# HOW TO PROMOTE RELATIONAL PROCESSING WHILE LEARNING WITH EXPOSITORY TEXTS

## INCORPORATING INTERLEAVED SEQUENCES AND GENERATION TASKS TO HIGHLIGHT AND BRIDGE THE COHESION GAPS

DISSERTATION

in partial fulfillment of the requirements
for the degree of Doktor der Philosophie
(Dr. phil.) in Psychology

submitted to
Fachbereich Humanwissenschaften
(FB 01) Universität Kassel

by
Roman Abel, Master of Science
Kassel, November 2020

| | |
|---|---|
| **Dean of the faculty of human sciences:** | Prof. Dr. Theresia Höynck |
| **Supervisor:** | Prof. Dr. Martin Hänze |
| **Thesis reviewers:** | Prof. Dr. Ralf Rummer |
| | Prof. Dr. Julian Roelle |
| **Date of disputation:** | 13.01.2021 |
| **Note:** | Summa cum laude |

# Content

## Summary

Learning from expository texts encompasses a high range of learning goals and demands. Learners have difficulties in organizing and integrating the content with previous knowledge (i.e., *relational processing*). Learners thus often fail to construct a coherent representation of the learning content by going rarely beyond a shallow text representation. The aim of the present dissertation is hence to provide recommendations on how to facilitate relational processing while reading expository texts.

Relational processing can be supported by manipulating text-characteristics – for example, by making the relations across information units in text explicit (i.e., increasing *cohesion*). Especially learners with a low level of prior knowledge depend on the guidance provided by cohesion devices (e.g., causal connectors). However, previous research also indicates a mismatch between cohesion and coherence – from henceforth we will refer to this mismatch as the *cohesion-coherence-mismatch*: A fully cohesive text provides the instructions for how to establish a coherent mental representation but lowers the necessity of relational processing due to the lack of *cohesion gaps*. *Cohesion gaps* stand for essential relations across information units that are not explicated in the text, but have to be bridged by readers themselves. A poorly written text thus does not provide the necessary instructions, but engages readers in relational processing in order to close the cohesion gaps.

Reading strategies that require active processing – i.e., *generative learning tasks* – also engage learners in relational processing. However, the learning success is strongly affected by learners' generation accuracy, which in turn depends on domain specific previous knowledge. Despite this link, high skilled learners do not require cognitive stimulation by a generation task because – different from low skilled learners – high skilled learners are spontaneously engaged in relational processing. Accordingly, there is a mismatch between learners' aptitude to accurately generate inferences and the necessity for engaging learners to do so: high skilled learners are capable of accurately generating inferences but do not require it, low skilled

learners in contrast require the stimulation by a generation task but lack the capability to accurately generate inferences. From henceforth we will refer to this mismatch as the *ability-requirement-mismatch.*

Against the background of the cohesion-coherence-mismatch and the ability-requirement-mismatch, we designed two learning tools to promote relational processing (especially) in less skilled learners – *interleaving of information units* and a *causal cohesion generation task*. Interleaving and generation are considered *desirable difficulties* in *initial* learning (which is different from spacing and testing, which promote consolidation processes). Interleaving of information units refers to the manipulation text sequence, whereas the causal cohesion generation task refers to the manipulation of learning instruction.

Both tools were designed based on the following common presumption, which frames the present dissertation: Learners might fail to take learning advantages of cohesion gaps due to two reasons – if they lack the domain specific knowledge necessary for generating inferences *and* (even if they have the necessary knowledge) if they fail to recognize a cohesion gap in the first place. If a cohesion gap were *invisible* for a learner, he or she would not make any efforts to retrieve related idea units from memory to make an *elaborative inference* (irrespective of the level of previous knowledge) or to connect information units from the text (i.e., making a *bridging inference*). Investigating learning tools that highlight the cohesion gaps and support learners in bridging information to close those gaps (especially if learners lack the domain specific previous knowledge) holds therefore educational value.

The theoretical possibility that learners might fail to detect the cohesion gaps is so far neglected in reading comprehension research and instructional science. The present work thus addresses this shortcoming: Across four experiments, the learning tools – interleaving (Experiments 1a and 1b) and cohesion generation (Experiment 2a and 2b) – were tested with respect to learning processes (e.g., inferences while reading and cognitive load via a dual task) and outcomes (e.g., text-based representation and situation model when immediately tested and delayed). Both tools were supposed to increase the visibility of cohesion gaps and reduce the dependency on the domain specific previous knowledge in closing those gaps.

