1

	Model 1		Model 2	
Fixed Effects	β	*SE*	β	*SE*
(Intercept)	.46 **	.17	.45 **	.16
Learning condition	-.16+	.09	-.10	.07
Retention interval	-.18 ***	.03	-.19 ***	.03
Domain	-.42 **	.15	-.42 **	.15
Learning condition*retention interval	.12 ***	.03	.12 ***	.03
Prior knowledge (z-standardized)			.42 ***	.05
Learning condition*prior knowledge (z-standardized)			.05	.04
Retention interval*prior knowledge (z-standardized)			.04	.03

Learning Condition*Retention interval*prior knowledge (z-standardized)			-.03	.03

Random Effects

σ^2	3.29	3.29
τ_{00}	0.57 $_{\text{subject}}$	0.33 $_{\text{subject}}$
	1.65 $_{\text{Item}}$	1.67 $_{\text{Item}}$
Marginal R^2 / Conditional R^2	0.04 / 0.43	0.08 / 0.43
Deviance	7197.98	7115.23

Note. Learning condition (contrast-coded: distributed = 1, massed = -1). Retention interval (contrast-coded: immediate = -1, delayed = 1). Domain (contrast-coded: biology = -1, physics = 1). Prior knowledge was included *z*-standardized. Directional hypotheses were tested one-tailed.

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

$^{+}p < .05$, $^{++}$ $p < .01$, $^{+++}$ $p < .001$ (one-tailed).

Discussion

In Experiment 1, we replicated the findings of C. E. Greving and Richter (2021). We found the same interaction between learning condition and retention interval: At the immediate test, distributed practice seems to have detrimental effects on performance, whereas one week later, no difference between the learning conditions was found. For participants in the massed condition, performance decreased between the texts, whereas for participants in the distributed condition, no such decrease was found. In sum, similar to C. E. Greving and Richter (2021), Experiment 1 showed that the effects of distributed reading depend on time of test. However, it remains unclear whether distributed reading alters memory or only hinders performance immediately after learning. In analogy to distributed practice, we would assume that the benefits of distributed learning would appear in the long run. As a retention interval of one week might not be long enough to show long-term effects, in Experiment 2 a longer retention interval of two weeks was implemented.

Experiment 2

Experiment 2 and its hypotheses have been preregistered (C. E. Greving & Richter, 2017). It extended the design of Experiment 1 with an additional delayed test of learning outcomes two weeks after learning. The hypotheses were the same as in Experiment 1.

Method

Participants. One-hundred-and-nineteen (89 women, 29 men, 1 other) undergraduates of the University of Kassel, 23.26 years old (*SD* = 3.93), participated in the experiment for course credit or monetary reward (25 €). Participants were randomly assigned to the massed (*n* = 60) or the distributed learning condition (*n* = 59). Two participants did not show up for one of the later sessions. Two participants rescheduled their appointments, resulting in altered distribution intervals. Therefore, their data was excluded from analysis. Additionally, three participants had to be excluded due to errors in the data collection, resulting in a sample size of 112 data sets included in the analyses (distributed condition: *n* =54; massed condition: *n* = 58).

Power Analysis. We preregistered a sample size of 60 participants per cell, which seems to be sufficient for unbiased estimates and standard errors (Maas & Hox, 2005). In a post-hoc analysis, we determined the power for finding a small effect (represented by an estimate of .05) by simulation with the R package simr (Green & MacLeod, 2016). The power to find an effect of this size was high (1.00, 95% CI [.99, 1.00]).

Assessment of Learning Outcomes. For Experiment 2, three tests (A, B and C) were constructed, each containing five short-answer questions and three multiple-choice-questions. The tests contained the same questions as in Experiment 1, but the three test forms were parallelized regarding the item difficulty assessed in Experiment 1. Two of five short-answer questions were presented repeatedly on each retention interval. The order of tests was counterbalanced between the conditions. All answers to the short-answer questions were scored as either incorrect (0) or correct (1) by two independent raters who were blind to the experimental conditions (Cohen's $\kappa = .79$, estimated for 30% of the answers). Questions asking for more than one piece of information (e.g. differences between bacterial and plant cell) were divided in sub questions to enable dichotomous coding. The mean item difficulty was .29 (*SD* = .21).

Assessment of Domain-Specific Prior Knowledge. Participants answered the same questions as in Experiment 1 to assess their domain specific knowledge in biology and physics. The order of domains and questions was randomized. The answers were scored by two independent raters, Cohen's $\kappa = .67$.

Assessment of Control Variables. As in Experiment 1, the participants were asked which school courses they had visited in biology and physics. Moreover, the RSPAN (Cronbach's $\alpha = .90$ computed in the current sample), WLST and ELVES (Cronbach's $\alpha = .91$ computed in the current sample) were assessed.

Procedure. The procedure of Experiment 2 was identical to Experiment 1, except for one additional session when the third test of learning outcomes took place. This fifth session was scheduled one week after the fourth session. A second difference to Experiment 1 was that the assessment of the learner characteristics (RSPAN, ELVES, WLST, prior knowledge) took place in the first session.

Design. Experiment 2 was based on a mixed 2x2x3 Design with the independent variables domain (biology vs. physics, within-subjects), distribution (massed vs. distributed by one week, between-subjects), retention interval (five minutes vs. one week vs. two weeks, within-subjects). The participants were randomly assigned to the experimental groups.

Data Analysis. Two nested generalized linear mixed models were estimated. In the first step, learning condition (contrast coded: distributed = 1, massed=-1) and retention interval (as Helmert contrasts, with one contrast testing the one week retention interval against two week interval: five minutes = 0, one week = -1, two weeks = 1; and the other contrast testing the short retention interval against both long retention intervals: five minutes = 2, one week = -1 and two weeks = -1), the interaction of both contrasts and the domain (contrast coded: biology=-1, physics=1) as well as the performance in the ELVES reading skills test (*z*-transformed) were included as predictors and the relative score as dependent variable. In the second step, we included the variable prior knowledge (*z*-transformed) and its interaction with learning condition and retention interval as additional predictors. In the following, we will only report the central results, for all estimates see Table 3.2.

Results

Group Differences. The learning conditions did not differ with regard to the abitur examination grades ($t(104.09) = -1.78$, $p = .077$), visits of school courses (Biology: $\chi^2(2, N = 112) = 1.43$, $p = .489$, Physics: $\chi^2(2, N = 112) = 1.62$, $p = .559$), working memory capacity ($t(107.73) = -0.13$, $p = .896$), knowledge of reading strategies ($t(109.62) = -0.58$, $p = .561$) or prior knowledge (Biology: $t(106.06) = -0.24$, $p = .810$, Physics: $t(102.44) = 0.16$, $p = .870$). However, a small, but non-significant difference in the reading ability was found between conditions ($t(108.87) = -1.93$, $p = .056$). Even though this difference was not significant at a Type-I error probability of .05, we decided to control for reading ability in all models, to prevent a bias by this potential confound.

Table 3.2

Parameter Estimates and Significance Tests for the General Mixed Model for Learning Outcomes in Experiment 2.

	Model 1		Model 2	
Fixed effects	β	*SE*	β	*SE*
(Intercept)	0.23	0.17	0.25	0.17
Learning condition	-0.07	0.09	-0.08	0.07
Retention interval (1)	-0.05	0.03	-0.05	0.03
Retention interval (2)	0.25 ***	0.02	0.25 ***	0.02
Domain	-0.53 ***	0.15	-0.55 ***	0.15
ELVES, z-standardized	0.55 ***	0.09	0.37 ***	0.08
Learning condition*retention interval (1)	-0.03	0.03	-0.03	0.03
Learning condition*retention interval (2)	-0.07 ***	0.02	-0.06 **	0.02
Prior knowledge (z-standardized)			0.53 ***	0.05
Learning condition*prior knowledge (z-standardized)			0.18 ***	0.05
Retention interval (1) *prior knowledge (z-standardized)			0.01	0.04

Retention interval (2) *prior knowledge (z-standardized)	-0.01	0.02
Learning Condition*Retention interval (1)*prior knowledge (z-standardized)	-0.04	0.04
Learning Condition*Retention interval (2)*prior knowledge (z-standardized)	0.02	0.02

Random Effects

σ^2	3.29	3.29
τ_{00}	0.75 $_{\text{subject}}$	0.50 $_{\text{subject}}$
	1.15 $_{\text{Item}}$	1.17 $_{\text{Item}}$
Marginal R^2 / Conditional R^2	0.12 / 0.44	0.17 / 0.45
Deviance	8587.52	8409.14

Note. Learning condition (contrast-coded: distributed = 1, massed = -1). Retention interval (1) (5 minutes = 0, 1 week = -1, 2 weeks = 1).Retention interval (2) (5 minutes = 2, 1 week = -1 and 2 weeks = -1). Domain (contrast-coded: biology = -1, physics = 1). Prior knowledge and ELVES were included z-standardized. Directional hypotheses were tested one-tailed.

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

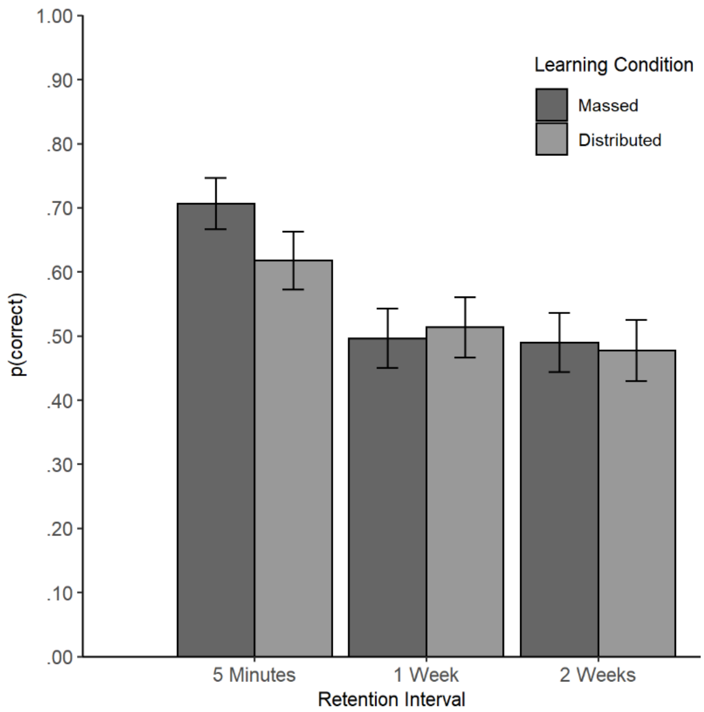
$^{+}$ $p < .05$, $^{++}$ $p < .01$, $^{+++}$ $p < .001$ (one-tailed).

Model 1. We found a significant interaction between the learning condition and the comparison between the short and the longer retention intervals (β = -0.07, *SE* = 0.02, 95% CI [-.11, -.03], z = -3.59, $p < .001$), whereas no interaction was found between learning condition and the comparison between the one and the two week retention intervals (β = -

0.03, *SE* = 0.03, 95% CI [-0.10, 0.03], *z* = -0.81, *p* = .418) (Figure 3.3). However, the interaction was in discordance with the pattern predicted by Hypothesis 1. Contrary to the results in Experiment 1, participants in both conditions showed a significant decrease between the immediate test (massed: *M* = .71, *SE* = .04, 95% CI [.63, 0.78]; distributed: *M* = .62, *SE* = .05, 95% CI [.53, 0.71]) and the test one week delayed (massed: *M* = .50, *SE* = .05, 95% CI [.41, 0.59]; *z* = -9.22, *p* < .001, *OR* = 2.54 (95% CI [2.11, 3.05]; distributed: *M* = .51, *SE* = .05, 95% CI [.42, 0.61]; *z* = -4.67, *p* < .001, *OR* = 1.58 (95% CI [1.31, 1.91]) and two weeks delayed (massed: *M* = .49, *SE* = .05, 95% CI [.40, .58]; *z* = -9.77, *p* = <.001, *OR* = 2.66 (95% CI [1.22, 3.19]; distributed: *M* = .48, *SE* = .05, 95% CI [.39, 0.57], *z* = -6.14 *p* = <.001, *OR* = 1.84 (95% CI [1.52, 2.23]). Between the tests that were delayed by one and two weeks, no decrease in performance was found in both learning conditions (for both |*z*| < 1.59, *p* > .809, OR < 0.95). Follow-up contrasts showed no difference between the learning conditions at any retention interval (for all |*z*| < 2.15, *p* > .439, OR <1.52).

Figure 3.3

Interaction of Learning Condition and Retention Interval in Experiment 2



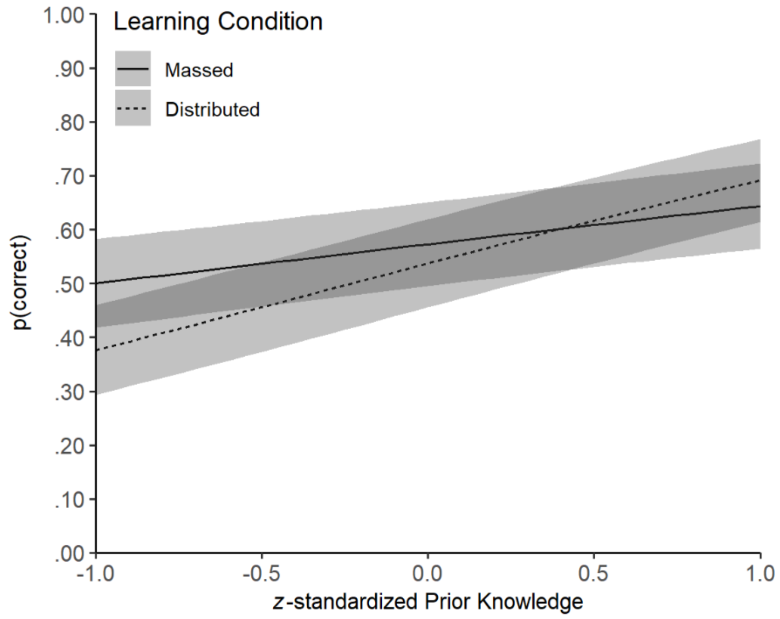
Note. Learning outcomes (text comprehension) by learning condition at the three retention intervals in Experiment 2 (estimated based on Model 1, back-transformed probability of a correct answer,). Error bars represent standard errors (+/− 1 *SE*).

Model 2. The results of Model 2 paralleled the results of Model 1 with an additional main effect for prior knowledge, as predicted by Hypothesis 2 (β = 0.53, *SE* = 0.05, 95% CI [.43, .61], $z = 11.32$, $p < .001$). A difference of one standard deviation in prior knowledge corresponded to an odds ratio of 1.70 (95% CI [1.55, 1.86]). Contrary to the results of Experiment 1, we additionally found a significant interaction between learning condition and prior knowledge (β =0.18, *SE* =0.05, $z = 3.99$, $p < .001$, 95% CI [0.10, 0.29]) (Figure 3.4). To interpret this relationship, comparisons for participants with low (-1 *SD*), medium (0 *SD*) and high (+ 1 *SD*) prior knowledge were calculated. With low prior knowledge, distributed learning seems to have detrimental effects on learning, leading to lower performance in the distributed condition (*M* = .38, SE = .04, 95% CI [.29, .46]) than in the massed condition (*M* =

.50, *SE* = .04, 95% CI [.41, 0.58], z = -3.10, p =.005, *OR* = 0.59, 95% CI [0.42, 0.83]). However, with medium (distributed: M = 0.54, *SE* = .04, 95% CI [.46, .62]; massed: M = .57, *SE* = .03, 95% CI [.50, .65]) and high prior knowledge (distributed: M = .69, *SE* = .04, 95% CI [.61, .77], massed: M = .65, *SE* = .04, 95 % CI [.57, .73]), no differences were found (for all $|z| < 1.12$, $p > .218$).

Figure 3.4

Interaction of Learning Condition and Domain-Specific Prior Knowledge in Experiment 2



Note. Learning outcomes (text comprehension) as back-transformed probability of a correct answer estimated as a function of learning condition and prior knowledge in Experiment 2, Model 2. Shaded areas around each line represent standard errors (−/+ 1 *SE*) for the point estimates at different levels of prior knowledge.

Discussion

The results of Experiment 2 parallelized the results in Experiment 1. Contrary to our assumption, that the detrimental effects of distributed learning at the short retention interval would turn into beneficial effects at the longest retention interval, we still found no differences between the two learning conditions. Thus, Hypothesis 1 had to be rejected in

Experiment 2, too. Furthermore, the participants showed a decrease between the short retention interval and the one week retention interval, but not between one week and two weeks after learning.

The results support Hypothesis 2 that predicted prior knowledge is positively correlated with learning outcomes. We also found an interaction between learning condition and prior knowledge, indicating that distributed learning might be detrimental for learners with low prior knowledge.

General Discussion

In our experiments, we addressed the question whether distributed reading of complementary texts is an effective didactic strategy for university students. In both experiments, we replicated the finding of C. E. Greving and Richter (2021) that the effect of distributed reading depends on time of test. In an immediate test, participants in the massed condition outperformed participants in the distributed condition, whereas no difference was found one or two weeks later. However, we did not find any beneficial effect of distributed reading, not even two weeks after learning. Nevertheless, the dependency of the learning outcomes of the time of test show that it is necessary to consider delayed measurements of learning outcomes to evaluate the effects of didactical measures, as the results can differ. This finding is consistent with the theoretical distinction between performance and learning (Soderstrom & Bjork, 2015) and aligns with other research regarding desirable difficulties in learning (R. A. Bjork & Bjork, 2020).

In Experiment 2, we also found an interaction between learning condition and prior knowledge. Participants with low prior knowledge showed a disadvantage if they read the texts in a distributed instead of a massed fashion. Different mechanisms might explain this effect. First, research has shown that readers with low prior knowledge are less likely to draw elaborative inferences in natural reading conditions (Calvo et al., 2003). Furthermore, more bridging inferences were found for readers with higher vocabulary knowledge (Dixon et al., 1988, Singer et al., 1992) and the amount of readers prior knowledge seems to affect whether readers generate global inferences (Finch-Kiefer, 1992). A lower likelihood of drawing inferences in low-knowledge readers might be especially detrimental in distributed reading, where bridging inferences require even more active processing than in massed reading. Second, we also found that prior knowledge is associated with test performance, a finding that is consistent with prior research (Ozuru et al., 2009; Schneider et al., 1989; Simonsmeier et al., 2021). Consequently, participants with low knowledge might have learned less from reading the first text and thus not have been able to retrieve the information during reading of the second text, which hindered learning from both. However, we did not find a benefit for distributed reading for readers with high prior knowledge.