*Interleaving of information units* means to rearrange the sentences within a text in a way that characteristics of various categories are juxtaposed, as opposed to a *blocked* text, in which categories are presented one by one. Interleaving should thus increase the contrast between categories and in turn enable comparisons. The expository texts used in Experiments 1a and 1b lacked cohesion; that is, the texts lacked any relational statements relating multiple information units (e.g., comparisons, co-occurring patterns, or underlying principles such as functionality), but consisted purely of factual statements. Those *missing* relational statements (=cohesion gaps) could be concluded based on comparisons across factual statements (that is, independently of domain specific previous knowledge). Because of the contrast, we expected the readers of an interleaved text to be more likely to detect the cohesion gaps (than readers of a blocked text) and bridge information to close these cohesion gaps. The results of Experiments 1a and 1b confirmed our expectations: Based on comparisons made while reading an interleaved text (i.e., *comparative inferences*), learners became aware of the missing relations across multiple information units and closed these gaps by concluding on underlying regularities across categories (*inductive inferences*). Both experiments demonstrated the superiority of interleaving over blocking in terms of short- and long-term learning advantages for young (less skilled) and more advanced readers.

For the implementation of the causal cohesion generation task, causal connectives (such as *therefore*, *because*, *however*, and *although*) were removed from the expository text, but *explicit* conjunction gaps were left over. Readers were then instructed to establish a causal relation between two clauses for each gap by selecting the appropriate connective. To be able to select the correct connective for each gap, readers were required to reflect on causal relations between clauses. We assessed learners' previous knowledge (and reading skill) to determine the extent of dependency between the previous knowledge and generation accuracy. In the control condition, readers received a fully cohesive text (Experiment 2a and 2b). Experiment 2b additionally used a non-cohesive text, which lacked not only the causal connectives, but also any indication of their absence (i.e., *implicit* gaps). Thus, only in the generation condition, the gaps were visible. We therefore expected the poor readers to benefit from being engaged in relational processing by visible cohesion gaps. Poor readers in the generation condition should thus outperform their counterparts in the control conditions. The results confirmed our expectations: Poor readers – and especially those who succeeded to accurately close the cohesion gaps – showed sustainable learning in terms of text-based representation and situation model. The cohesion gaps in a non-cohesive text, in contrast, remained invisible to poor readers and consequently did not engage them in relational processing. Only the high skilled readers with a high level of previous knowledge benefited from reading a non-cohesive text because they were able to detect and close the cohesion gaps.

Based on the pattern of results across the four experiments, the potential of learning tools that highlight the cohesion gaps in expository texts and support learners in closing those gaps could be demonstrated. As a rule of thumb: Cohesion gaps promote relational processing if learners have the necessary proficiencies to overcome the demands imposed by a non-cohesive expository text, that is, to detect and close the cohesion gaps – if not, learners require learning aids that compensate for the invisibility of cohesion gaps and the lack of ability to close them.

# Introduction

## Reading Expository Texts: Demands and Struggles

Expository texts are a major medium of scientific knowledge. Expository texts are composed of factual descriptions and detailed explanations of scientific phenomena such as the *greenhouse effect* or *life of marine mammals*, which we used for the present work. As a medium of scientific knowledge, expository texts convey its informational density, complexity, interconnectedness across concepts, and multi-causal relations (cf. Britt et al., 2014), inherently resulting in a high element interactivity (Sweller, 2010).

Learning from expository texts encompasses a range of goals and demands. To understand scientific contents, readers are required to connect and integrate the concepts into a coherent mental representation (cf. van den Broek et al., 2015; Zwaan & Radvansky, 1998). Furthermore, because the extent of zoom-in into the network of connections has its limitations, several relations across concepts are not explicitly addressed in the text, but remain implicit. Thus, several sentences are interconnected by implicit links, imposing the demand on readers to detect the gaps and infer the links by themselves. To explore how readers deal with such demands, we used expository texts that either consisted purely of factual but lacked any relational statements (Experiments 1a and 1b) or were manipulated with regard to the extent that relations are explicit (Experiment 2b).

To conclude on implicit relations among information units, readers are required to close the cohesion gaps. This can be achieved by accessing and integrating information with previous knowledge, that is, making *elaborative inferences*. To link remotely placed idea units, readers are required to navigate among sentences and make *bridging inferences* (McNamara et al., 1996). To achieve deep comprehension, readers are also required to make generalizations based on explicit ideas in the text, that is, to discovery related ideas, underlying regularities, general patterns, and principles.

Unexperienced readers usually struggle with expository texts because the content and the macrostructures of the text are unfamiliar to them (Cook & Mayer, 1988; Lorch, 2015; Meyer, 1975). That is – differently than with narratives – readers lack the superstructural knowledge of expository texts. Apart from informational density of expository texts, the multi-causality of scientific phenomena is especially challenging for readers (cf. Britt et al., 2014). Consequently, readers struggle with selecting, organizing, and integrating the main text ideas. Readers thus often fail to construct a coherent representation of scientific phenomena. Studies on reading comprehension show that most learners stringently follow the text linearly, make no efforts of looking back on text contents (Hyönä et al., 2002), and simply focus on the immediate context (Cook & Mayer, 1988; Coté et al., 1998). Readers then fail to establish links between distant sentences. Poor readers may especially struggle in establishing coherent representations of scientific phenomena. As opposed to higher skilled readers, poor readers struggle with bridging inferences based on distant sentences and integrating novel content with previous knowledge (Hannon & Daneman, 2001). These processes are though essential for the situation-model construction (Kintsch, 1988). Providing recommendations for increasing the readability of expository texts and facilitating relational processing while reading is therefore one essential aim of instructional science in general and of this work in particular.

## Instructional Science: Ways to Support Relational Processing while Reading Expository Texts

There are two ways to promote rational processing while learning from expository texts, either by manipulating the expository text-characteristics or by directly engaging readers in relational processing via a generative learning instruction.

### *Via Text-characteristics*

Relational processing can be supported by providing readers with a well-designed text. We consider a text as well designed if it supports readers in making inferences essential for the learning objective. In the following, we will consider two broad clusters of text-characteristics, *text sequence* and *text cohesion*. It is important to note that sequence and cohesion are independent text-characteristics. Studies on impact of cohesion accordingly manipulate the level of cohesion in text without changing its sequence. Analogously, a study on impact of text sequence manipulates only the sequence, but holds the content constant across conditions. However, it should be noted that a random sentence order (de Jonge et al., 2015) and mixing multiple lines of argumentation (Roelle & Nückles, 2019) apparently preclude cohesion by making the relations more difficult to establish.

**Sequencing.** To relate information (i.e., to make a bridging inference), information units need to be processed simultaneously. Because of the working memory capacity constraints adjacent information units are more likely to be processed simultaneously than distant information units (Kintsch, 1988). The degree to which learners are supported in relational processing may thus depend on the sequence in which information units are presented (Wiley & Myers, 2003). In this sense, the likelihood of bridging two information units can be considered a function of their proximity (McKoon & Ratcliff, 1992). For example, due to a close succession of object characteristics in a canonical – *object-oriented* – text sequence, the object characteristics should be simultaneously processed and integrated into a coherent representation (cf. Kintsch & van Dijk, 1978). In contrast, if the learning objective requires learners to discriminate among categories, an *aspect-oriented* presentation sequence (i.e., *interleaved*), which juxtaposes the categories, may be superior to the canonical object-oriented sequence (i.e., *blocked*), which maintains the continuity of category presentation (cf. Schnotz, 1984).

Based on the assumption that information that is processed simultaneously is likely to be related (Wiley & Myers, 2003), manipulating the sequence of information units on the local level seems to be the basic way of supporting readers in establishing essential links. Against the background of this rationale, it might appear surprising that the learning impact of sequencing is barely investigated by the research on text comprehension. The present work addresses this shortcoming by extending the literature with Experiments 1a and 1b.