Statistical power was very high in both experiments. Consequently, even a small positive effect of distributed reading on the learning outcomes, if it exists in the population, should have been detected. Thus, it seems warranted to raise the question why we did not find any beneficial effects. We assumed that distributed learning in analogy to distributed practice would lead to a more difficult retrieval of the information given in the first text. Because of the longer time between the texts, passive retrieval of previously read information by spreading-activation becomes less likely, which means that learners have to rely on coherence-based retrieval (Van den Broek et al., 1996). The amount of activation of text and background information depends on several features as the similarity between the concepts, the causal relatedness, and the elaboration or distance of the information (Albrecht & Myers, 1995, 1998; Long & Chong, 2001; Albrecht & O'Brian, 1993). Thus, even if the text materials were constructed in a way to enhance retrieval and activation of concepts of the first text during reading the second text, it is possible that these concepts were not retrieved. Most of the university students who participated in our experiments should be used to learning from text and, thus, their ability to integrate even less accessible information during text reading should be higher compared to less selective groups of learners, such as the secondary school students in the experiments by C. E. Greving and Richter (2021). However, one key factor in the active retrieval of information are the readers' standards of coherence, thus, the extent to which the reader strives for (global) coherence (van den Broek et al., 2011). Some properties of our experimental setting might have influenced the standards of coherence and consequently the amount of coherence-based retrieval during reading. First, we chose to present texts that were developed for school students. Even if the participants have been informed that they should take the materials still as serious as possible, this might have led to different reading patterns. For example, those texts might have been perceived as easy and familiar, leading to higher feeling of coherence. Second, the overall motivation to learn from the texts might be reduced in this experimental setting. The contents of the experimental texts were largely unrelated to the contents of participants' study program. Participants received a monetary reward regardless of their learning success, thus, no external reward was provided to increase learning effort. In further research, success-dependent payment and more difficult learning materials might overcome those motivational issues. Furthermore, it might be fruitful to support the retrieval by prompts that engage the students to retrieve information of the first text during reading of the second text.

Another account for the lack of effects of distributed reading might be the relatively short retention intervals. Even with an extended retention interval of two weeks in Experiment 2, this retention interval was still relatively short. For example, Putnam and colleagues (2017) recommend a retention interval of two months for a lag of one week. It seems promising to investigate distributed reading with even longer retention intervals as two months or even longer, to determine if the detrimental effects of distributed reading might turn into benefits in the long run.

Additionally, potential beneficial effects of distributed reading could also have been undermined by repeated testing. In our experiments, we chose a within-subject manipulation of the retention interval. This design was chosen to reach a more economic procedure while maintaining high statistical power, but it also brings along some issues. Repeated testing is well known as an effective learning strategy (Rowland, 2014), but influenced by retrieval success. More successful retrieval leads to better learning outcomes than less successful retrieval. As distributed reading was detrimental in immediate tests, participants in the distributed condition might have been at a disadvantage at the later retention intervals. Thus, in further research, a between-subject realization of the retention interval would be recommended.

Finally, there might be a fundamental feature of distributed learning that might ward against beneficial effects. In contrast to distributed practice, distributed learning is characterized by the fact that the learned information is not presented again and refreshed at the second learning occasion. Therefore, if readers have forgotten the information until the second learning session, it is very likely that this information will not be restored during the second learning occasion. Thus, in distributed learning, forgetting influences the performance even at the short retention interval, whereas the performance of the massed condition at the short retention interval is influenced by text comprehension alone. This leads to a systematic underestimation of performance in the distributed condition. However, as this confound is a design feature of distributed reading, it is questionable if the confound could be resolved by a different operationalization. One possibility was explored in Experiment 2 of C. E. Greving and Richter (2021), in which the interstudy interval was held very short (15 min). However, such short interstudy intervals might also underestimate the benefits of distributed learning, as retrieval gets more difficult with more time between the texts. Therefore, with long and short interstudy intervals, distributed learning has to be very effective to overcome its drawback.

In real-life learning settings, it could be assumed that most learning occasions are not an exact repetition of a previous learning occasion, but build upon previous learning occasions, containing a mixture of repetition and new, often related, contents. However, to date, only beneficial effects for distribution of learning in repetitive learning are supported by the state of research.

Conclusion

In our experiments, we found no beneficial effects of distributed reading. Instead, even detrimental effects were found immediately after reading and for students with low prior knowledge. Despite these inauspicious results, our research again demonstrated that delayed measurement is necessary to evaluate the effectiveness of didactical strategies. Furthermore, the results indicate that learner characteristics as domain-specific prior knowledge might moderate the effectiveness of distribution. As distributed learning seems to be a common learning strategy, further research should investigate whether the recommendation to “Stop cramming!” also holds for all learners and non-repeated learning.

Chapter IV

Experiment 5

Distributed Practice in the Classroom: Effects of Rereading Schedules Depend on Time of Test

A version of this chapter is published as:

Greving, C. E., & Richter, T. (2019). Distributed learning in the classroom: Effects of rereading schedules depend on time of test. *Frontiers in Psychology,* 9, 2517. https://doi.org/10.3389/fpsyg.2018.02517

Distributed Practice in the Classroom: Effects of Rereading Schedules Depend on Time of Test

Carla E. Greving & Tobias Richter

Abstract. Research with adults in laboratory settings has shown that distributed rereading is a beneficial learning strategy but its effects depend on time of test. When learning outcomes are measured immediately after rereading, distributed rereading yields no benefits or even detrimental effects on learning, but the beneficial effects emerge two days later. In a preregistered experiment, the effects of distributed rereading were investigated in a classroom setting with school students. Seventh-graders ($N = 191$) reread a text either immediately or after one week. Learning outcomes were measured after 4 minutes or 1 week. Participants in the distributed rereading condition reread the text more slowly, predicted their learning success to be lower, and reported a lower on-task focus. At the shorter retention interval, massed rereading outperformed distributed rereading in terms of learning outcomes. Contrary to students in the massed condition, students in the distributed condition showed no forgetting from the short to the long retention interval. As a result, they performed equally well as the students in the massed condition at the longer retention interval. Our results indicate that distributed rereading makes learning more demanding and difficult and leads to higher effort during rereading. Its effects on learning depend on time of test, but no beneficial effects were found, not even at the delayed test.

Introduction

Learning from text is essential for learning in school and academic settings. But how should we read to foster long-term learning? Distributing learning episodes of study material over a longer time instead of cramming in one session has shown to be a beneficial learning strategy, especially for longer retention intervals (spacing effect; Cepeda et al., 2006). Given that distributed practice is usually perceived as more difficult by the learners than massed learning, distributed practice is categorized as a desirable learning difficulty (E. L. Bjork and Bjork, 2011). The assumption that distributed practice benefits long-term learning seems to hold also for learning with texts. Research with adults in laboratory settings has repeatedly shown that distributed rereading of a text is more effective for long-term retention than massed rereading (Glover and Corkill, 1987; Krug et al., 1990; Rawson, 2012; Rawson and Kintsch, 2005; Verkoeijen et al., 2008). However, the effect of distributed vs. massed rereading has not yet been investigated with younger learners in real-world educational settings. In a preregistered experiment (C. E. Greving & Richter, 2017), we investigated distributed rereading in a school environment with seventh-graders. In this article, we first discuss desirable difficulties and distributed practice in general. We also provide an overview of empirical findings on the effects of distributed rereading and then introduce the current experiment.

Distributed Practice and Desirable Difficulties

Distributed practice is one of several learning strategies labelled as desirable difficulties (R. A. Bjork, 1994; E. L. Bjork & Bjork, 2011; Lipowsky et al., 2015). These learning strategies share two key features. They seem to make learning more difficult during learning, but they enhance learning outcomes in the long term. One factor assumed to make learning more difficult but foster long-term retention is the time between repetitions of learning material.

Distributed practice refers to learning schedules in which repetitions of the information to be learned (e.g., a new word in a foreign language) is distributed over several (at least two) learning sessions instead of learning in only one session. For example, when using flashcards to learn vocabulary of a new language, the inter-study interval should be increased between the repetitions of the same flashcard. The term distributed practice encompasses the spacing and the lag effect. The spacing effect refers to the finding that any inter-study interval leads to better learning than massed learning (i.e., learning with an inter-study interval of zero). However, in studies investigating the lag effect, learning outcomes are compared between learning schedules with different inter-study intervals.

The spacing effect, as defined by Cepeda and colleagues (2006), is a robust effect that is not moderated by the retention interval. That is, distributed practice is usually better than massed learning. In contrast, the lag effect designates a non-monotonic effect of the inter-study interval. Learning performance increases with longer inter-study intervals until the effect reaches a peak, after which the performance decreases with even longer inter-study intervals. Moreover, the lag effect depends on the retention interval. Learning over longer retention intervals seems to benefit from longer inter-study intervals (Cepeda et al., 2006, 2008; Glenberg, 1976). Different processes might account for the spacing and the lag effect (Cepeda et al., 2006; Küpper-Tetzel & Erdfelder, 2012). Retrieval processes during retention tests have been discussed as explanations for the spacing effect, whereas the lag effect may be explained by different encoding strategies (e.g., retrieval of the first encoding of an item) during learning or maintenance after learning. One key mechanism might be the retrieval of stored information from the first learning occasion during the second learning occasion. Study-phase retrieval theories suggest that successful retrieval of the first learning occasion is needed to strengthen the memory trace and thus prevent forgetting (Cepeda et al., 2008; Delaney et al., 2010; Thios & D’Agostino, 1976).

Distributed Rereading as Desirable Difficulty in Learning

The long-term benefits of distributed practice have been shown for a wide range of materials, from simple motoric tasks (e.g., Baddeley & Longman, 1978) and simple materials such as vocabulary (e.g., Kornell, 2009) to complex learning materials such as texts (Rawson and Kintsch, 2005). Rereading texts clearly seems to be a common learning strategy widely used by students (e.g., Karpicke et al., 2009, Gagnon & Cormier, 2018). Contrary to common sense, rereading a text immediately after the first reading often provides at best marginal gains in the learning outcome compared to reading the text only once (Callender & McDaniel, 2009). However, rereading the text in a distributed fashion might be a better strategy (Glover and Corkill, 1987; Krug et al., 1990; Verkoeijen et al., 2008), but its effectiveness seems to depend on the retention interval (Gordon, 1925; Rawson, 2012; Rawson and Kintsch, 2005).

Rawson and Kintsch (2005, Experiment 1) investigated the rereading and retention interval effects by comparing recall and text comprehension performance of undergraduates who read an expository text about carbon sequestration (1730 words) once or twice either immediately after the first reading or one week later. Recall and text comprehension performance were measured either immediately after reading or after a delay of two days. When learning outcomes were assessed immediately after reading, students who had read the text twice in the massed condition outperformed students in the single reading condition in recall and text comprehension performance, whereas no differences were found between the distributed reading and single reading conditions. Thus, at the short retention interval, no benefit of distributed rereading was found. In the recall performance, students in the massed condition even outperformed students in the distributed condition. But when learning outcomes were measured two days later, a different pattern emerged. Students in the distributed condition outperformed those in the massed and single reading condition in recall and comprehension performance. Thus, the benefits of distributed rereading depended on time of test.

Rawson (2012; also see Rawson and Kintsch, 2005, Experiment 2) replicated the interaction between the rereading and retention intervals in three experiments with undergraduates and a text about the portrayal of historical events in Hollywood films (1541 words in length). In all experiments, they found no difference between the rereading conditions at the short retention interval, whereas students in the distributed condition outperformed students in the condition with immediate rereading at the long retention interval. In addition, Rawson and Kintsch (2005) as well as Rawson (2012) measured the reading times and found a decrease in reading time between the first reading and the rereading. The decrease was greater for the group with immediate rereading. Thus, participants in the distributed condition spent more time reading the second text than participants in the massed condition.

In sum, the interaction between rereading schedules and retention intervals on testing performance seems to be robust in college students in laboratory settings. The differences in reading times suggest that readers spend greater cognitive effort in distributed vs. massed rereading.

Meta-Cognitive Judgements of the Learning Process and Distributed Practice

Although distributed practice is an effective learning strategy, students seem to underrate the effectiveness of distributing their learning time in their metacognitive judgments of the learning process (for a review see Son and Simon, 2012). One core type of meta-cognitive judgements of the learning process, which is often assessed immediately after learning, is the estimated proportion of correctly recalled items. These judgements are influenced by many cues, as for example the perceived difficulty of a to-be-learned item (Koriat, 1997; see also Vössling et al., 2017, for the influence of difficulty on the accuracy of those judgements). For example, Kornell (2009) investigated distributed vs. massed learning of vocabulary with flashcards. Despite the objective advantage of a distributed practice strategy, participants estimated a higher percentage of correct recalled items of the massed learned items than of the distributed learned items. A possible explanation for this pattern is the lower experienced fluency during distributed practice (Alter and Oppenheimer, 2009; R. A. Bjork et al., 2013). As distributed practice should induce a (desirable) difficulty, learners might also perceive learning as more difficult when the materials are presented in a distributed instead of a massed fashion. Thus, distributed rereading might not only affect the learning outcome and the reading time, it might also alter the meta-cognitive judgements of the learning process. However, to our knowledge, the effects of distributed rereading on the meta-cognitive judgements of the learning process have not yet been investigated. As texts are more complex learning materials than single words, the question arises whether distributed rereading also induces a perceivable difficulty and if so, whether meta-cognitive judgements of learning are affected by the difficulty induced by distributed rereading.

Distributed Practice in Real-World Educational Settings

The effects of distributed practice are well investigated in laboratory settings but only few studies have been conducted that examine distributed practice in real-world educational settings (Küpper-Tetzel, 2014). However, to give recommendations to teachers to apply distributed practice, studies are needed to investigate whether this teaching strategy is indeed beneficial in real-world educational settings. Such settings differ in a number of respects from laboratory settings. For example, distributed practice occurs embedded in other instructional activities, learning usually is usually more self-regulated and is based on more complex materials.

Furthermore, the studies introduced above have been conducted with adult learners, especially with undergraduates. Experimental settings in school and with younger learners might confront researchers with more heterogeneous samples. Whereas undergraduate university students often represent a highly selected group of learners on a relatively high level of ability, in a secondary school setting, high-capacity students often visit the same class as low-capacity learners. Interestingly, advantages of distributed practice were shown for vocabulary learning with school students in classroom settings (Bloom & Shuell, 1981; Küpper-Tetzel et al., 2014; Sobel et al., 2011), and distributed practice of scientific concepts and laws seems to foster long-term learning (Gluckman et al., 2014; Grote, 1995; Kapler et al., 2015; Vlach, 2014; Vlach & Sandhofer, 2012). However, in an experiment conducted by Goossens et al. (2016), a longer lag failed to facilitate primary school vocabulary learning in a classroom learning scenario compared to a shorter lag condition.

Additionally, learning abilities (for example skill learning, Schiff and Vakil, 2015), general cognitive prerequisites for learning such as working memory capacity (Gathercole et al., 2004) and reading comprehension skills (Perfetti, Landi, & Oakhill, 2005) that are especially important for learning from text underlie huge developmental changes. Thus, as matter of principle, a learning method that has been shown to be beneficial for adult learners is not guaranteed to work for younger learners. However, some studies suggest that distributed practice seems to be as beneficial for young children as for young adults (Seabrook et al., 2005; Toppino et al., 1991).

To summarize, despite the contrary findings regarding the lag effect of Goossens et al. (2016), distributed practice promises to be a beneficial learning strategy even for school-aged learners and in real-world educational settings. However, distributed rereading of expository texts has not yet been investigated with younger learners and it is unclear whether the findings for adult learners generalize to this population.

The Role of Prior Knowledge in Distributed Rereading

Prior knowledge is arguably the most important learner characteristic for learning from text (e.g., Kintsch, 1998), even more important than verbal abilities (Schneider et al., 1989). Moreover, prior knowledge has been shown to moderate the effects of text difficulty on learning from texts. McNamara and Kintsch (1996) demonstrated that the comprehension of junior high school students with low prior knowledge benefited from more coherent and thus easier texts, whereas the comprehension of students with higher prior knowledge benefit from less coherent and thus more difficult texts. As distributed rereading should also lead to higher difficulty in rereading, the question arises whether distributed rereading is also only beneficial for students with high(er) prior knowledge. In their experiment with university students, Rawson and Kintsch (2005) measured prior knowledge but did not find an interaction with the rereading schedule. Still, prior knowledge might play a role for distributed rereading in a school context, where the distribution of prior knowledge is likely to differ from the distribution typically found at universities.

The Current Experiment

In this preregistered experiment (Greving and Richter, 2017), we investigated the effects of massed and distributed rereading on short- and long-term retention with seventh-graders in the classroom. In addition to reading times, metacognitive judgements were obtained to gain insights into the learning process.

Based on the experimental design of Rawson and Kintsch (2005), participants twice read curriculum-orientated texts about the bacterial cell. The rereading occurred either immediately after the first reading or one week later. Recall and text comprehension performance were measured five minutes after rereading (short retention interval) or one week later (long retention interval). Thus, the present experiment is the first to investigate the effects of distributed rereading on the learning outcomes of school students but to also concurrently expand the research on the effects on metacognitive processes.

Following the findings of Rawson and Kintsch (2005), we expected that distributed rereading would have beneficial effects on learning in recall and text comprehension performance. However, the expected beneficial effect of distributed rereading was expected to depend on time of test. No differences were expected at the short retention interval, whereas the benefits of distributed rereading was expected to be significant at the longer retention interval (Hypothesis 1). In addition, we expected that because of forgetting, the learning outcome should decrease between the retention intervals (Hypothesis 2). Considering the significant influence of prior knowledge on learning with texts, we also estimated the effects of domain-specific prior knowledge. We first assumed that students would learn more from the texts the higher their prior knowledge (Hypothesis 3). This hypothesis is backed up by a large body of research demonstrating the importance of prior knowledge in learning from text (e.g., Schneider et al., 1989). Although the hypothesis is not novel, testing it in the present experiment is important to ensure that students indeed used their prior knowledge to understand and learn from the text. Furthermore, we addressed the exploratory research question whether the effects of distributed rereading would depend on prior knowledge (similar to other measures that make text comprehension more difficult, such as low-coherence texts, McNamara et al., 1996)

We also hypothesized that distributed rereading would lead to greater cognitive effort and hence longer reading times in the second text presentation (Hypothesis 4). Regarding metacognitive judgements of learning, we expected that students would perceive distributed rereading as more difficult (Hypothesis 5) and rate the learning process as less successful (Hypothesis 6). Despite the perceived disadvantage, we expected that students would be more focused on the task (Hypothesis 7) during distributed rereading.