Sequence effects on learning from textual materials are not restricted to the local sentence-by-sentence level *within* a text. Especially the research on interleaving demonstrated the impact of sequencing *across* textual segments and whole texts on category learning. *Interleaving* stays for an alternating order of category presentation. Thereby, studies have shown beneficial effects of interleaving on categorization of e.g., psychological disorders (Zulkiply et al., 2012; Zulkiply, 2013) and crime cases (Helsdingen et al., 2011). A recent study by Maier et al. (2018) provided further evidence that the positive impact of interleaving goes beyond categorization, but improves also the processing and comprehension of belief-inconsistent information.

The discriminative contrast hypothesis attributes the categorization advantage of interleaving to the *discriminative contrast* among categories: When categories share many characteristics (e.g., psychological disorders share many symptoms), juxtaposition of categories via an interleaved sequence highlights the subtle differences that are predictive for the category membership (Birnbaum et al., 2013; Kang & Pashler, 2012; Mitchell et al., 2008; Zulkiply & Burt, 2013). Accordingly, the discriminative contrast can be created between sentences (on the local level) and larger segments (on the global level). However, a discontinuous sequence alone is not sufficient to establish the discriminative contrast. The alternating information should be semantically related and comparable – like the cases of two psychological disorders are. Studies that mix semantically unrelated textual materials yielded consequently no interleaving effects (Dobson, 2011; Hausman & Kornell, 2014; Mandler & DeForest, 1979).

***Aptitude x Sequence Interaction.*** Against the background that there is generally little research on the impact of sequencing on learning from expository texts, the number of studies investigating the moderating role of learners' proficiencies on sequence effects must inherently be low. Schnotz (1982) reported a positive correlation between the previous knowledge and learning outcomes only for readers of an aspect-oriented (=interleaved) text, but not for readers of an object-oriented (=blocked) text. He explains this interaction pattern with additional navigating demands imposed by a discontinuous layout of an aspect-oriented text; previous knowledge then is supposed to compensate for additional navigating demands. In line with this reasoning, Wiley and McGuinness (2004) found an advantage of reading a *compare/contrast* (=interleaved) text only for readers with a high level of previous knowledge; but no sequence compensated for a low level of previous knowledge.

If considering interleaving in general (applied for visual category learning), there is one article by Sana et al. (2018) showing no moderating impact by the working memory capacity. Hence, further research on interaction patterns between learners' proficiencies and study sequence is needed.

**Cohesion.** Given a particular sequence, clear references and transitions among the information units should be established to increase cohesion, that is, the extent that relations in text are explicit. If a transition between two information units is made explicit – e.g., by a lexical device such as *because* – readers are likely to make the conclusion that the succeeding information unit is the reason for the preceding information unit.

Graesser et al. (2011) differentiate cohesion metrics on the local (sentence-by-sentence) text level. Overlap of words is tapped by *referential cohesion* – for using the same nouns and stems, and avoiding the usage of ambiguous pronouns – and *verb cohesion* for using the same verbs throughout the text. *Logical cohesion* implies the usage of additive and adversative conjunctions. *Temporal cohesion* refers to an unambiguous usage of tense, implying the clarity of chronology by means of lexical devices such as *later* and *before*. *Intentional cohesion* implies the usage of intentional verbs and lexical devices (e.g., *in order to*) signaling actions that are motivated and supposed to serve a purpose. Lexical devices guide readers' expectations to how to integrate the upcoming statement into a coherent mental representation (Gernsbacher, 1990; Zwaan & Radvansky, 1998).

One essential cohesion metric for coherence construction is *causal cohesion* (Louwerse, 2001; Noordman & Vonk, 1997; Sanders & Noordman, 2000). Causal connectives convey information about a relation over and above the additive and temporal connectives. To establish causal cohesion, it is necessary to explicate the causal links among phenomena. Causal links can reflect either consequence-cause (i.e., *objective*) relations or claim-reason (i.e., *subjective*) relations (Canestrelli et al., 2013; Traxler et al., 1997). According to the taxonomy by Sanders et al. (1992) and Louwerse (2001), causal connectives systematically vary in the dimensions, *direction* and *polarity*. Connectives such as *therefore* (and *however*) signal a *forward* direction of the relation, indicating that a cause leads to a consequence (or a reason supports a claim). *Because* (and *although*), in contrast, signals a *backward* direction of the relation, indicating that a consequence can be attributed to a cause (or a claim is supported by a reason). *Positive* connectives such as *therefore* and *because* suggest a congruency with expectations, whereas *negative* connectives such as *however* and *although* convey a violation of expectations (Lagerwerf, 1998). When encountering a causal connective in text, readers retrieve general premises from long-term memory to validate a particular relation (Noordman et al., 1992). If readers lack the particular knowledge, general premises can be concluded and integrated (Cozijn et al., 2011). Studies demonstrated better recall of causally linked sentences than of disconnected sentences (Fletcher & Bloom, 1988; Myers et al., 1987; Trabasso & van den Broek, 1985). Increasing causal cohesion was also shown to improve comprehension (Degand et al., 1999; Degand & Sanders, 2002; Maury &

Teisserenc, 2005; Sanders et al., 2007; van Silfhout et al., 2015; van Silfhout, Evers-Vermeul, Mak et al., 2014; van Silfhout, Evers-Vermeul, & Sanders, 2014).

In addition to the local level, cohesion can be established at the global level by using macrosignals indicating the underlying text structure such as subtopic headers and macropropositions. Particularly the macropropositions at the beginning/end of the respective paragraphs serve the purpose of linking the subtopics among each other and to the superordinate topic (van Dijk & Kintsch, 1983).

What researchers on text comprehension consider *cohesion* is actually treated diversely across studies. Beyond making relations among information units explicit by means of lexical devices and macrostructural signals, providing background information necessary for understanding is also considered a substantial cohesion improvement and is often used. For example, Vidal-Abarca and Sanjose (1998) manipulated the degree of supporting readers in linking the text ideas with their previous knowledge by adding missing premises and concrete images for clarifying abstract information. This improvement resulted in better recall and – combined with the revision regarding the relations in text – in better problem solving. Further, in two studies conducted by McKeown et al. (1992), pupils benefited from reading revised expository texts, which have been enriched with background information, irrespective of whether they received background knowledge in a preceding session or not. It is important to note that the studies by McNamara and Kintsch (1996) and McNamara et al. (1996), which reported moderating effects of previous knowledge on the impact of cohesion, also revised original texts by adding background information. It seems reasonable to assume that low knowledge learners especially rely on additional background information that is necessary to establish a coherent representation, whereas the lack of cohesion devices might be probably easier compensated by making bridging inferences.

Different from the research on text sequencing, the research investigating the impact of cohesion/gaps is far more advanced. However, studies investigating the impact of cohesion usually have one of two limitations in common: Studies that use full-length expository texts usually change numerous text-characteristics at once to manipulate cohesion (i.e., applied research with real world learning materials), but studies that do not confound text-characteristics by manipulating only one text-characteristic (e.g., usage of a lexical device *because*) usually use very short text passages as learning materials at the expense of external validity (i.e., basic research) (cf. van Silfhout, Evers-Vermeul, Mak et al., 2014). The present work addresses this shortcoming by manipulating only one text-characteristic – namely the *causal cohesion* – in full-length expository texts in Experiment 2b.

***Aptitude x Cohesion Interaction.*** As the research has demonstrated, learners with a low level of prior knowledge fail to maintain coherence when reading non-cohesive texts; they thus depend on the guidance provided by linguistic markers and background information (Kamalski et al., 2008; McNamara et al., 1996). Learners with a high level of previous knowledge, in contrast, get engaged in elaborating upon the contents when the relations in text are implicit (Kamalski et al., 2008; Kintsch, 1990; McNamara et al., 1996; McNamara & Kintsch, 1996; Ozuru et al., 2009). The proposed principle behind this pattern is that the low knowledge learners need cohesion to repair their mental model and establish coherence, whereas the high knowledge learners use their coherent mental representation to close the cohesion gaps in text and by doing so integrate the content in their mental representation.

That learners with a high level of previous knowledge take advantage of closing the cohesion gaps puts emphasis on the substantial importance of triggering the construction and integration processes. The *construction-integration model* of text comprehension by Kintsch (1988) and the *desirable difficulties account* by Bjork and Bjork (2014) suggest that increasing the fluency of reading via text cohesion improvements may be the wrong strategy because it may lead to superficial processing. Cohesion gaps in contrast engage learners in effortful processing to overcome the difficulty, which fosters coherence formation and increases the number of retrieval routes (Anderson, 1983; O’Brien & Myers, 1985).