Method

Participants

The sample included 191 (53 % female) seventh-grade students from eight classes and three different schools (German Gymnasium and comprehensive schools). The average age of participating students was 12.94 years (*SD* = 0.39). Students participated only if their parents had given their permission (97% permission; students without permission took quizzes during sessions). Students were randomly assigned to the four experimental learning conditions: massed learning condition with delayed measurement (*n* = 49), massed learning with immediate measurement (*n* = 47), distributed learning with delayed measurement (*n* = 48), and distributed learning with immediate measurement (*n* = 47). As a reward, the students received sweets after each session and a magic cube puzzle after the last session.

Twenty students missed at least one of the learning sessions, thus did not read the texts twice. Therefore, their data were excluded from all analyses. This participant loss resulted in the following group sizes: massed/delayed (*n* = 47), massed/immediate (*n* = 47), distributed/delayed (*n* = 37), and distributed/immediate (*n* = 40). Additionally, 26 students missed the test or the assessment of prior knowledge, resulting in the following group sizes in the analysis of free recall and text comprehension performance: massed/delayed (n = 36), massed/immediate (n = 45), distributed/delayed (n = 26), and distributed/immediate (n = 38).

Text Materials

The experimental text was an expository text about the bacterial cell (length 74 sentences, 977 words). The bacterial cell structure is part of the extended curriculum of biology science classes in the State of Hessen (Germany) where the study was conducted. However, the bacterial cell structure is usually not covered in class because it is too small to be microscopable in school contexts. Thus, it was unlikely that the students had prior knowledge about the bacterial cell itself, but they might have had prior knowledge about cells in general. A complementary image illustrating the structure of the cell that was also explained in the text was presented adjacent to the text. The image was presented stable, thus the reader could always integrate text and image. This is comparable to the typical layout of text books of biology, in which the information about the respective cell is mostly accompanied by illustrations of its structures. The text had a Flesh reading ease score of 54 (German formula, Amstad, 1978;

Assessment of Learner Characteristics

Participants' first language and diagnosed reading and writing disability were reported by their teachers in a teacher questionnaire. Moreover, further learner characteristics were assessed via standardized tests. Besides the domain-specific prior knowledge, we assessed reading ability, working memory capacity, and knowledge about reading strategies as further abilities which are associated with reading and learning skills and thus can be seen as prerequisites for learning (see section 1.4). A randomized block design was used to ensure that the experimental groups are matched with respect to these abilities.

Domain-Specific Prior Knowledge. Participants were asked to answer five open-ended questions and to label the components of the plant cell and the bacterial cell in a schematic image. The questions covered knowledge related to the bacterial cell (e.g. function of cells, knowledge about genetic information), but were asked in a way to promote the students to write up any prior knowledge. For example, one question was "What is a plant or animal cell? Please write down everything you know about those cells." The questions have been used in two other studies as well (two experimental and one pilot study). In these studies, the scores were highly correlated with the recall performance after reading a preliminary version of the text used in this experiment (pilot study: r= .72, 95% CI [.51, .84).The questions were presented in randomized order. Additional to the knowledge questions, we also asked the participants to indicate whether they had encountered the topic before in class or at home. The protocols were scored by two independent raters following a coding scheme. Any answer which was correct even at low level, as for example "something inside an animal", was given a point, with more points given for more elaborated answers as "An animal or plant cell is a tiny unit of a plant or an animal.", ICC (2,1) = .93, 95% CI [.923, .932] (Shrout and Fleiss, 1979).

Knowledge About Reading Strategies. Participants completed the Würzburger Lesestrategie-Wissenstest für die Klassen 7-12 (WLST 7-12; Würzburg Reading Strategy Knowledge Test, Schlagmüller & Schneider, 2007; split-half reliability, r = .90, estimated in a sample of 4490 students in Grades 7-11). The WLST includes six items that require participants to grade the utility of different reading strategies in a given learning situation (on a scale from 1 to 6, corresponding to the German grading system, where 1 is the highest achievement and 6 the lowest).

Reading Ability. Participants completed the subtest sentence verification of ELVES, a German-speaking test that assesses the efficiency of basic reading processes at the word and sentence level (Richter and van Holt, 2005). In this task, 16 statements are judged as true or false (verification task). The test score combines reading speed and verification accuracy into an integrated score (Cronbach's α = .58, estimated in the current sample). The reliability of this measure was lower than in previous studies (e.g., Richter and van Holt, 2005, report a Cronbach's α of .87), indicating a relatively high amount of measurement error. However, given that the purpose of the reading ability measure was to match the experimental groups according to this criterion, the reliability of the measure may still be sufficient.

Working Memory Capacity. Working memory capacity for text was assessed with a computerized version of the Reading Span Task (RSPAN; Oberauer et al., 2000). The task involves verification judgements for sequentially presented sentences that increase in number throughout the test and the memorization of the final word of each sentence. The test score is the average proportion of correctly recalled words (Cronbach's $\alpha = .89$, estimated in the current sample).

Assessment of Learning Outcomes

Recall Performance. Recall performance was assessed with a free recall task. Participants were asked to write down as much information that could be recalled from the first part of the text. The participants were given a time limit of two minutes. The free recall protocols were scored by two independent raters, ICC (2,1) = .92, 95% CI [.866, .948] (Shrout and Fleiss, 1979).

Text Comprehension Performance. Text comprehension performance was assessed with eight short-answer and six single-choice questions (one correct response option and three distractors). For example, one short-answer question was, "A bacteria cell does not have a cell nucleus. But where can you find the genome of the bacteria cell?", and one single-choice questions was, "To which kind does the bacteria cell belong?", with the response options (a) Prokaryots, (b) Eukaryots, (c) Plasmid, and (d) Organelle. The additional single-choice questions (compared to Rawson and Kintsch, 2005) were chosen because younger learners in previous (yet unpublished) experiments tended to forego answering the open-ended questions. All questions were literal questions asking for information explicitly stated in the text. The questions had originally been developed for these previous experiments and were optimized for the present experiments regarding item difficulties. The item difficulty (calculated averaged about all learning and retrieval conditions) ranged between .01 and .72, with a mean difficulty of .30 (SD = .19) in the short-answer questions as well in the single-choice questions (SD = .10) (corrected for chance success). Answers to the short-answer questions were scored as either incorrect (0) or correct (1) by two independent raters who were blind to the experimental conditions (Cohen's $\kappa = .87$).

Assessment of Learning Processes

Reading Time. The students read the text in a self-paced fashion with the moving-window method. The text was presented on screen with all sentences blurred except the one the student was currently reading. The students could return to previously read sentences to reread them. Reading times per sentence were assessed and divided by the number of letters in the sentence to account for different sentence lengths.

Metacognitive Judgements of the Learning Process. After reading the text for the second time, participants judged the following aspects of the learning/reading process on 5-point Likert scales. They predicted their learning success and rated the perceived reading difficulty. In addition, the perceived on-task focus (three items, one reversed, Cronbach's $\alpha = .64$, estimated in the current sample) was assessed. Furthermore, they rated the perceived similarity of the two (identical) texts for exploratory purposes. The results for this measure are not reported as they do not contribute to answering the research questions.

Procedure

All materials were presented on notebook computers with 15.6" screens. The experiment was created and presented with the software Inquisit 3 (Version 3.0.6.0, 2011).

The experiment consisted of four sessions (Figure 4.1). The pretest took place at the first session, in which the experimental parts were administered collectively in the classroom, supported by instructions on screen. The students completed the prior knowledge test, the WLST, the ELVES, and the RSPAN tests, in this order.

In the further sessions, instructions were given on screen after a short instruction delivered by the experimenter to the whole group.

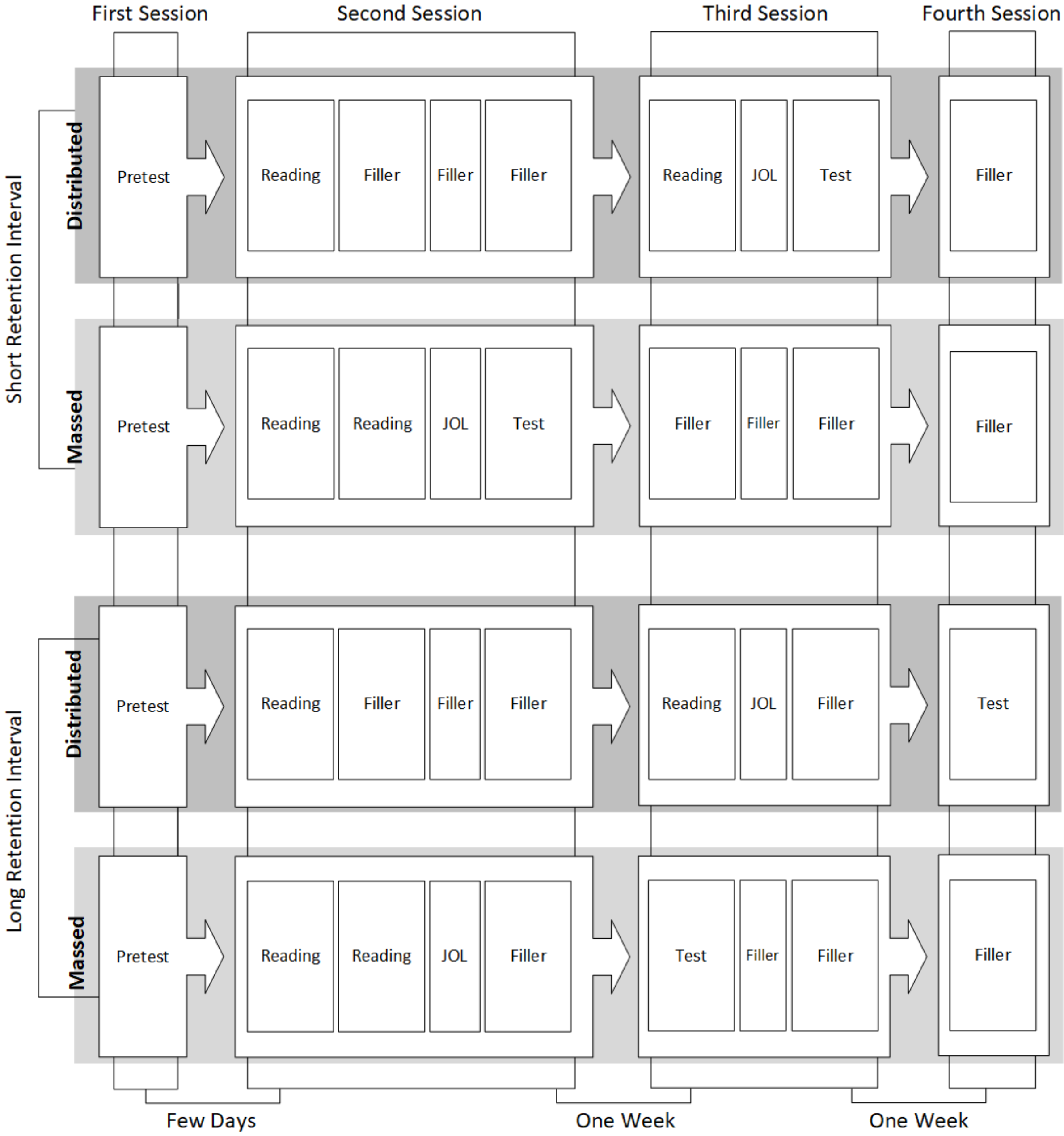
In the second session, the students either read the experimental text once (distributed) or twice (massed). In the distributed condition, the students received filler tasks after reading. All filler tasks consisted of questions about social media usage and were not analyzed. In the massed condition, the students completed the metacognitive judgements of the learning process. Afterwards, they either were tested (short retention interval) or received a filler task (long retention interval).

In the third session, students read the second text (distributed condition) or received a filler task (massed condition, short retention interval), or the recall test (massed condition, long retention interval). Afterwards, students in the massed condition received a filler task. In the distributed condition, the students completed the metacognitive judgements of the learning process and were then either tested (short retention interval) or received a filler task (long retention interval).

In the fourth session, students were tested (distributed condition, long retention interval) or received a filler task (all other experimental groups).

Figure 4.1

Overview of the Experiment Procedure



Note. Reading, reading the test; Filler, filler task; JOL, metacognitive judgments of the learning process; Test, assessment of recall performance and text comprehension performance

Design

We employed a 2 x 2 between-subjects design with matched (parallel) groups and the independent variables learning condition (massed vs. distributed by one week) and retention interval (immediate vs. one week delayed). To ensure similar capabilities in all learning conditions, we first formed homogeneous blocks of students matched according to first language, reading and writing disabilities, prior knowledge and reading ability. The students from these groups were then randomly assigned to the experimental conditions. No differences were found between the two learning conditions in working memory capacity, $F(1,155) = 0.26$, $p = .611$, and reading ability, $F(1,155) = 0.41$, $p = .521$, and between the two groups tested at different retention intervals in working memory, $F(1,155) = 1.42$, $p = .236$, and reading ability, $F(1,155) = 0.081$, $p = .777$. Likewise, the interaction of the two independent variables was not significant for working memory, $F(1,155) = 0.01$, $p = .940$, or reading ability, $F(1,155) = 0.36$, $p = .548$. Finally, we found no differences between learning conditions in prior knowledge, $F(1,155) = 1.69$, $p = .195$, and reading strategy knowledge, $F(1,155) = 0.33$, $p =.565$, or between the two groups tested at different retention intervals in prior knowledge, $F(1,155) = 0.01$, $p = .942$, and reading strategy knowledge, $F(1,155) = 0.01$, $p =.930$. The interaction was also not significant for prior knowledge, $F(1,155) = 0.54$, $p = .464$, or reading strategy knowledge, $F(1,155) = 2.28$, $p = .133$.

Results

We used linear models (recall performance and judgements of learning), linear mixed-effect models (LMM, reading time) and generalized linear mixed-effect models (GLMM, text comprehension performance) with the R packages lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017) and lsmeans (Lenth, 2016) in the R environment in version 3.4.4 (R Developmental Core Team, 2018). Mixed effect models are the method of choice for analyzing data in educational contexts, which are often characterized by a hierarchical multilevel structure (students nested in classes nested in schools). Moreover, these models are advantageous in experimental contexts when participants and experimental items form a crossed (imperfect) hierarchy (Baayen et al., 2008). We included school, class, student, or item as random effect (random intercept) if the intra-class correlation of the dependent variable exceeded .05. Unstandardized regression weights are reported. For interpreting the GLMM results, predicted probabilities (back-transformed from the log odds) for experimental conditions are reported. For all models, the distribution of residuals was inspected visually for normality. All available data points were analyzed; no outliers were excluded. Type 1 error probability was set at .05. Directed hypotheses were tested with one-tailed tests.

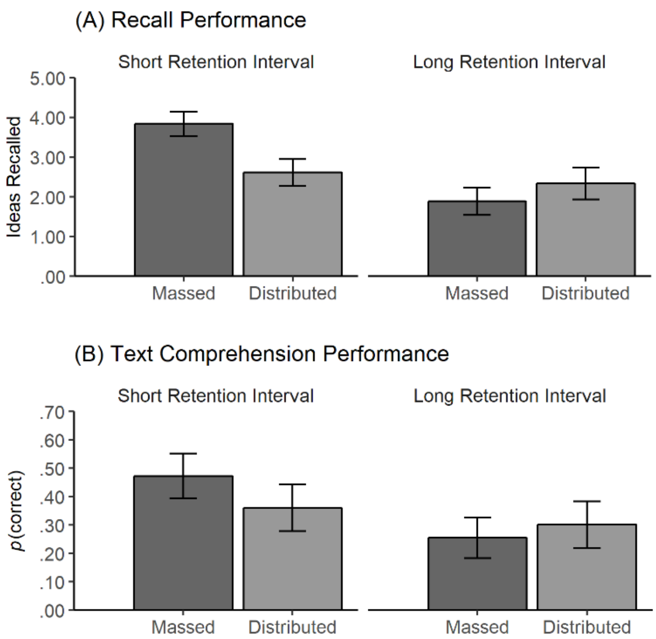
Recall Performance

We estimated a linear model with learning condition (contrast coded: massed = -1, distributed = 1), retention interval (contrast coded: short = -1, long = 1), prior knowledge (*z*-standardized), and the two- and three-way interactions of these variables as predictors and recall performance as dependent variable. A main effect of retention interval emerged, $\beta = 0.57$, $SE = 0.17$, $t(137) = 3.34$, $p < .001$, one-tailed, $\Delta R^2 = .08$. As expected, recall performance was better at the short interval ($M = 3.21$, $SE = 0.22$) than at the long retention interval ($M = 2.06$, $SE = .26$). Additionally, students' recall performance was positively related to their prior knowledge, $\beta = 0.55$, $SE = 0.19$, $t(137) = 2.86$, $p = .002$, one-tailed, $\Delta R^2 = .05$. Thus, a difference of one standard deviation in prior knowledge corresponded to a 0.55 difference in the free recall task. No main effect of learning condition was found on recall performance, $\beta = -0.19$, $SE = 0.17$, $t(137) = -1.09$, $p = .140$, one-tailed, $\Delta R^2 = 0.03$, performance in the massed condition ($M = 2.82$, $SE = .22$) did not differ from that in the distributed condition ($M = 2.45$, $SE = .26$). However, a significant interaction between learning condition and retention interval emerged, $\beta = -0.46$, $SE = 0.17$, $t(137) = -2.687$, $p = .004$, one-tailed, $\Delta R^2 = .04$ (Figure 4.2a). In the massed condition, recall performance decreased from the short ($M = 3.86$, $SE = 0.30$) to the long retention interval ($M = 1.79$, $SE = 0.34$), $t(137) = -4.60$, $p < .001$. In contrast, no difference was found between the short interval ($M = 2.56$, $SE = 0.33$) and long retention interval ($M = 2.34$, $SE = 0.40$) in the distributed condition, $t(137) = -0.43$, $p = .667$. Conversely, at the shorter retention interval, students in the massed condition outperformed students in the distributed condition, $t(137) = -2.93$, $p = .004$. At the longer retention interval, a slight and non-significant difference was found in the opposite direction between the massed (*M* = 1.79, *SE* = 0.34) and distributed condition (*M* = 2.34, SE = 0.40), *t*(137) = 1.05, *p* = .297.