In a nutshell, a fully cohesive text provides the instructions for establishing coherence but lowers the necessity to do so, whereas a poorly written text forces readers to engage in compensatory (repair) processing to infer unstated relations in the text, but does not provide the necessary instructions for establishing these relations. These considerations underscore a mismatch between cohesion as a text-characteristic and coherence as the situation model of text content. From henceforth we will refer to this mismatch as the *cohesion-coherence-mismatch*.

Against the background of the cohesion-coherence-mismatch, we consider the cohesion gaps an effective way to engage readers in relational processing as long as the readers meet the necessary conditions for closing the cohesion gaps. Not only a high level of previous knowledge might be sufficient for closing the gaps. We assume that learners' chances to close a gap independently of their previous knowledge may depend on the type of cohesion indices (e.g., a background information that serves as a premise might be more difficult to conclude from the text than a referent of a pronoun) and whether the text provides adjacently placed basic information for making a bridging inference.

Further studies interested in aptitude x cohesion interaction investigated the interaction of cohesion with the reading skill. Herein, the pattern of findings is inconsistent. On the one hand, there is evidence that cohesion compensates for poor reading skills when reading difficult expository texts (Linderholm et al., 2000). This pattern corresponds to the pattern of that low previous knowledge readers take advantage of reading a fully cohesive text. On the other hand, there are studies suggesting a reversed pattern, namely that reading skill helps readers to deal with additional processing demands imposed by cohesion – especially a higher complexity due to a higher number of explicit relations (O'Reilly & McNamara, 2007; Ozuru et al., 2009; Voss & Silfies, 1996).

The inconsistency of interaction patterns between cohesion and reading skill raises the question about the processing components of the reading skill while reading a non-cohesive text compared to reading a fully cohesive text. In the present work, we pursue an original idea that poor readers might take *no* advantage of reading a non-cohesive text – as has been shown by Linderholm et al. (2000) – because they fail to detect the cohesion gaps in the first place. Such a view is consistent with the investigations by Hannon and Daneman (2001) on component processes of the reading skill. Their investigations revealed the moderating role of reading skill in making elaborative inferences. In other words, whether readers use their previous knowledge to close a cohesion gap might depend on their reading skill. In light of this reasoning, it appears plausible that in the studies by McKeown et al. (1992), students from the 5th grade benefit from text revisions also when provided with substantial background knowledge: Unexperienced readers – even with a high level of background knowledge – may easily overlook the cohesion gaps in the unrevised text and consequently fail to engage in relational processing.

### ***Via Generative Activity***

Overcoming the cohesion-coherence-mismatch means on the one hand to instruct learners on how to establish relations across text ideas and between text ideas and the own previous knowledge. This can be achieved by providing cohesion devices and background information. On the other hand, it means to engage learners in establishing relations. Applying generative learning instructions while reading a cohesive expository text appears to be a promising way of combining the functions of instructing and engaging learners (Ainsworth & Burcham, 2007).

The so called *generation effect* refers to learning advantages of learning strategies engaging learners in active processing and is well known in the literature beginning with the experiments conducted by Slamecka and Graf (1978) (Bertsch et al., 2007; McNamara, 1992). There are several accounts emphasizing the positive impact of generative instructions engaging learners in relational processing on learning such as the *construction-integration* framework by Kintsch (1988), Wittrock's (1989) *generative model of learning*, and Mayer's (2014) *select-organize-integrate* framework. These accounts attribute the positive impact of generative learning instructions to establishing relations across text ideas (i.e., organization processes) and between text ideas and the previous knowledge (i.e., integration processes). All three accounts consider *deep comprehension* an interconnected representation (cf. Zwaan & Radvansky, 1998). Accordingly, generative learning instructions should focus readers' attention to relations such as causes and consequences as well as reasons and claims (cf. McCrudden et al., 2007). In line with this reasoning, research confirmed the link between relating text ideas while reading and learning outcomes (Allen et al., 2015; Kurby et al., 2012; Magliano & Millis, 2003).