These results showed that the predicted differential effects of massed vs. distributed learning at the short and long retention intervals were only partially supported. When students reread the text in a distributed fashion, no decrease in recall performance occurred from the short to the long retention interval. Nevertheless, the benefit of distributed rereading at the longer retention interval predicted in Hypothesis 1 did not occur.

Figure 4.2

Interaction between Learning Condition and Retention Interval in Recall Performance and Text Comprehension Performance



Note. Estimated recall and text comprehension performance in the two learning conditions (massed vs. distributed) at the short and the long retention interval; (A) mean number of recalled ideas in the free recall task, (B) back-transformed probability of a correct answer in the text comprehension test. Error bars represent standard errors.

Comprehension Performance

We estimated a generalized mixed model with students and items as random effects (random intercepts) and learning condition (contrast coded: massed = -1, distributed = 1), retention interval (contrast coded: short = -1, long = 1), prior knowledge (*z*-standardized), and item type (contrast coded: CR = 1, MC = -1) and their interactions as predictors with fixed effects and comprehension performance as dependent variable (Table 4.1).

Similar to the model for recall performance, retention interval ($\beta = 0.30$, $SE = 0.10$, $z = 3.10$, $p < .001$, one-tailed,) and prior knowledge ($\beta = 0.48$, $SE = 0.11$, $z = 4.31$, $p < .001$) exerted main effects on comprehension performance. Participants performed better at the short retention interval (probability = .41, $SE = .07$) than at the long retention interval (probability = .28, $SE = .07$). A difference of one standard deviation in prior knowledge corresponded to an 11% difference in the probability to provide a correct response. The main effect of learning condition was not significant, $\beta = -0.08$, $SE = 0.10$, $z = -0.84$, $p = .201$, one-tailed. Performance in the massed condition (probability = .37, $SE = .07$) did not differ from perfomanc0 in the distributed condition (probability = .33, $SE = .08$). However, the model revealed a significant interaction between learning condition and retention interval, $\beta = -0.19$, $SE = 0.10$, $z = -1.95$, $p = .026$, one-tailed (Figure 4.2b). Consistent with the findings from the recall performance analysis, students in the massed condition showed a decrease in the text comprehension performance from the short (probability = .47, $SE = .08$) to the long retention interval (probability = .26, $SE = .07$), $z = -3.97$, $p < .001$. In contrast, no significant decrease in text comprehension performance was found in the distributed condition from the short (probability = .35, $SE = .08$) to the long retention interval (probability = .30, $SE =.08$), $z = -0.81$, $p = .420$. At the short retention interval, the difference between massed and distributed condition was statistically significant, $z = -2.16$, $p = .031$, whereas the difference at the long retention interval was not significant, $z = 0.75$, $p = .455$.

Additionally, we found a significant three-way interaction between learning condition, prior knowledge, and item type, β = 0.15, *SE* = 0.06, *z* = 2.59, *p* = .010. The performance of students in the distributed condition was more strongly associated with prior knowledge than in the massed condition, but only with short-answer questions (Figure 4.3).

Table 4.1

Estimated Coefficients, Standard Errors and z-Values for the Generalized Linear Mixed Model with Text Comprehension as Dependent Variable.

	Est.	*SE*	*z*	
(Intercept)	-0.70	0.32	-2.20	*
Learning condition (LC)	-0.08	0.10	-0.84	
Retention interval (RI)	0.31	0.10	3.10	***
Item type	-0.54	0.30	-1.78	
Prior knowledge (PK)	0.48	0.11	4.31	***
Learning condition x retention interval	-0.19	0.10	-1.95	*
Learning condition x item type	-0.01	0.05	-0.12	
Learning condition x prior knowledge	0.09	0.11	0.79	
Retention interval x item type	-0.07	0.05	-1.38	
Retention interval x prior knowledge	0.01	0.11	0.10	
Prior knowledge x item type	0.17	0.06	2.95	**
Learning condition x retention interval x item type	0.03	0.05	0.56	

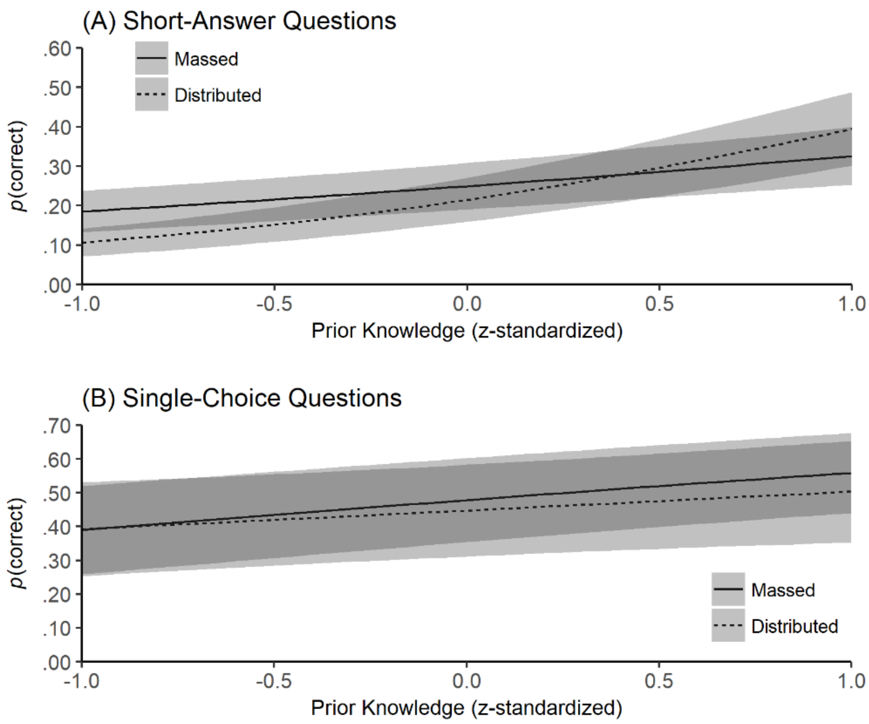
Learning condition x retention interval x prior knowledge	-0.09	0.11	-0.81	
Learning condition x prior knowledge x item type	0.15	0.06	2.59	**
Retention interval x prior knowledge x item type	-0.12	0.06	-1.99	*
Learning condition x retention interval x item type x prior knowledge	-0.01	0.06	-0.18	

Note. Learning condition (contrast coded: distributed = 1, massed=-1). Retention interval (contrast coded: immediate=-1, delayed=1). Item type (contrast coded: CR=1, SC=-1). Prior Knowledge was included z-standardized. * $p < .05$, ** $p < .01$, *** $p < .001$ (one-tailed for directional hypotheses).

To further interpret the interaction, we estimated and tested the effect of learning condition on the performance with short-answer questions for students with low prior knowledge (1 *SD* below the sample mean) and for students with high prior knowledge (1 SD above the sample mean; see Aiken and West, 1991, for a discussion on post-hoc probing of continuous moderators). The analyses revealed that students with low prior knowledge showed lower comprehension performance in the distributed condition (probability = .10, *SE* = .03) than in the massed condition (probability = .18, *SE* = .05), $z = -2.10$, $p = .036$, whereas for students with high prior knowledge, no such difference was found between massed (probability =.33, *SE* = .07) and distributed conditions (probability =.39, *SE* = .09), $z = 0.88$, $p = .380$. The pattern of results for this type of question suggests that only students with lower prior knowledge were impeded by distributed rereading.

Figure 4.3

Interaction between Learning Condition and Prior Knowledge in Short-Answer and Single-Choice Questions



Note. Back-transformed text comprehension performance (probability of correct answer) in (A) short-answer questions and (B) single-choice questions estimated as a function of prior knowledge and learning condition (massed vs. distributed). Shaded areas around each line represent standard errors.

Summarizing the results of recall and text comprehension performance, Hypothesis 1, which stated that distributed rereading would have beneficial effects on learning in long-term retention, was not supported. In both learning outcomes, we found the interaction between learning condition and retention interval predicted in Hypothesis 1, but contrary to our assumptions, we found no benefit of distributed rereading at the longer retention interval. We found the decrease in both learning outcomes predicted in Hypothesis 2 but only in the massed condition. As predicted in Hypothesis 3, participants with higher prior knowledge showed better recall and text comprehension performance. Finally, our exploratory findings showed that participants with low prior knowledge seemed to be impeded by distributed rereading, whereas participants with higher prior knowledge benefitted equally from both reading conditions.

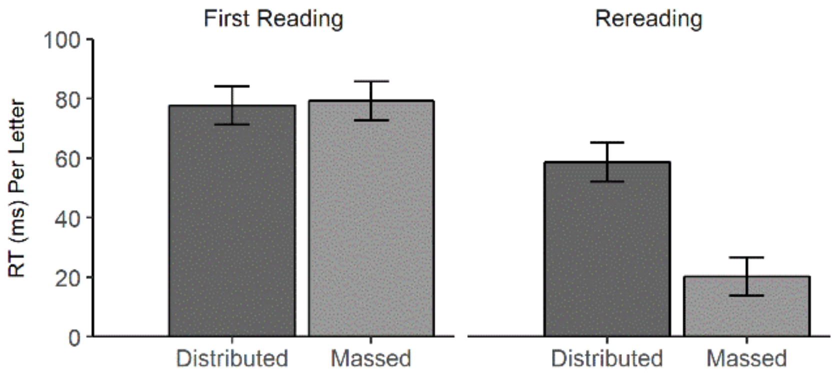
Reading Behavior

Reading times (first pass reading) were analyzed in a linear mixed model with sentences and students as random effects (random intercepts) and the fixed effects of learning condition (contrast coded: massed = -1, distributed = 1) and text presentation (contrast coded: first presentation = 1, second presentation = -1) and their interactions. It should be noted that the intraclass correlation for students missed the criterion value of 0.05, but it was included as random effect to achieve normal distribution of residuals. This model revealed a significant main effect of learning condition, $\beta = 9.72$, $SE = 1.53$, $t(169) = 6.34$, $p < .001$, indicating slower reading times in the distributed condition ($M = 70.07$, $SE = 6.78$) than in the massed condition ($M = 50.62$, $SE = 6.72$), and a main effect of text, $\beta = 20.36$, $SE = 0.65$, $t(25062.99) = 31.26$, $p < .001$. The second text presentation was read faster than the first, in the massed condition, $t(25062.99) = 33.63$, $p < .001$, and in the distributed condition, $t(25062.99) = 11.72$, $p < .001$. However, this difference was larger in the massed condition, as indicated by the significant interaction between learning condition and text presentation, $\beta = -9.04$, $SE = 0.65$, $t(25062.99) = -13.88$, $p < .001$ (Figure 4.4). Follow-up tests revealed that the reading times in the first presentation did not differ between the massed condition ($M = 80.02$, $SE = 6.77$) and the distributed condition ($M = 81.38$, $SE = 6.85$), $t(235.51) = 0.41$, $p = .682$. In contrast, in the second presentation, participants in the distributed condition ($M = 58.75$, $SE = 6.85$) read the text more slowly than participants in the massed condition ($M = 21.23$, $SE = 6.77$), $t(235.51) = 11.27$, $p < .001$.

In sum, the findings support Hypothesis 4 that distributed rereading would lead to longer reading times in the second text.

Figure 4.4

Reading Times in the Learning Conditions for First Reading and Rereading



Note. Estimated reading times per letter in the two learning conditions (massed vs. distributed) for the first reading and the rereading of the text. Error bars represent the standard error of the mean.

Judgements of the Learning Process

For the perceived reading difficulty, predicted learning success analyses and on-task focus, we estimated linear models with the respective item(s) as dependent variable and learning condition (contrast coded: massed = -1, distributed = 1) as predictor.

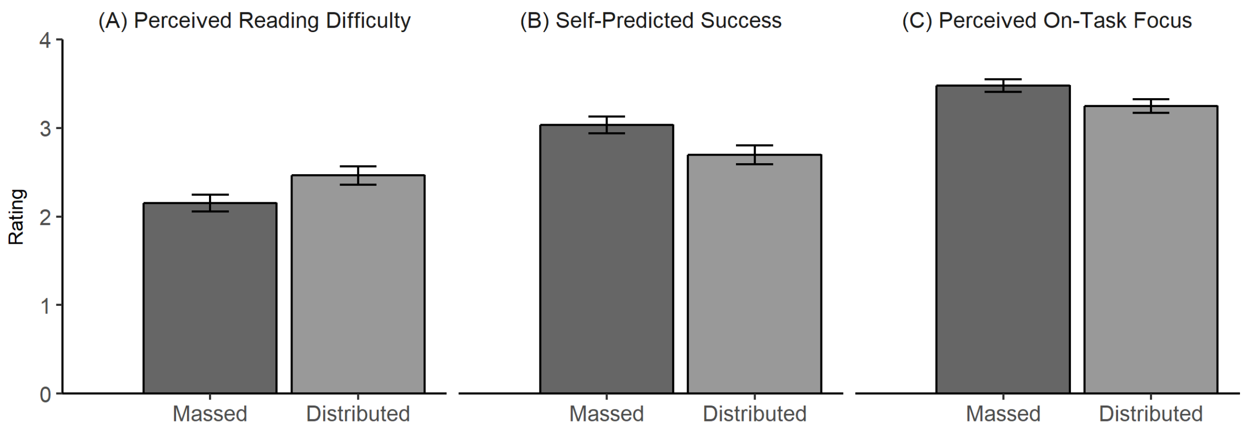
Perceived Reading Difficulty. The effect of learning condition on perceived reading difficulty was not significant (Figure 4.5a); it failed to reach significance by a narrow margin, $\beta = 0.11$, $SE = 0.07$, $t(169) = 1.65$, $p = .051$, one-tailed, $\Delta R^2 = .02$. Despite a descriptive difference between students in the distributed condition ($M = 2.38$, $SE = 0.10$) and students in the massed condition ($M = 2.15$, $SE = 0.09$) in the predicted direction, Hypothesis 5 was not supported.

Predicted Learning Success. Learning condition exerted an effect on predicted learning success, $\beta = -0.16$, $SE = 0.07$, $t(169) = -2.46$, $p = .008$, one-tailed, $\Delta R^2 = .03$. In line with Hypothesis 6, students in the distributed condition ($M = 2.71$, $SE = 0.10$) predicted lower learning success than students in the massed condition ($M = 3.04$, $SE = 0.09$) (Figure 4.5b).

Perceived On-Task Focus. Learning condition also had an effect on the perceived on-task focus during learning, β = -0.13, *SE* = 0.05, $t(511) = -2.70$, $p = .004$, one-tailed, $\Delta R^2 = .01$, but in the opposite direction than predicted in Hypothesis 7. Students reported higher on-task focus when rereading in a massed fashion ($M = 3.49$, $SE = 0.07$) compared to a distributed fashion ($M = 3.23$, $SE = 0.07$) (Figure 4.5c).

Figure 4.5

Judgements of the Learning Process in the Learning Conditions



Note. Estimated means in the judgements of the learning processes for the massed and distributed condition. Part (A) shows the perceived reading difficulty, (B) the resulting self-predicted success, and (C) the perceived on-task focus during reading.

Discussion

In this experiment, we investigated the effects of massed vs. distributed rereading on learning outcomes (recall and text comprehension performance) at two retention intervals, immediately after reading the text and one week later. We found a benefit for massed rereading at the short retention interval. At the longer retention interval, we found no difference between the learning conditions because of the lower forgetting rate in the distributed condition. In fact, the learning outcomes decreased between the retention intervals only in the massed condition, whereas students in the distributed condition showed no forgetting from the immediate to the delayed test of recall and comprehension performance. As a result, learning outcomes at the longer retention interval were on par for massed and distributed rereading but the distributed rereading condition did not show the expected advantage.

The main finding was that the effects of distributed rereading for secondary students depend on time of test, which parallels results found in earlier studies with college students. Distributed rereading seems to be detrimental when learning outcomes are assessed immediately, but it leads to a lower rate of forgetting that results in performance at least as good one week after learning. The difference in forgetting rates is in line with the previous studies by Rawson and Kintsch on distributed rereading (Rawson, 2012; Rawson and Kintsch, 2005). For example, Rawson (2012) found a decline of 49% for the massed (short-lag) condition, but only a decline of 3% for the distributed condition. By comparison, we found a decline of 50% in the massed condition and only 10% in the distributed condition. The difference in the decline of the distributed conditions might be explained by the length of the retention interval. Rawson’s (2012) delayed test was two days after learning, whereas the delayed test in the present study took place after one week.

The different patterns of learning outcomes at the two retention intervals raises the question of the underlying cognitive processes. Soderstrom and Bjork (2015) argue that short-term retention, assessed during learning or immediately after learning, rests on retrieval strength, i.e. on the currently accessible memory representations, whereas long-term retention relies on storage strength, which depends on the degree of interconnectedness of the learned information with other representations in long-term memory. For the latter, an unlimited capacity and no decrease over time is assumed (R.A. Bjork and Bjork, 1992). According to this approach, the goal of teaching and learning should be to increase storage strength and not retrieval strength. Importantly, a learning method which increases retrieval strength might even lead to lower increase in storage strength. To illustrate this assumption, Soderstrom and Bjork (2015) review several manipulations of learning situations which might have contrary effects on short- and long-term retention – and one of these might be distributed practice.

The pattern of effects for long- and short-term retention is also reminiscent of previous meta-analytic findings of the lag effect (Cepeda et al., 2006). As described above, the spacing effect does not depend on the retention interval, whereas the lag effect does. Moreover, Cepeda and colleagues (2006) reported from their meta-analysis spacing effects even for short retention intervals, and they found no evidence for the so-called Peterson paradox in which massed repetition is beneficial at short retention intervals. However, the distinction between spacing and lag effects depends on the definition of massed repetitions. For example, in Donovan and Radosevich's (1999) definition, a massed repetition may be interrupted by items or time when necessary for the experimental design, whereas Cepeda and colleagues (2006) specified that massed learning means that the learning should not be interrupted at all. This evokes the question whether massed rereading is a massed repetition of learning materials as defined by Cepeda and colleagues (2006).

Massed rereading means that a text is read (e.g., 977 words in the present experiment), and immediately following the last sentence, the reader starts again with the first sentence. Thus, the repetition of each sentence is distributed by several sentences before the reader encounters the same sentence in the second reading. Consequently, Rawson (2012) used the term short-lag rereading instead of massed rereading. Although we agree with Rawson (2012, p. 870) that the term "massed is somewhat of a misnomer" as it is applied to rereading, we are not certain whether the term should be changed. Naming this condition short-lag would imply that a shorter lag is possible, but it is not with text materials. Additionally, when learning from text, the comprehension of the coherent text is essential, which depends not only on the information given within one sentence but also on its relation to other sentences in the text. Thus, the text should be considered as the unit of learning, and rereading always includes the text as a whole. Hence, the difference between massed rereading and other massed repetitions (e.g., single words) is clearly due to the nature of the materials. Moreover, the massed conditions employed in numerous studies in educational contexts do not fit the definition according to Cepeda and colleagues (2006) (Bloom and Shuell, 1981; Fishman et al., 1968; Grote, 1995; Harzem et al., 1976; Kornell, 2009; Paik and Ritter, 2016). In real-world learning settings, some didactical strategies exclude pure massed repetitions, for example, when changing the repetition mode from reading to testing, as it was done in the study of Küpper-Tetzel and colleagues (2014). All of this implies that the pure spacing effect as defined by Cepeda and colleagues (2006) does not occur outside the laboratory. The research regarding distributed practice in real-world educational settings seems to investigate the lag effect rather than the spacing effect and thus might lead to a differential pattern regarding the learning outcomes at different times of test

Several theories (e.g., the one-shot account of spacing, Delaney et al., 2010), use retrieval processes to explain the effects of distributed learning. This mechanism might be especially important for the explanation of the lag effect in which forgetting between the repetitions of learning materials is essential. Because of the inter-study interval, the last presentation of an item must be retrieved from memory, which is more difficult when the inter-study interval is longer. Furthermore, the more difficult the retrieval, the stronger the memory trace (R. A. Bjork, 1975). Generalizing these ideas to rereading, the information acquired during the first reading of a text has to be retrieved from memory when rereading the text. In a massed presentation of the text, information acquired during the first reading is easily retrieved, whereas in distributed presentation, the retrieval is more difficult. This might result in a stronger memory trace, which is more resistant to forgetting compared to massed rereading. This interpretation is well in line with our finding that distributed rereading prevented forgetting.

Further research might additionally address the question whether the retrievability of information acquired during the first reading plays a crucial role in beneficial effects of distributed rereading and contrast its effect on short- and long-term retention.

Bearing the assumption in mind that distributed rereading is more related to the lag effect than to the spacing effect, the proportion of the inter-study and retention intervals might appear to not have been well chosen in the present study. According to Cepeda et al. (2008), the optimal inter-study interval for a one-week retention interval would have been one or two days (20-40% of the retention interval), or the optimal retention interval for a one-week inter-study interval would have been 18-35 days (note that these recommendations are based on experiments with simple verbal materials, not texts). In this experiment, we decided to use a retention interval which was as long as the inter-study interval. This was chosen for two different but related reasons. Most topics in school are taught on a weekly basis. Hence, testing the content of the previous lesson is often conducted one week later. For a more pragmatic reason, we also chose a schedule that fits well in the class learning schedule. Nevertheless, in further experiments, a longer retention interval and a better fit between retention interval and lag should be considered. Given the finding that distributed rereading changed forgetting from the short to the long retention interval, an effect of distributed rereading could emerge with a longer retention interval.

The findings from the prior knowledge analysis support the general assumption that students learn more from texts when their prior knowledge is already high (Kintsch, 1998; Schneider et al., 1989). For text comprehension performance assessed with short-answer questions, we also found a hint that the effects of distributed rereading depend on prior knowledge. Students with higher prior knowledge benefitted equally from distributed and massed rereading, whereas students with lower prior knowledge were hindered by the difficulties of distributed rereading. This finding is consistent with the idea that the learning difficulty introduced by distributed reading cannot be overcome by learners if the prior knowledge is too low. However, this relationship was not found in the free recall task and for single-choice questions.

We also found longer reading times in the second text presentation in the distributed condition compared to the massed condition. This pattern is comparable with the findings of Rawson (2012). In both experimental groups, the reading times during rereading were shorter than during the first reading. However, this decline was higher for the massed condition (74 %) than for the distributed condition (24 %) and both conditions declined to a greater extent compared to the rates reported by Rawson (2012), who found a decline of 14% for the distributed condition and 22% for the massed (short-lag) condition. School-aged students (at least in the age group that we looked at) might be even more vulnerable for rereading effects, especially when rereading takes place immediately. From this perspective, the extent that seventh-graders in the massed condition engaged in meaningful processing of the text during rereading is questionable. Apparently, though, at least some of the students engaged in meaningful processing at least to some extent. Otherwise, the superior performance of students in the massed condition at the short retention interval compared to students in the distributed condition would be difficult to explain. Nevertheless, given the results regarding the reading times, distributing the time of rereading might be even more essential in younger learners than in adults to prevent superficial processing of the text.

Students' meta-cognitive judgements of the learning process might indicate that distributed rereading is perceived as more difficult than massed rereading. Consistent with our assumptions, students in the distributed condition predicted lower learning success. However, the descriptive difference between the conditions regarding the perceived difficulty showed a trend in the predicted direction but missed statistical significance. Furthermore, contrary to our initial assumption, students in the massed condition perceived higher on-task focus during reading. This is especially surprising considering the shorter rereading times in the massed condition. Maybe the longer session in the massed condition was perceived as more demanding and difficult, but the students confused this feeling with being on-task. Thus, distributed rereading might be perceived as more difficult, but this was not fully reflected by differences in the judgement of reading difficulty. Nevertheless, in sum, the results regarding the judgements of learning fit well with the assumption that distributed rereading is qualified as desirable difficulty.

Its informative results notwithstanding, this study suffers from certain limitations. As discussed above, maybe the biggest limitation (that is shared with other experiments on distributed rereading) is that we compared only two retention intervals and two learning intervals. Such a design provides only a snapshot of learning and may generate results that are not easy to interpret. In future research, it would be desirable to contrast several learning and retention intervals to get an insight into lag-effects in distributed rereading. However, an experiment based on such a complex design would not be easy to implement in a school setting. Further limitations are associated with the greatest advantage of our study, its implementation in the classroom. Of course, the real-world educational setting may lead to compromises regarding the control of potential distractors and interruptions of the individual learning process, which might have added some noise to our results (although systematic confounds are unlikely given the rigorous experimental design). Last but not least, the participants in this experiment read just two texts. The text topic was chosen carefully to match typical contents of the school curriculum and the texts were carefully designed to match typical expository texts for secondary school students. Nevertheless, the generalizability of results to other topics and texts is not entirely clear.

To conclude, this experiment was the first to replicate a central finding of distributed rereading with school-aged learners in a real-world learning setting: The effects of distributed rereading depend on the time of the test. The findings for meta-cognitive judgements highlight that learners perceive distributed rereading of text as difficult, and the findings for reading times suggest that the cognitive effort of readers is increased in distributed rereading. However, our results leave open the question of whether distributed rereading is also a desirable difficulty that should be promoted in school learning.

Chapter V
General Discussion

This dissertation aims to answer the research questions whether (1) distributed learning might be beneficial in learning from complementary texts, (2) whether prior knowledge moderates the effects of distribution on learning and (3) whether distribution affects the judgements of learning similar to distributed practice. Additionally, (4) I wanted to investigate whether distributed practice is beneficial in learning from single texts for school students.

Research Question 1

Four experiments have been reported to answer the first research question. Experiment 1 was conducted in the classroom with seventh graders and showed that distributed learning is not beneficial for retention and even detrimental for the performance in an immediate test. However, an interaction between learning condition and retention interval was found. In the distributed condition, no decrease of learning performance between immediate and delayed test was found, whereas participants in the massed condition showed such a significant decrease. This interaction effect was also found in Experiment 2, which was conducted in the same setting and with a similar sample. Despite this interaction, and contrary to Experiment 1, the massed condition did not outperform the distributed condition at an immediate test. Nevertheless, in the massed condition, a significant decrease between immediate and delayed test was found while no decrease was found in the distributed condition. Experiment 3 was conducted in the laboratory with university students and showed a very similar pattern to Experiment 1: The massed condition outperformed the distributed condition at the immediate test, however this advantage for the massed condition disappeared at a delayed test. Furthermore, a significant decrease between immediate and delayed test was found for the massed condition, but not for the distributed condition. Thus, the experiments show a very similar pattern of results. The results might indicate that distributed practice changes memory , as no or comparably smaller decreases in performance between immediate and delayed tests were observed in all experiments, even in Experiments 2 and 4 where no disadvantage of the distributed condition was found immediately after learning. In line with the assumption of retrieval as "memory modifier" (R. A. Bjork, 1975), one might deduce that distributed learning modified memory in the way that it increases storage strength. As higher storage strength might be beneficial especially in the long run, it could be argued that the retention interval was too short in Experiment 1-3 to show benefits of distributed learning. Therefore, in Experiment 4 an additional test was implemented. The results were very similar to the results of Experiment 1-3. However, no significant differences were found between the massed and distributed condition at any retention interval. Thus, no disadvantage of distributed learning at immediate test nor an advantage two weeks later was found. Furthermore, a decrease between immediate and one-week-delayed test was found for both conditions, but no decrease between the two delayed tests. Thus, even if the interaction effect was significant, the differences between the conditions were small and no beneficial effects of distributed learning emerged after two weeks.

Research Question 2

The second research question was addressed in Experiments 3 and 4. Additional to the main effect of the learning condition, the moderating effects of domain-specific prior knowledge were explored. While main effects of domain-specific knowledge on learning were found in both experiments (as well as in Experiment 1 and 2), moderating effects were only found in Experiment 4. Distributed learning seems to decrease learning outcomes for participants with low prior knowledge.

Research Question 3

The third research question was addressed in Experiment 1 and 2. Beside the effects of distributed learning on learning outcomes, the effects on meta-cognitive judgements were investigated. However, effects of learning condition were only found in Experiment 1, were participants in the distributed condition perceived learning as *less* difficult but predicted lower success. Furthermore, they indicated lower learning success and perceived the texts as being less similar. In Experiment 2, no effects of learning condition on meta-cognitive judgements were found, neither on the perceived difficulty and predicted success, nor on the perceived similarity.

Therefore, the results of Experiment 1-4 suggest that (1) distributed learning is not a beneficial learning strategy in learning with multiple complementary texts. Furthermore, at least in one experiments it seems that (2) even with high prior knowledge, massed learning is as beneficial as distributed learning, whereas distributed learning seems to be even detrimental for participants with low prior knowledge. Additionally, (3) distributed learning is not perceived as more difficult (respectively even as less difficult), but seems to be evaluated as less effective than massed learning, at least in Experiment 1 with a long lag of one week. It should be noted that this meta-cognitive evaluation is only accurate for the immediate test, as no differences were found between massed and distributed learning one and two weeks after learning.

Research Question 4

To evaluate the results and theoretical considerations of distributed learning, it might be fruitful to have a look at the results regarding the fourth research question, whether distributed practice is beneficial for learning from single texts of school students.

This research question was investigated in Experiment 5. Interestingly, the results are quite similar to the results of Experiment 1-4. As Experiment 5 was designed as conceptual replication of Rawson and Kintsch (2005), free recall was assessed, which was not present in Experiments 1-4. However, the results in free recall and the learning comprehension test (which corresponds to the learning outcome measure in Experiments 1-4) were quite similar. In free recall as well as in text comprehension, participants in the massed condition outperformed participants in the distributed condition immediately after learning. However, in a test one week after learning, no differences were found between conditions. In the massed condition, the performance significantly decreased between the tests, whereas in the distributed condition, the performance remained stable. In the comprehension test, a three-way interaction between learning condition, item type and domain-specific prior knowledge was found. In short-answer questions, distributed learning was detrimental for participants with low prior knowledge. As in Experiment 1 and 2, the effects of learning condition on meta-cognitive judgements were investigated. No differences between the condition were found in the perceived difficulty, however, participants in the distributed condition predicted their learning success to be lower than participants in the massed condition. Additionally, the perceived on-task-focus during reading was assessed. Participants indicated lower on-task focus in the distributed condition. Interestingly, this result is contrary to the reading behavior: longer reading times were found for participants in the distributed condition.

In sum, it has to be stated that (4) even distributed practice with single texts was not found to be beneficial for school students.

Explaining The Lack of Beneficial Effects of Distribution

Summarizing the findings of all experiments, the results of Experiment 1-4 and 5 are quite similar, even if they are distinct in the investigated effect, as Experiment 1-4 investigated distributed learning and Experiment 5 investigated distributed practice. Despite the missing beneficial effects of distributed learning and practice in the delayed test, the results of the experiments reported in this dissertation are also quite similar to the results of Rawson and Kintsch (2005) and Rawson (2012): the effects of distributed learning and distributed practice depend on time of test. Therefore, why we did find this interaction between learning condition and retention interval in each experiment, but not the assumed beneficial effects of distributed learning and practice?

Explanation by Field Setting and Sample

As discussed in Chapter II-IV, the lack of beneficial effects in Experiment 5 and Experiment 1 and 2 might be explained by the field setting and the age of the participants. All three experiments were conducted within the classroom and with seventh graders. Therefore, it might be assumed that the field setting is detrimental for the benefits of distributing learning time (either of repeated or complemented learning materials), for example because of a noisier learning environment as argued by Goossens et al. (2016). Additionally, as mentioned in Chapter III, young learners might be less susceptible to the benefits of distributing learning time, especially when applied to learning from texts, because of lower reading skills (Barth et al., 2015; Perfetti et al., 2005). First, the lower reading skills might have resulted in less appropriate encoding of text. Son (2010) reported that distributed practice effects are not found when items have not been appropriately encoded. The low over-all reading comprehension found in Experiment 1, 2 and 5 (e.g. a mean of 38% correct answers immediately after learning in Experiment 2) might indicate that at least some information in the text was not encoded properly. Teachers also gave the feedback that they would think that the texts have been too difficult for “their” students. Second, in learning from text, domain-specific prior knowledge has to be integrated during reading to construct a situation model (Kintsch, 1988; van den Broek et al., 1996). For readers with low reading skills, this prior specific knowledge has to be highly accessible to enable the reader to integrate this prior knowledge during reading (E. R. Smith & O’Brien, 2016). In massed rereading as well as massed reading of complementary texts, the prior knowledge can be assumed to be more accessible than in distributed (re)reading. In massed rereading, the information retrieved during first reading is still accessible and the second reading gives a second opportunity to retrieve more relevant information from memory. In massed reading of complementary texts, the first text provides background information, which is still accessible during reading the second text. Additionally, the prior knowledge might be generally low in seventh graders. This might have affected the encoding in general, but also moderated the effect of distributed learning and practice. Regarding text cohesion, McNamara and colleagues (McNamara, 2001; McNamara et al., 1996; McNamara & Kintsch, 1996; O’reilly & McNamara, 2007; Ozuru et al., 2009) found that for participants with low prior knowledge, low-cohesion texts lead to lower performance, while beneficial effects were found for participants with high prior knowledge. As mentioned in Chapter III, distributed learning might also reduce the text cohesion, because the related sentences become more distant. Thus, prior knowledge might also moderate the effects of distributed learning.

However, these explanations do not apply to Experiment 3 and 4, which were conducted in the laboratory (with less noise) and with university students. Not only that the higher age of participants can indicate better reading abilities, university students are also a selected sample in which reading abilities can be assumed to be high. This is also reflected in the performance of the participants. While in Experiment 1 and 2 a low performance was found in both experimental groups, in Experiment 3 and 4, the participants showed a better general performance, with for example a mean of 59% correct answers for the inferior distributed condition in Experiment 3. Thus, at least for distributed learning, explaining the lack of beneficial effects by the inferior reading skills and the suboptimal learning environment is not sufficient.

Explanation by Domain-Specific Prior Knowledge

Regarding the explanation by lower domain-specific prior knowledge, it can be noted that an interaction effect between learning condition and domain-specific prior knowledge was found in Experiment 4 (but not in Experiment 3). However, even if low prior knowledge was associated with lower performance in the distributed condition, we did not find a benefit of distribution for participants with high prior knowledge. This moderating effect then seems to be insufficient to explain lacking distribution effects in Experiment 1 and 2. Nevertheless, as domain-specific prior knowledge was included as a quasi-experimental variable, the interpretation of results is limited in generalizability. In further experiments, designs with manipulations of domain-specific prior knowledge, e.g. with selection of participants in dependency of domain-specific knowledge, might be fruitful. Additionally, reading ability should also be included because it was also found that reading ability affects the moderating effects of prior knowledge (O'reilly & McNamara, 2007; Ozuru et al., 2009).

Explanation by Inhibiting Limitations

In the discussion sections in Chapters II, III, and IV, I already discussed several further limitations of our experiments that might have inhibited beneficial effects. As mentioned in Chapter II and III, the lack of effects might be explained by a retention interval that was too short, as beneficial effects might appear only a long time after learning. Although we extended the retention interval to two weeks in Experiment 4, this retention interval might still be too short, as for example four weeks or longer are recommended by Rohrer (2015). Thus, in further research I would recommend a longer retention interval to investigate whether beneficial effects of distributed learning and distributed practice (with school students) appear only after a long time. The beneficial effects might then rely on higher retrieval strength in the distributed condition, which leads to a slower decrease over time.

The issue of retention interval is related to the issue of lag. In the five experiments reported here, we used lags of one week (Experiments 1,3,4, and 5) and 15 min (Experiment 2). Therefore, we can evaluate the effects of a very short lag and a quite long lag. However, a lag of 15 min might be too short, as the information given in Text 1 might still be highly accessible during reading of Text 2, while a lag of one week might be too long, as the information of Text 1 might not be retrievable during reading of Text 2. Furthermore, the interdependency of lag and retention interval should be kept in mind. In the reported experiments, my coauthor and I wanted to use a lag and retention interval which resembles those of learning schedules in school and university, in which a course/class is given once a week at the same time. However, as mentioned in Chapter I, the optimal length of retention intervals depend on the length of the lag, as for example Putnam et al. (2017) recommend a retention interval of two month for a lag of one week. It could be argued that the proportion of lag and retention interval might not only affect the amount of benefits, but also the ability to detect any beneficial effects of distribution. Consequently, I would recommend a better-proportioned realization in further research.

Additionally, in Experiments 1-4 presented in this dissertation, the retention interval was implemented as a within-subjects factor, thus, the participants were tested repeatedly. However, repeated testing is a highly effective learning strategy (Bangert-Drowns et al., 1991; Dunlosky et al., 2013; Rowland, 2014). Furthermore, the benefits of repeated testing seems to depend on the retrieval success and, maybe even more important in the context of the current experiments, on the retrievability. S. Greving and Richter (2018) investigated the testing effect in university learning and found that a testing effect was only found for items with high retrievability. Transferred to the context of the current experiments, it could be derived that the implemented repeated testing was only beneficial for the massed conditions, in which the retrievability was high, but not for the distributed conditions, in which the retrievability was low. Moreover, the retrievability, thus, the retrieval strength, is also designated to be low in distributed learning and practice as expounded in Chapter I (and displayed below). Thus, I would recommend between-subjects manipulations in further research of distributed learning and practice. However, it should be noted that a between-subjects realization of retention interval was chosen in Experiment 5 – where we still did not find beneficial effects of distributed practice.

Finally, the standards of coherence might have been low in all experiments. As displayed in Chapter I, the standards of coherence are seen as a benchmark that influences which processes are used by readers to build a situation model. Furthermore, the standards of coherence are a function of the reader, the text and the reading situation (van den Broek et al., 2011). In Experiment 1, 2 and 5, school students read a text that should resemble text book chapters for seventh graders. Thus, the text materials should be perceived as common learning materials and therefor elicit different standards of coherence than for example a narrative text. However, younger readers in general, might have lower or different standards of coherence. If the reading ability is low, the standard might only be to read the text, thus, decode the text base and reach local coherence, not to learn from the text or to integrate the text with prior knowledge. Furthermore, the learning task was highly demanding, while we used a low-stakes assessment of learning outcomes, as already mentioned in Chapters II and III. Therefore, this combination might have also resulted in low standards of coherence. Even if the participants had the opportunity to reread sentences and thus, could use this to ensure encoding, the readers might not have been motivated or might have failed to detect insufficient encoding due to low standards of coherence. In Experiments 4 and 5, the readers were older and more proficient. However, in this sample, the text might have been perceived as less demanding, as the wording was adjusted to seventh graders. This might also have reduced the standards of coherence. Therefore, in further research the factors text, reader and learning situation might be adapted to ensure high standards of coherence.

That said it should be highlighted again that all five experiments came to very similar results, although they differed in sample population, lag, retention interval and repetition character, thus raising the question whether distributed practice and distributed learning from texts can be evaluated as effective learning strategy in real-world learning scenarios, which also might differ in these features.

Theoretical Considerations

Taken together, one could especially cast doubt on the assumption that distributed learning with complementary texts is a desirable difficulty in analogy to distributed practice. The robust interaction effect found in all experiments might alternatively be interpret in terms of the difference between performance and learning (E. L. Bjork & Bjork, 2011; Soderstrom & Bjork, 2015). The tests immediately after reading might have assessed the performance, thus the retrieval strength. The retrieval strength seems to be higher in massed than in distributed practice, as the massed condition outperformed the distributed condition in Experiment 1 and 3. However, this retrieval strength is not a reliable measure of learning, as can be seen by the decrease in performance between immediate and delayed test. Contrary to retrieval strength, storage strength is reliable and stable. Thus, one might deduce that massed and distributed learning lead to a similar amount of storage strength, but massed learning resulted in higher retrieval strength. However, this implies the assumption that retrieval strength is lower in distributed learning with complementary text. As displayed in Chapter I, lower retrieval strength is one prerequisite of a learning situation that strengthens storage strength the most. According to Bjork and Bjork (1992), a lower retrieval strength should lead to an accelerated accumulation of storage strength. In this regard, the manipulation of retrieval strength by adding a lag between complementary texts seems to be successful, however, no beneficial effects on retention were found.

However, lower retrieval strength is a necessary, but not a sufficient condition for beneficial effects of desirable difficulties. Additionally, the storage strength has to be high. However, if the storage strength is too low, even distribution with repetition might fail to enhance learning. I presented evidence for this assumption in the in the form of the relationship between domain-specific prior knowledge in Experiment 4 and Experiment 5, in which participants with low prior knowledge showed lower learning results in the distributed condition. Thus, it can be said that the lower the storage strength, the more detrimental the effects of distributed learning and distributed practice. However, following the theories displayed in Chapter I, it has also to be assumed that the effects of distribution are more beneficial the higher the storage strength is—however, for this effect, I did not find any proof in the current experiments.

This lack of evidence for the assumed link between storage strength and beneficial effects of distributed learning and distributed practice might be explained by the method of assessing prior knowledge. In a recent meta-analysis, it was found that prior knowledge is highly correlated with learning outcomes, as also found in Experiment 1-5, but barely with knowledge gains (Simonsmeier et al., 2021). The authors argue that the knowledge is stable—as indicated by the correlation between prior knowledge and learning outcomes—but the amount of knowledge does not facilitate learning. This might be explained by the multiple mediation hypotheses: prior knowledge might sometimes enhance learning and sometimes impede learning, for example by negative transfer (Simonsmeier et al., 2021). The crucial factor might then not be *amount* of prior knowledge, but its *quality*. Beker and colleagues state that the “ultimate goal of learning is creating a high quality knowledge representation” (2017, p.21). For example, well-structured and interconnected prior knowledge might facilitate knowledge gain by facilitating the comprehension and integration of new information, and prevent negative transfer, while a huge amount of rarely connected prior knowledge might not. From a theoretical point of view, this assumption might be especially true in learning from text. In learning from text, the activation of prior knowledge can assumed to be especially helpful, if this knowledge is well-structured and equipped with rich interconnection. Such prior knowledge enables the reader to (1) activate more appropriate knowledge and (2) integrate the new knowledge in this well-structured network, which helps to retrieve information at a later point of time. If only the amount of prior knowledge is assessed, the effects of high prior knowledge might be lower, as this high prior knowledge might be well-structured or not. However, in low prior-knowledge, it is likely that this knowledge is also less structured (or non-existent at all). Consequently, an assessment of the quality and not the amount of prior knowledge might be promising to detect beneficial effects of distributed learning and practice. Additionally, future research might assess knowledge gains instead of learning achievements to give a better insight in the effects of prior knowledge on learning.

Furthermore, in distributed learning, a lack of (in)direct feedback could explain the lack of beneficial effects. Even if reminding effects might be sufficient to elicit beneficial effects of distribution (Benjamin & Tullis, 2010), the reminding despite repeating has the disadvantage that it provides no feedback. Thus, contrary to distributed practice, the second text cannot be used to relearn the information given in the first text, which might be especially necessary when the storage strength acquired in the reading of the first text is low. As stated in Chapter I, reminding can be perceived as indirect retrieval practice (Tullis et al., 2014). Research regarding retrieval practice has shown that providing feedback is associated with greater testing effects (Kang et al., 2007; Pashler et al., 2005; Rowland, 2014). Moreover, feedback might especially be necessary in learning from texts, as it may compensate for lower or inaccurate comprehension in the first learning occasion. Furthermore, in distributed learning, it might be argued that more time passes between initial learning (e.g. reading the first text) and the test than in massed learning. As this is a design feature of distributed learning, the benefits of distribution had to overcome this confounding factor. Providing feedback, for example by providing the possibility to look back in Text 1 during reading Text 2, might then also be a possibility to reduce the influence of more time between initial reading. On the other hand, participants could be supported in learning from text by learning strategies as for example elaborative retrieval and testing (Dunlosky et al., 2013), to ensure that the information of the first text is sufficiently comprehended and learned before participants read the second text.

Additionally, in the experiments presented in this dissertation, effects of distributed learning on learning outcomes were assessed. This was justified by the assumption of direct effects of distributed learning by more difficult retrieval processes during reading the second text. Nevertheless, as noted in Chapter I, comprehending texts and learning from texts is a very complex, multi-faceted process (Toyama, 2019; van den Broek & Espin, 2012). Consequently, the direct effects of retrieval difficulty might be smaller and less reliable as they are for less complex learning materials as for example vocabulary learning (Donovan & Radosevich, 1999). It might then be fruitful to have a look at the indirect effects on learning, namely, the effects on meta-cognitive monitoring. For example, Thiede et al. (2003) investigated the effects of immediate versus delayed keyword finding after reading on the accuracy of cognitive monitoring, self-regulated learning and learning outcomes. The participants read six texts and were afterwards asked to write down five keywords for each texts either immediately after reading the respective text or delayed after reading all texts. Afterwards, they indicated their learning success and fulfilled a first comprehension test. After a feedback about the comprehension test, the participants had the opportunity to reread the texts. Participants' monitoring accuracy was higher in the delayed-keywords group and furthermore, they were more likely to reread less-learned texts. In consequence, the participants in the delayed-keyword condition outperformed the immediate-keyword condition in the comprehension test *after the rereading*, but not in the test after the first reading. Transferred to distributed learning with texts, one might assume that distributed (re-)reading of texts might also affect self-regulated learning, but only if the reader detect lower comprehension during distributed (re)reading.

For this cognitive monitoring, the retrieval difficulty during reading could be used as indicator. However, participants in our experiments seem to have failed to use the higher retrieval difficulty to adapt their reading. The results regarding the meta-cognitive judgements in Experiments 1, 2, and 5 might be an indicator that the participants—consistent with other research of distributed practice effects (Son & Simon, 2012)—were not able to evaluate the effectiveness of distributed learning during learning. As meta-cognitive monitoring is effortful in text reading (de Bruïne et al., 2021), this might be especially true for readers with lower working memory capacities and lower reading abilities (as the school students in Experiments 1, 2, and 5). Thus, further research might be interested in the research question how readers can be encouraged to use retrieval difficulty during reading for meta-cognitive monitoring. Furthermore, because of the complexity of learning from texts, it might be encouraging to investigate the effects of distributed learning and practice on self-regulated learning. Distributed learning and practice with text might then benefit learning from two sources: direct as represented and investigated in the current dissertation and indirect by different learning patterns as result of distributed learning and practice.

Practical Implications

In summary, it can be assumed that distribution alone is not sufficient to induce beneficial effects on learning from texts. A study of Linderholm et al. (2016) of interleaving and testing in multiple texts might underline this assumption. They investigated interleaved presentation and testing at once in learning from multiple texts and found that testing effects were found for the more complex recall task only when combined with interleaved texts. They concluded that especially in complex tasks it would need more than one desirable difficulty to benefit learning. As this applies to learning from texts, is might be necessary to use at least two effective learning strategies to find reliable beneficial effects on learning outcomes.

Consequently, learning from texts might be a special learning situation, in which the effects of common learning strategies might be less potent than in other learning situations.

It is difficult to deduce clear practical implications from the research reported in this dissertation. As I did neither find evidence for beneficial nor for (stable) detrimental effects of distributed learning and practice with texts, both strategies might be implemented in realistic learning scenarios. However, the generalized advice to distributed learning has to be viewed with some mistrust. This means that recommendations for practitioners should be carefully formulated and restricted to the applied settings and materials for which the beneficial effects of distribution are well proven: for repeated materials with low complexity (Cepeda et al., 2006; Donovan & Radosevich, 1999).

Conclusion

As displayed in Chapter I, text comprehension and learning from texts depend on many factors, which include features of the text, the reader, the reading situation and the task used to assess comprehension and learning. Similar to the effects of text cohesion (McNamara, 2001; McNamara et al., 1996; McNamara & Kintsch, 1996; O'reilly & McNamara, 2007; Ozuru et al., 2009), the effects of distribution (and other desirable difficulties) might not only depend on the domain-specific prior knowledge, but also the reading ability, the standards of coherence, the goal of reading, or on text features as cohesion and the time of test. Further research is needed to investigate under what conditions distributed practice and learning might be beneficial in learning from texts or not. For now, I have to conclude that the general recommendation to space all studying is not applicable to learning from single and multiple texts.

References

References

Agarwal, P. K., & Roediger, H. L. (2018). Lessons for learning: How cognitive psychology informs classroom practice. *Phi Delta Kappan*, *100*(4), 8–12. https://doi.org/10.1177/0031721718815666

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks.

Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1459–1468. https://doi.org/10.1037/0278-7393.21.6.1459

Albrecht, J. E., & Myers, J. L. (1998). Accessing distant text information during reading: Effects of contextual cues. *Discourse Processes*, *26*(2–3), 87–107. https://doi.org/10.1080/01638539809545040

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1061–1070. https://doi.org/10.1037/0278-7393.19.5.1061

Alexander, P. A., & Jetton, T. L. (2003). Learning from traditional and alternative texts: New conceptualizations for the information age. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 199–242). Routledge.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Amstad, T. (1978). *Wie verständlich sind unsere Zeitungen? [How understandable are our newspapers?]* [Unpublished dissertation thesis]. University of Zürich.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behaviour*, *22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, *21*(8), 627–635. https://doi.org/10.1080/00140137808931764

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, *85*(2), 89–99. https://doi.org/10.1080/00220671.1991.10702818

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barth, A. E., Barnes, M., Francis, D., Vaughn, S., & York, M. (2015). Inferential processing among adequate and struggling adolescent comprehenders and relations to reading comprehension. *Reading and Writing*, *28*(5), 587–609. https://doi.org/10.1007/s11145-014-9540-1

Barzagar Nazari, K., & Ebersbach, M. (2019). Distributing mathematical practice of third and seventh graders: Applicability of the spacing effect in the classroom. *Applied Cognitive Psychology*, *33*(2), 288–298. https://doi.org/10.1002/acp.3485

Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting integration of multiple texts: A review of instructional approaches and practices. *Educational Psychology Review*, *30*(3), 973–999. https://doi.org/10.1007/s10648-018-9436-8

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beker, K., Jolles, D., Lorch, R. F., & van den Broek, P. (2016). Learning from texts: Activation of information from previous texts during reading. *Reading and Writing*, *29*, 1161–1178. https://doi.org/10.1007/s11145-016-9630-3

Beker, K., Jolles, D., & Van den Broek, P. (2017). Meaningful learning from texts: The construction of knowledge representations. In J. A. León & I. Escudero (Eds.), *Reading Comprehension in Educational Settings* (pp. 29–62). John Benjamins. https://doi.org/10.1075/swll.16.02bek

Beker, K., van den Broek, P., & Jolles, D. (2019). Children's integration of information across texts: Reading processes and knowledge representations. *Reading and Writing: An Interdisciplinary Journal*, *32*(3), 663–687. https://doi.org/10.1007/s11145-018-9879-9

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Bigot, L. L., & Rouet, J.-F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research*, *39*(4), 445–470. https://doi.org/10.1080/10862960701675317

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Lawrance Erlbaum Associates.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MA: MIT Press. https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/07/RBjork_1994a.pdf

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* (pp. 35–67). Erlbaum. https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/07/RBjork_EBjork_1992.pdf

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

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

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, *74*(4), 245–248. https://doi.org/10.1080/00220671.1981.10885317

Boscolo, P., & Mason, L. (2003). Topic knowledge, text coherence, and interest: How they interact in learning from instructional texts. *The Journal of Experimental Education*, *71*(2), 126–148. https://doi.org/10.1080/00220970309602060

Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval, and time in working memory: Evidence. *Memory*, *6*(1), 37–66. https://doi.org/10.1080/741941599

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J.-F. (1999). Content integration, source separation in learning from multiple texts. In S. R. Goldman, Graesser, Arthur C., & P. Van den Broek (Eds.), *Narrative Comprehension, Causality, and Coherence* (pp. 209–233). Lawrance Erlbaum Associates.

Britt, M. A., Rouet, J.-F., & Braasch, J. L. (2013). Documents as entities: Extending the situation model theory of comprehension. In M. A. Britt, S. R. Goldman, & J.-F. Rouet (Eds.), *Reading–From words to multiple texts* (pp. 160–179). Routledge.

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, *145*(11), 1029–1052. https://doi.org/10.1037/bul0000209

Budé, L., Imbos, T., van de Wiel, M. W., & Berger, M. P. (2011). The effect of distributed practice on students' conceptual understanding of statistics. *Higher Education*, *62*(1), 69–79. https://doi.org/10.1007/s10734-010-9366-y

Bussmann, H. (1996). *Routledge dictionary of language and linguistics*. Routledge.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31–42. https://doi.org/10.1037/0022-0663.96.1.31

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, *34*(1), 30–41. https://doi.org/10.1016/j.cedpsych.2008.07.001

Calvo, M. G., Estevez, A., & Dowens, M. G. (2003). Time course of elaborative inferences in reading as a function of prior vocabulary knowledge. *Learning and Instruction*, *13*(6), 611–631. https://doi.org/10.1016/S0959-4752(02)00055-5

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378. https://doi.org/10.1007/s10648-012-9205-z

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236–246. https://doi.org/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. https://doi.org/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. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Lawrance Erlbaum Associates.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–426. https://doi.org/10.1037/0033-295X.82.6.407

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, *48*, 1227–1237.

D'Agostino, P. R., & DeRemer, P. (1973). Repetition effects as a function of rehearsal and encoding variability. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 108–113. https://doi.org/10.1016/S0022-5371(73)80066-0

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

de Bruïne, A., Jolles, D., & van den Broek, P. (2021). Minding the load or loading the mind: The effect of manipulating working memory on coherence monitoring. *Journal of Memory and Language*, *118*, 104212. https://doi.org/10.1016/j.jml.2020.104212

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. Ross (Ed.), *Psychology of Learning and Motivation: Advances in research and theory* (Vol. 53, pp. 63–147). Elsevier Academic Press. https://doi.org/10.1016/S0079-7421(10)53003-2

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

Dirkx, K. J. H., Kester, L., & Kirschner, P. A. (2014). The testing effect for learning principles and procedures from texts. *The Journal of Educational Research*, *107*(5), 357–364. https://doi.org/10.1080/00220671.2013.823370

Dixon, P. (2008). Models of accuracy in repeated-measures design. *Journal of Memory and Language*, *59*(4), 447–456. https://doi.org/10.1016/j.jml.2007.11.004

Dixon, P., LeFevre, J.-A., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology*, *80*(4), 465–472. https://doi.org/10.1037/0022-0663.80.4.465

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. https://doi.org/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. https://doi.org/10.1177/1529100612453266

Ebbinghaus, H. (1885). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie [Memory: A contribution to experimental psychology]*. Duncker & Humblot.

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, *20*(3), 586–592. https://doi.org/10.3758/s13423-012-0367-y

Fincher-Kiefer, R. (1992). The role of prior knowledge in inferential processing. *Journal of Research in Reading*, *15*(1), 12–27. https://doi.org/10.1111/j.1467-9817.1992.tb00018.x

Fishman, E., Keller, L., & Atkinson, R. C. (1968). Massed versus distributed practice in computerized spelling drills. *Journal of Educational Psychology*, *59*(4), 290–296. https://doi.org/10.1037/h0020055

Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. *The R Journal*, *5*(1), 39. https://doi.org/10.32614/RJ-2013-004

Gagnon, M., & Cormier, S. (2018). Retrieval practice and distributed practice: The case of french canadian students. *Canadian Journal of School Psychology*, *34*(2), 83–97. https://doi.org/10.1177/0829573518773225

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, *4*(3), 49–59. https://doi.org/10.1016/j.tine.2015.01.001

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

Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology*, *79*(2), 198–199. https://doi.org/10.1037/0022-0663.79.2.198

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum: Spacing, memory, and generalization. *Applied Cognitive Psychology*, *28*(2), 266–273. https://doi.org/10.1002/acp.2997

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., Bouwmeester, S., & Zwaan, R. A. (2016). Distributed practice and retrieval practice in primary school vocabulary learning: A multi-classroom study. *Applied Cognitive Psychology*, *30*(5), 700–712. https://doi.org/10.1002/acp.3245

Gordon, K. (1925). Class results with spaced and unspaced memorizing. *Journal of Experimental Psychology*, *8*(5), 337–343. https://doi.org/10.1037/h0064623

Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Greving, C. E., & Richter, T. (2017). *Learning from text: The effects of time between learning episodes*. https://osf.io/9fz72

Greving, C. E., & Richter, T. (2019). Distributed learning in the classroom: Effects of rereading schedules depend on time of test. *Frontiers in Psychology*, *9*, 2517. https://doi.org/10.3389/fpsyg.2018.02517

Greving, C. E., & Richter, T. (2020). *Learning from complementary texts: Is distributed reading a beneficial learning strategy?* Manuscript submitted for publication.

Greving, C. E., & Richter, T. (2021). Beyond the distributed practice effect: Is distributed learning also effective for learning with non-repeated text materials? *Frontiers in Psychology*, *12*, 685245. https://doi.org/10.3389/fpsyg.2021.685245

Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, *9*, 2412. https://doi.org/10.3389/fpsyg.2018.02412

Grote, M. G. (1995). Distributed versus massed practice in high school physics. *School Science and Mathematics*, *95*(2), 97–101. https://doi.org/10.1111/j.1949-8594.1995.tb15736.x

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, *19*(1), 126–134. https://doi.org/10.3758/s13423-011-0181-y

Harzem, P., Lee, I., & Miles, T. R. (1976). The effects of pictures on learning to read. *British Journal of Educational Psychology*, *46*(3), 318–322. https://doi.org/10.1111/j.2044-8279.1976.tb02328.x

Hintzmann, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium.* Lawrance Erlbaum Associates.

Hou, J., Rashid, J., & Lee, K. M. (2017). Cognitive map or medium materiality? Reading on paper and screen. *Computers in Human Behavior*, *67*, 84–94. https://doi.org/10.1016/j.chb.2016.10.014

Inquisit 3 (Version 3.0.6.0). (2011). [Computer software]. http://www.millisecond.com

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

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. https://doi.org/10.1086/374692

Johnston, W. A., & Uhl, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(2), 153–160. https://doi.org/10.1037/0278-7393.2.2.153

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4/5), 528–558. https://doi.org/10.1080/09541440601056620

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, *36*, 38–45. https://doi.org/10.1016/j.learninstruc.2014.11.001

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*(2), 317–326. https://doi.org/10.1007/s10648-015-9309-3

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition*, *38*(1), 116–124. https://doi.org/10.3758/MC.38.1.116

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, *35*(7), 1567–1577. https://doi.org/10.3758/BF03193491

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (1992). A cognitive architecture for comprehension. In H. L. Pick, P. W. van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues.* (pp. 143–163). American Psychological Association. https://doi.org/10.1037/10564-006

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294–303. https://doi.org/10.1037/0003-066X.49.4.294

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology*, *126*(4), 349–370. https://doi.org/10.1037/ 0096-3445.126.4.349

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

Krug, D., Davis, T. B., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology*, *82*(2), 366–371. https://doi.org/10.1037/0022-0663.82.2.366

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

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory*, *20*(1), 37–47. https://doi.org/10.1080/09658211.2011.631550

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. https://doi.org/10.1007/s11251-013-9285-2

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lechuga, M. T., Ortega-Tudela, J. M., & Gómez-Ariza, C. J. (2015). Further evidence that concept mapping is not better than repeated retrieval as a tool for learning from texts. *Learning and Instruction*, *40*, 61–68. https://doi.org/10.1016/j.learninstruc.2015.08.002

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1). https://doi.org/10.18637/jss.v069.i01

Linderholm, T., Dobson, J., & Yarbrough, M. B. (2016). The benefit of self-testing and interleaving for synthesizing concepts across multiple physiology texts. *Advances in Physiology Education*, *40*(3), 329–334. https://doi.org/10.1152/advan.00157.2015

Linderholm, T., Virtue, S., Tzeng, Y., & van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the landscape model. *Discourse Processes*, *37*(2), 165–186. https://doi.org/10.1207/s15326950dp3702_5

Lipowsky, F., Richter, T., Borromeo-Ferri, R., Ebersbach, M., & Hänze, M. (2015). Wünschenswerte Erschwernisse beim Lernen [Desirable difficulties in learning]. *Schulpädagogik Heute*, *6*(11), 1–10.

Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology*, *27*(6), 1424–1429. https://doi.org/10.1037/0278-7393.27.6.1424

Lüdecke, D. (2018). *Sjplot: Data visualization for statistics in social science*. R package version 2.6.2. https://doi.org/10.5281/zenodo.1308157

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 828–835. https://doi.org/10.1016/S0022-5371(69)80050-2

Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, *58*, 61–68. https://doi.org/10.1016/j.ijer.2012.12.002

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mayer, R. E. (1996). Learners as information processors: Legacies and limitations of educational psychology's second metaphor. *Educational Psychologist*, *31*(3), 151–161. https://doi.org/10.1080/00461520.1996.9653263

McCrudden, M., Schraw, G., Hartley, K., & Kenneth, A. K. (2004). The influence of presentation, organization, and example context on text learning. *The Journal of Experimental Education*, *72*(4), 289–306. https://doi.org/10.3200/JEXE.72.4.289-306

McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review*, *27*(6), 1139–1165. https://doi.org/10.3758/s13423-020-01762-3

McDaniel, M. A., Hines, R. J., & Guynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language*, *46*(3), 544–561. https://doi.org/10.1006/jmla.2001.2819

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *55*(1), 51–62.

McNamara, D. S., & Graesser, A. C. (2012). Coh-metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution.* IGI Global. ftp://129.219.222.66/Publish/pdf/McNamara_Graesser_Coh-Metrix.pdf

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247–288. https://doi.org/10.1080/01638539609544975

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In B. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 51, pp. 297–384). Elsevier Science. https://doi.org/10.1016/S0079-7421(09)51009-2

McRae, K., & Jones, M. (2013). Semantic memory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 206–219). Oxford University Press. https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1124&context=psychologypub

Mills, C., D'Mello, S. K., & Kopp, K. (2015). The influence of consequence value and text difficulty on affect, attention, and learning while reading instructional texts. *Learning and Instruction*, *40*, 9–20. https://doi.org/10.1016/j.learninstruc.2015.07.003

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, *91*(3), 488–496. https://doi.org/10.1037/0022-0663.91.3.488

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017–1045. https://doi.org/10.1016/S0191-8869(99)00251-2

O'Brien, E. J., Plewes, P. S., & Albrecht, J. E. (1990). Antecedent retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 241–249. https://doi.org/10.1037//0278-7393.16.2.241

O'reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, *43*(2), 121–152. https://doi.org/10.1207/s15326950dp4302_2

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003

Paik, J., & Ritter, F. E. (2016). Evaluating a range of learning schedules: Hybrid training schedules may be as good as or better than distributed practice for some tasks. *Ergonomics*, *59*(2), 276–290. https://doi.org/10.1080/00140139.2015.1067332

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8. https://doi.org/10.1037/0278-7393.31.1.3

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*, *14*(2), 187–193. https://doi.org/10.3758/BF03194050

Pearson, P. D., & Cervetti, G. N. (2015). Fifty years of reading comprehension theory and practice. In P. D. Pearson & E. H. Hiebert (Eds.), *Research-Based Practices for Common Core Literacy* (pp. 1–24). Teachers College Press.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76. https://doi.org/10.1037/bul0000124

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 227–247). Blackwell. https://doi.org/10.1002/9780470757642.ch13

Priebe, S. J., Keenan, J. M., & Miller, A. C. (2012). How prior knowledge affects word identification and comprehension. *Reading and Writing*, *25*(1), 131–149. https://doi.org/10.1007/s11145-010-9260-0

Primor, L., & Katzir, T. (2018). Measuring multiple text integration: A review. *Frontiers in Psychology*, *9*(2294). https://doi.org/10.3389/fpsyg.2018.02294

Putnam, A. L., Nestojko, J. F., & Roediger III, H. L. (2017). Improving student learning: Two strategies to make it stick. In J. Hattie, J. C. Horvath, & J. M. Lodge (Eds.), *From the Laboratory to the Classroom: Translating the Science of Learning for Teachers* (pp. 94–121). Routledge.

Putnam, A. L., Sungkhasettee, V. W., & Roediger III, H. L. (2016). Optimizing learning in college: Tips from cognitive psychology. *Perspectives on Psychological Science*, *11*(5), 652–660. https://doi.org/10.1177/1745691616645770

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Randler, C., Kranich, K., & Eisele, M. (2008). Block scheduled versus traditional biology teaching—An educational experiment using the water lily. *Instructional Science*, *36*(1), 17–25. https://doi.org/10.1007/s11251-007-9020-y

Rapp, D. N., Broek, P. van den, McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, *11*(4), 289–312. https://doi.org/10.1080/10888430701530417

Rawson, K. A. (2012). Why do rereading lag effects depend on test delay? *Journal of Memory and Language*, *66*(4), 870–884. https://doi.org/10.1016/j.jml.2012.03.004

Rawson, K. A., & Kintsch, W. (2002). How does background information improve memory for text content? *Memory & Cognition*, *30*(5), 768–778. https://doi.org/10.3758/BF03196432

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, *97*(1), 70–80. https://doi.org/10.1037/0022-0663.97.1.70

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, *80*(1), 16–20. https://doi.org/10.1037/0022-0663.80.1.16

Richter, T. (2006). What is wrong with anova and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes*, *41*(3), 221–250. https://doi.org/10.1207/s15326950dp4103_1

Richter, T., Münchow, H., & Abendroth, J. (2020). The role of validation in integrating multiple perspectives. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of Learning from Multiple Representations and Perspectives* (1st ed., pp. 259–275). Routledge. https://doi.org/10.4324/9780429443961-18

Richter, T., & van Holt, N. (2005). ELVES: Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von Teilprozessen des Leseverstehens [ELVES: A computer-based test for measuring efficiency of processes involved in reading comprehension]. *Diagnostica*, *51*(4), 169–182. https://doi.org/10.1026/0012-1924.51.4.169

Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Educational Psychology Review*, *27*, 635–643. https://doi.org/10.1007/s10648-015-9332-4

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. https://doi.org/10.1037/a0037559

Schaffner, E., & Schiefele, U. (2013). The prediction of reading comprehension by cognitive and motivational factors: Does text accessibility during comprehension testing make a difference? *Learning and Individual Differences*, *26*, 42–54. https://doi.org/10.1016/j.lindif.2013.04.003

Schiff, R., & Vakil, E. (2015). Age differences in cognitive skill learning, retention and transfer: The case of the Tower of Hanoi Puzzle. *Learning and Individual Differences*, *39*, 164–171. https://doi.org/10.1016/j.lindif.2015.03.010

Schindler, J., & Richter, T. (2021). *Does generation benefit learning for texts? A meta-analytic review.* Manuscript submitted for publication.

Schlagmüller, M., & Schneider, W. (2007). *Würzburger Lesestrategie-Wissenstest für die Klassen 7-12 (WLST 7-12) [Würzburg reading strategy knowledge test for Grades 7-12]*. Hogrefe.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schneider, W., Korkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, *81*(3), 306–312. https://doi.org/10.1037/0022-0663.81.3.306

Schwartz, B. L., Son, L. K., Kornell, N., & Finn, B. (2011). Four principles of memory improvement: A guide to improving learning efficiency. *International Journal of Creativity and Problem Solving*, *21*(1), 7–15.

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. https://doi.org/10.1002/acp.1066

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2021). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, 1–24. https://doi.org/10.1080/00461520.2021.1939700

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. *Memory & Cognition*, *20*(5), 539–548. https://doi.org/10.3758/BF03199586

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Smith, E. R., & O'Brien, E. J. (2016). Enhancing memory access for less skilled readers. *Scientific Studies of Reading*, *20*(6), 421–435. https://doi.org/10.1080/10888438.2016.1214590

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *1*(3), 341–358. https://doi.org/10.1207/s1532690xci0103_4

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. https://doi.org/10.1002/acp.1747

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance. *Perspectives on Psychological Science*, *10*(2), 176– 199. https://doi.org/10.1177/1745691615569000

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

Son, L. K., & Simon, D. A. (2012). Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review*, *24*(3), 379–399. https://doi.org/10.1007/s10648-012-9206-y

Strømsø, H. I. (2017). Multiple models of multiple-text comprehension: A commentary. *Educational Psychologist*, *52*(3), 216–224. https://doi.org/10.1080/00461520.2017.1320557

Svihla, V., Wester, M. J., & Linn, M. C. (2018). Distributed practice in classroom inquiry science learning. *Learning: Research and Practice*, *4*(2), 180–202. https://doi.org/10.1080/23735082.2017.1371321

Theobald, M., Bellhäuser, H., & Imhof, M. (2018). Identifying individual differences using log-file analysis: Distributed learning as mediator between conscientiousness and exam grades. *Learning and Individual Differences*, *65*, 112–122. https://doi.org/10.1016/j.lindif.2018.05.019

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66–73.

Thios, S. J., & D’Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 529–536. https://doi.org/10.1016/0022-5371(76)90047-5

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing and abstraction. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 113–189). https://doi.org/10.1016/B978-0-12-800090-8.00004-4

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. https://doi.org/10.1016/0022-0965(91)90079-8

Toyama, Y. (2019). *What makes reading difficult? An Investigation of the contribution of passage, task and reader characteristics on item difficulty, using explanatory item response models*. University of California, Berkeley.

Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General*, *143*(4), 1526–1540. https://doi.org/10.1037/a0036036

van den Broek, P., Beker, K., & Oudega, M. (2015). Inference generation in text comprehension: Automatic and strategic processes in the construction of a mental representation. In E. J. OBrien, A. E. Cook, & J. Lorch (Eds.), *Inferences during Reading* (pp. 94–121). Cambridge University Press. https://doi.org/10.1017/CBO9781107279186.006

van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. McCrudden, J. Magliano, & G. Schraw (Eds.), *Text Relevance and Learning from Text* (pp. 123–139). Information Age Publishing.

van den Broek, P., & Espin, C. A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review*, *41*(3), 315–325. https://doi.org/10.1080/02796015.2012.12087512 van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081–1087. https://doi.org/10.3758/BF03206376

van den Broek, P., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A "landscape" view of reading: Fluctuation patterns of activation and the construction of a stable memory representation. In B. K. Britton & A. C. Graesser (Eds.), *Models of Understanding Text* (pp. 165–187). Lawrance Erlbaum Associates.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373–403. https://doi.org/10.1016/j.cognition.2014.01.007

van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*(2), 247–264. https://doi.org/10.1007/s10648-015-9310-x

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., Morris, R. G. M., & Fernández, G. (2014). Building on prior knowledge: Schema-dependent encoding processes relate to academic performance. *Journal of Cognitive Neuroscience*, *26*(10), 2250–2261. https://doi.org/10.1162/jocn_a_00630

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Özsoy, B. (2008). Distributed rereading can hurt the spacing effect in text memory. *Applied Cognitive Psychology*, *22*(5), 685–695. https://doi.org/10.1002/acp.1388

Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives*, *8*(3), 163–168. https://doi.org/10.1111/cdep.12079

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*(4), 1137–1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x von Dülmen, M., Koliussis, C., Ludwig, R., & Schulze, J. (2006). Naturwissenschaftliche Vorkenntnisse deutscher Studienanfänger in der Humanmedizin. *GMS Zeitschrift Für Medizinische Ausbildung*, *23*(4), Doc64.

Voss, J. F., & Silfies, L. N. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction*, *14*(1), 45–68. https://doi.org/10.1207/s1532690xci1401_2

Vössing, J., Stamov-Roßnagel, C., & Heinitz, K. (2017). Text difficulty affects metacomprehension accuracy and knowledge test performance in text learning: Text difficulty affects metacomprehension. *Journal of Computer Assisted Learning*, *33*(3), 282–291. https://doi.org/10.1111/jcal.12179

Weissgerber, S. C., & Reinhard, M.-A. (2017). Is disfluency desirable for learning? *Learning and Instruction*, *49*, 199–217. https://doi.org/10.1016/j.learninstruc.2017.02.004

Wittrock, M. C. (2010). Learning as a generative process. *Educational Psychologist*, *45*(1), 40–45. https://doi.org/10.1080/00461520903433554

Wright, J. (2018). *Definition of the terms cohesion and coherence*. https://www.english.cam.ac.uk/elor/lo/cohesion/index.html

Wylie, J., & McGuinness, C. (2004). The interactive effects of prior knowledge and text structure on memory for cognitive psychology texts. *British Journal of Educational Psychology*, *74*(4), 497–514. https://doi.org/10.1348/0007099042376418

Young, S. H. (2018). Why you cram for exams (and how to stop). *Scotthyoung*. https://www.scotthyoung.com/blog/2018/12/24/why-cram/

Yue, C. L., Castel, A. D., & Bjork, R. A. (2012). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*(2), 229–241. https://doi.org/10.3758/s13421-012-0255-8

Appendix

Appendix A

Text Materials Used in Experiment 1 (English Translations, the Original Texts are in German)

A.1 Biology

Text 1: Plant Cell

Page 1

The plant cell

Today, you will meet the plant cell. In the illustration, you can already see how the plant cell is structured.

As you may already know, cells are the basic components of all living beings. There are many different cells, which differ in shape, size, and function. In your body, there are 100,000,000,000,000 cells, including 210 different cells types that look different and fulfill different tasks. For example, liver cells are responsible for the detoxification of your body, and red blood cells provide your body with oxygen. Every cell in your body has a nucleus, the cell nucleus. All animal and plant cells also have a cell nucleus. Cells with a cell nucleus are called eukaryotes. However, some cells have no cell nucleus. Those cells are called prokaryotes. You will meet such a cell in the next chapter.

Page 2

In this chapter, you will get to know the typical structure of a plant cell, which belongs to the eukaryotes.

If you examine, for example, a membrane of an onion under a microscope, you will see how the cells are separated by a thin wall. This wall is called the cell wall, which mainly consists of the most important building material of plants called cellulose. All plant cells are characterized by such a cell wall made of cellulose. The cell wall provides strength for the cell and separate the cell from the surrounding. You can see in the illustration, that the cell wall has some openings. These openings are called pits. At these places, substances can enter or leave the cell.

The interior of the cell is mostly filled with a liquid or gel-like mass called the cell plasma. This cell plasma is surrounded by a thin layer called the cell membrane.

Page 3

In the cell plasma, you can find the cell nucleus, which you already briefly encountered above. The cell nucleus is the carrier of the genetic information of the cell. Thus, it contains the "construction plan" for the entire cell. The cell nucleus is also surrounded by a membrane, which separates the cell nucleus from the cell plasma.

The cell plasma, of course, contains the cell nucleus but also different organelles, which means small organ. Organelles undertake independent tasks and thus are somewhat comparable with organs of multicellular organisms. One of those organelles is the Golgi apparatus, which is also called the "post office" of the cell. Here, substances are sorted, packed in small bubbles and sent to their destinations. One substance sorted by the Golgi apparatus is protein. Protein is produced by another organelle called a ribosome. Proteins can be used, for example, to build membranes.

Page 4

Cells also need a "power plant" to provide the cell with energy. For this task, another organelle is responsible, called the mitochondria.

The biggest cell organelles are the vacuoles. Vacuoles are small bubbles that merge over time into one big bubble. This bubble stores important nutrients and water for the cell.

One last organelle is the chloroplast, which is the solar plant of the cell in which photosynthesis takes place. Through photosynthesis, plant cells can produce energy from light.

In the illustration, you can see all components of the cell again.

Text 2: Bacterial cell

Page 1

The bacterial cell

In the last chapter, you already learned several facts about cells. Today, you will meet a new cell: the bacterial cell. In the illustration, you can already see how the bacterial cell is structured.

Many people only know bacteria from illnesses, for example, from infected tonsils. If you have such an illness, the doctor will most likely prescribe antibiotics. Antibiotics are a type of medication that kills disease-causing bacteria, which helps ill people get well again.

But not all bacteria are harmful. Many useful bacteria exist in the world that clean water and decompose organic waste. Without them, human life would not be possible. Countless good bacteria also live on and in our bodies. In the intestine, for example, numerous bacteria help us digest our food.

Page 2

Bacteria are the simplest life forms on earth. They are unicellular organisms, which means that they consist of only one cell. Bacterial cells differ in size and structure from plant cells, which you have already seen in the last chapter. In contrast to plant cells, bacteria belong to the prokaryotes. Thus, the genetic material of the bacterial cell "floats" in the cell plasma without a sheath. In addition, bacteria are much smaller than plant cells. They are only 2-3 micrometers in size, which is 2-3 millionths of a meter.

Some bacteria also have ring-shaped genetic components floating in the cell plasma, the so-called plasmids. The plasmids contain very important information and enable the bacterium to develop resistance to antibiotics, which means that the bacterium becomes well protected from the antibiotic to the extent that the bacterium can no longer be harmed by the antibiotic. It can happen that the medicine prescribed by the doctor will no longer help the patient with tonsillitis because it cannot fight the disease-causing bacteria.

Page 3

The plasmid rings can even be exchanged between two bacteria and thus spread. This exchange is of course a problem when fighting diseases caused by bacteria. In the beginning, it may only be a certain bacterium against which the drug is no longer effective, but eventually an increasing number of different bacteria also become resistant to the drug.

The exchange of the plasmids takes place via so-called pili. The pili are small cell appendages with which bacteria can dock to other bacteria to exchange plasmid rings.

All reactions that are important for the bacterial cell, such as nutrient degradation and energy production, take place in the cell plasma. There are some similarities but also differences between eukaryotic plant cells and prokaryotic bacteria. One common feature is that bacterial cells, like plant cells, also have ribosomes, which you already learned about in the last chapter.

Page 4

To protect themselves from external attacks, bacteria have a cell wall that also provides strength. The cell is additionally surrounded by a kind of mucous layer to prevent dehydration. Inside, the cell wall is coated with a cell membrane, which has a very important function. It determines which substances are allowed to leave the bacterial cell or enter and remain in the cell.

Although a bacterial cell does not have organelles such as the Golgi apparatus, mitochondria or chloroplasts, a few species - like plant cells - can carry out photosynthesis.

In addition, bacteria often have bacterial flagella, which help them to move. The flagella are thread-like structures on the surface of the cell. They rotate like a propeller, creating suction or pushing the bacteria forward.

In the picture you can see all parts of the bacteria cell again.

A.2 Physics

Text 1: The relationship between heat and mechanical work

Page 1

The relationship between heat and mechanical work

Have you ever thought about what "heat" is? In the 18th century, people were still convinced that heat was a weightless substance called Caloricum. At that time, for example, the expansion of a thermometer fluid was explained as follows: at higher outside temperatures, more of this substance, (i.e. Caloricum) enters the pores of the thermometer fluid and the fluid therefore expands.

In the middle of the 19th century, however, a number of scholars concluded that heat is not a substance but has something to do with energy. Perhaps you have already learned the term energy. According to Lord Kelvin, energy is the ability of a body to perform mechanical work. Thus, it can be briefly described as the working ability or effectiveness of the body. For example, energy is motion energy, which is also called kinetic energy or potential energy. Potential energy, for example, is positional energy, like the apple that hangs from the tree has gravitational potential energy that is released when it falls from the tree.

Page 2

In 1842, the German physician, Robert Meyer, summarized his findings as follows: “Falling mass (i.e. the potential energy), motion (i.e. the kinetic energy), heat, light and electricity are one and the same object in different manifestations.”

About the same time, the English brewer, James Prescott Joule, suspected the existence of an equivalence between mechanical motion and heat. To investigate this relationship, he conducted an experiment in 1843. Mechanical energy was generated in a quantity of water that was thermally insulated, that is, protected from the outside temperature. Then, the temperature increase of the water was measured. In the figure you can see the test procedure. In a container is the insulated water. In the water, fins are connected to a paddle wheel, which when turned agitate the water.

Page 3

The shaft of the paddle is then turned by descending weights. The weights thus use their positional energy to turn the shaft. The fins of the paddle wheel then rotate in the water and heat it up. The temperature change can then be precisely measured in the tank. Based on the design, Joule was able to establish a very precise relationship between the potential energy, that is, the potential energy of the weights and the temperature increase of the water.

Page 4

The Principle of Conservation of Energy

This result made it possible to transfer assumptions that had already been formulated about mechanical energy to thermal energy.

For mechanical processes, it had already been determined that the energy of a closed system remains constant. In 1848, the German physicist Helmholtz formulated the general principle of conservation of energy in his writing, "On the Preservation of Force," which is formulated today as:

In a closed system in which any (mechanical, thermal, electrical, chemical) processes take place, the available total energy is retained.

Therefore:

$E = E1 + E2 + E3 + ... + En$

E	Total energy
E1, E2, ...	Energies in their different forms

Page 5

This sentence suggests that energy in a closed system is never lost and never generated: It can only be converted.

A closed system

What is a closed system? A closed system is a system from which energy is neither supplied from outside, nor is energy withdrawn from inside the system, that is, the system is sealed off and isolated from the environment. Therefore, no interaction can occur between the system and the environment.

In this closed system, however, the existing energy can be transformed, that is, from one form into another. An example of this transformation are the processes in a power plant. The chemical energy contained in crude oil is first converted into thermal energy in a power plant. This heat energy, which is also called thermal energy, is then converted in the generator into motion energy of rapidly rotating components, which is then converted into electrical energy.

Text 2: The inner energy of a system

Page 1

The internal energy of a system

In the last text, you already learned about energy and a closed system. In this text, you will now learn about the internal energy of a system.

As you know, energy occurs in different forms and can be transformed from one form into another. The energy can also be transferred from one system to another. For example, one system can release heat while reducing its own thermal energy, and another system can absorb that heat and increase its thermal energy.

The internal energy of a system is stored energy. It is therefore composed of the different forms of energy that exist within a system.

Page 2

Determining the internal energy of a system is not easy, but fortunately in most cases, this value is not needed. To describe the processes taking place, the changes in the energies of the bodies and systems involved can be recorded. For example, when a piston moves through a glass as the gas expands and cools, the temperature of the gas (i.e. its thermal energy) changes. But all other forms of energy that are part of its internal energy remain unchanged and have nothing to do with the flask being moved. For this reason, knowing the other forms of energy or the internal energy of the system is not important.

But as you can see from the example with the piston, the internal energy of a system can change. When work is performed by a system, its internal energy decreases. In contrast, when work is performed on a system, its internal energy increases.

Page 3

Not only the performance of work can change the internal energy of a system. When heat is added to or removed from the system, its internal energy changes. For example, if a gas is heated, its internal energy increases. If the gas releases heat to another system, the internal energy of the system decreases.

As you have already learned, all forms of energy can be converted into each other. The principle of conservation of energy you learned in the last chapter applies. In the principle of conservation of energy, one assumes a closed system. Thus, it is important that the system is closed. For example, when the electricity generated in the power plant (remember how that happened?) is used to heat a pot of soup, not all the chemical energy that was originally in the oil will arrive as heat in the soup. Why does this loss occur? When one form of energy is transformed into another, energy will always be lost, for example, through friction.

Page 4

This loss of energy means that a little heat is always released into the environment during this conversion. Therefore, the system, which consists of the power plant and the pot of soup, is not closed.

The fundamental theorem of thermodynamics

When a system in which only thermal and mechanical forms of energy occur, a special form of the principle of the conservation of energy applies. This principle is also called the fundamental theorem of thermodynamics.

The fundamental theorem of thermodynamics refers to the change of the internal energy U. Since only thermal and mechanical forms of energy occur in the special system, only the performance of work or the removal or supply of heat can change the internal energy of this system.

Page 5

The first fundamental theorem of thermodynamics

The change of the internal energy U is equal to the sum of the mechanical work A performed by or on the system and the heat Q supplied to or removed from the system.

The following applies:

$\Delta U = A + Q$

The "triangle" in front of the U is the Greek letter Delta, which denotes a change, that is, a change in internal energy U.

Appendix B

Table B1

Intercorrelations in the Massed Learning Condition (Lower Triangular Matrix) and the Distributed Condition (Upper Triangular Matrix) of All Measured Variables in Biology in Experiment 1

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Learning outcome (immediate)	—	.51	.24	.4	.33	-.10	.30	-.42	.37	-.02	-.03
2. Learning outcome (delayed)	.56	—	.49	.49	.41	.18	.54	-.29	.51	.23	.07
3. Perceived difficulty	.36	.42	—	.44	.20	.20	.32	-.28	.21	-.18	-.24
4. Self-predicted success	.24	.07	.37	—	.20	.14	.16	-.29	.26	-.04	.20
5. Perceived similarity	.00	.17	.25	.39	—	.34	.27	-.36	-.02	.09	-.21
6. Perceived learning coherence	-.08	-.18	-.17	.30	-.01	—	-.06	-.14	-.05	.17	.00
7. Prior knowledge	.56	.43	.45	.34	.14	.17	—	-.43	.21	.12	-.03
8. Grades	-.32	-.12	-.13	-.08	-.15	-.13	-.37	—	-.24	-.07	.21

9. Reading ability	-.14	.22	.26	-.21	-.11	-.16	.11	-.01	—	.63	.42
10. Working memory capacity	.18	.04	.49	.27	.11	.20	.24	-.36	.15	—	.57
11. Reading strategy knowledge	.07	-.17	.07	.11	-.33	.21	.14	-.34	.17	.32	—

Table B2

Intercorrelations in the Massed Learning Condition (Lower Triangular Matrix) and the Distributed Condition (Upper Triangular Matrix) of All Measured Variables in Physics in Experiment 1

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Learning outcome (immediate)	—	.23	.02	-.06	.27	.07	-.11	-.34	.59	.34	.07
2. Learning outcome (delayed)	.68	—	.04	.11	.19	-.31	.39	-.19	.28	.23	.22
3. Perceived difficulty	.12	.17	—	.45	.28	.37	-.04	-.01	-.17	-.39	-.28
4. Self-predicted success	-.10	.03	.57	—	.41	.33	-.08	.35	-.24	-.19	.02
5. Perceived similarity	-.28	.07	.17	.42	—	.43	-.27	-.01	.07	.01	-.1

6. Perceived learning coherence	-.32	-.06	.51	.37	.24	—	-.28	.21	-.2	-.37	-.03
7. Prior knowledge	.34	.58	-.03	-.07	.14	.02	—	-.17	.17	.1	-.19
8. Grades	-.30	-.24	.03	.09	-.14	.04	-.56	—	-.26	-.22	.07
9. Reading ability	.09	.19	.08	-.27	.12	-.24	-.12	-.01	—	.63	.4
10. Working memory capacity	.14	.02	.28	-.13	.05	.16	.33	-.17	.17	—	.56
11. Reading strategy knowledge	.20	.09	-.04	-.27	-.18	.12	.08	-.22	.17	.32	—

Appendix C

Instructions Used in Experiment 1 (English Translations, the Original Instructions Were Provided in German)

The main instructions with explanations of the tasks were provided in Session 1. The instructor read the instructions out loud, while the participants read the instruction on screen and solved some examples. In the experimental Session 2-4, the participants received short instructions by the computer program before the respective task, for example “Now you are asked to read a text. Do you remember how you can use the keys to read the text?” or “Now we are going to ask you some questions about the text you just read.”. No further instructions were given. Furthermore, the participants did not know about the schedule, thus, when they will get which task. The students were informed about the domains (physics, biology) by the teachers ahead of the start of the experiments.

In the following, we will present a translation of parts of the main instructions given in Session 1.

C1. General Instructions.

“[…] Your participation is very important to us. We need your honest opinion and your best performance because we want to find out how students learn best. […]”

C2. Instruction of Questions

“In some tasks we will ask you questions. I will now show you an example of what this looks like. Please press the “R” key to get to the example.

In the middle of the screen, you can see the question “Who is Harry Potter’s best friend?”. Below the question, you see a white box to type in your answer. When you have finished typing your answer, click on the button marked “Continue” to move on to the next question. You cannot go back to the questions once you clicked “Continue”.

Try to answer this question. Type your answer in the box below the question. When you are satisfied by your answer, click on the button marked “Continue”.

[…]

Very good. Please press the key with the “E” on it to go to the next page. In the next type of question you will get the question “What is the name of Harry Potter’s best friend?” and four possible answers: “Lucius Malfoy”, “Ron Weasley”, “Fred Weasley” and “Samuel Jackson”. Only one of these four answers is correct! Click on the answer you believe to be correct. If you are not quite sure, click on the answer you think is most likely to be correct. Again, if you click on the “Continue” button, you will get to the next question, but you cannot go back to the question afterwards.”

C3. Instruction of Text Reading.

“In the next sessions, we ask you to read texts. But these text will be presented to you in a special way. How this looks exactly, I will show you with an example. Now press the key “E” to get to the text. As you can see, you can only read the headline of the text yet. If you want to read the next sentence, press the right pointing arrow on the keyboard.

Now press the arrow pointing to the right. This will lead you to the next sentence. Very good, now you can read this sentence. […] As you can see, you can’t read the headline now. But if you want to read the headline again, you can press the key with the arrow pointing to the left. Press now the key with the arrow pointing to the left. Do you see? You can now read the headline again. By pressing repeatedly the key with the arrow pointing to the right, you can read through the text till the last sentence. If you would like to read a sentence again, you can use the key with the left pointed arrow to go back to this sentence.”

Acknowledgements

First, I would like thank my supervisor, Professor Dr. Tobias Richter for his support and guidance in all phases of this dissertation. Even if I did not follow each career advice, I learned very much from every single one. I would also like to offer my thanks to Professor Dr. Mirjam Ebersbach and Professor Dr. Wolfgang Lenhard for their time and interest spent in this dissertation.

I am also very grateful to the Hessian Ministry for Science and Art, who supported the research presented in this dissertation. Moreover, I would like to thank my student assistants Melanie Dreher, Alice Hodapp, Magdalena Hofmann, Dorothee Gantert, Marina Klimovich, Vanessa Michalski, Daria Mundt, Fabienne Sielaff, Lara Tebbe, Lea Mara Vorreiter and Sarah Wittkopf for their help in data collection and scoring. Without this womanpower, the conduction of the school experiments would not have been possible.

As it took some time (some would say too much time) to complete my research and this dissertation, many colleagues from the University of Kassel and University of Würzburg accompanied my way. All of them made this way enjoyable– and I am very grateful for the lunches, the laughs, the dances, the teatime, the fruitful discourses, and, of course, the coffees. I really adored working with and learning from such inspiring people. Particularly outstanding was my time in the “Systembau” and I would like to thank my fellows from the LOEWE-project, especially Katharina Barzagar Nazari, Irina Streich, Stella Pede and Anne Cohonner for the support and shared experiences. Aside from that, I am especially grateful for the Wednesday writing sessions with Anita Körner, as I don’t know if I would sit here, writing my acknowledgments (on a Wednesday evening), if it were not for our regular appointments in the last years.

I would also like to thank Dres. Schindler, not only for the always encouraging and inspiring scientific discussions and maybe even more inspiring trash TV sessions, but also for emotional support and for sharing the challenges of compatibility of family and career. I am also so grateful for the love and support, the comforting words, the honesty of Denise Dinnebier, for the blissfully happy times with Seifbiers and Schrenks, for the Corona-Gang with Hanske-Vogts, which helped me to go through the last years.

Finally, I would like to thank my parents, Hermann und Susanne Greving, for their love and support. I am also grateful for the support of my parents-in-law, Birgit und Steffen Platzek, as it is a special gift to have not only one but two families behind me. Last but not least, I would like to thank my closest ones: Sven Greving, my husband and favorite colleague, for having my back, for his serenity, for being there in the most challenging moments; and my wonderful children, Kornelius and Theodora, for bringing the joy and so much love in my life, for your patience and for showing me what really matters.