There is a high range of generative learning instructions, varying from very simple word generation tasks (deWinstanley & Bjork, 2004) to more complex forms such as generating self-explanations (Wylie & Chi, 2014), elaborative interrogations (Seifert, 1994), and questions (Bugg & McDaniel, 2012). Finally, some generative instructions engage learners in processing contents at the global text level such as re-ordering scrambled texts (McDaniel & Butler, 2011) and generating concept maps (Nesbit & Adesope, 2006).

**Aptitude x Task Interaction.** In general, the pattern of results across studies investigating the moderating impact of learners' proficiencies on generation effect is inconsistent.

There are on the one hand studies that found the generation effect only for high knowledge learners (e.g., Kalyuga et al., 2003), attributing it to the disadvantage of redundant instructional support in control conditions (i.e., the *expertise reversal effect*). Whereas, low knowledge learners do not possess the schemes necessary to solve the task and thus benefit from instructional support (i.e., the *randomness as genesis principle*).[1] Analogous reasoning

[1] McNamara and colleagues argue in a similar way to explain the advantage of cohesion gaps only for high knowledge readers.

prevails in studies investigating the moderating role of task element interactivity, that is, whether the skill level of learners meets the requirements imposed by the task (low element interactivity) or not (high element interactivity) (cf. Sweller, 2010). Herein, studies demonstrated the generation effect only when the element interactivity was low (Chen et al., 2015, 2016). Furthermore, Ionas et al. (2012) showed that problem solving benefited from self-explanation prompts when learners possessed a high level of previous knowledge. It seems reasonable that learners must fulfill necessary conditions to be able to master the generation task and in turn to take learning advantage of it. If learners lack the necessary previous knowledge to generate a solution on their own, they obviously cannot benefit from generating.

On the other hand, there are also studies demonstrating the generation effect especially for poor readers (McDaniel et al., 2002; McDaniel & Butler, 2011). Schindler et al. (2019) demonstrated the generation effect also for leaners with a low level of need for cognition. To explain this pattern of results, McDaniel and Butler (2011) refer to the tetrahedral model of Jenkins (1979), which suggests among others that the learning instruction should compensate for learners‘ deficiencies, particularly for the lack of spontaneous engagement in relational processing. Accordingly, learners who are not spontaneously engaged in relational processing while reading may benefit from a generative instruction that drives their attention to relations across text ideas and to their previous knowledge. That poor readers are not engaged in processes of bridging information and especially integration with the previous knowledge was shown by Hannon and Daneman (2001).

Taken together, there are findings showing that skilled learners benefit from generation but the less skilled learners do not, and vice versa. This apparent inconsistency emphasizes a mismatch*:* Skilled learners are able to master the generation task, but do not require it because they are spontaneously engaged in relational processing anyway. Less skilled learners in contrast require a generative instruction engaging them in relational processing, but they are not capable of accurately performing on the task. From henceforth we will refer to this mismatch as the *ability-requirement-mismatch.*

Accordingly, to make accurate predictions with respect to the impact of a particular generative instruction on learning for learners with various levels of proficiencies (i.e., aptitude-treatment-interaction), it must be considered how well different learners perform on the particular generation task and whether they are spontaneously engaged in relational processing. Over and above, it may be important to determine, which proficiencies enable accurate performance and spontaneous engagement in relational processing respectively. Across Experiments 2a and 2b, we directly addressed the questions to which extent the generation accuracy depends on previous knowledge, to which extent spontaneous engagement in relational processing depends on reading skill, and to which extent generative accuracy and engagement in relational processing affect learning.

**Generative Instruction: The Weak Link.** For a real learning setting, the following problem arises: The generative activity remains the weak link because performing on the task strongly depends on learners' proficiencies and probably their *willingness* to follow the generative instruction in the first place. Generative instructions such as self-explanation prompts may be an effective way in engaging learners in relational processing – however, generative instructions may be in vane if learners do not spontaneously follow them when they learn on their own.

Previous research indicates that many learners may habitually disregard generative instructions. Generation tasks are associated with a higher mental effort (Paas & van Merriënboer, 1993). Many learners are also unaware of the benefits of generation because learners' assessment of their learning progress is affected by their subjective sense of fluency (cf. Bjork et al., 2013). Studies further suggest that less than 50% of college students (Karpicke et al., 2009) and secondary school students (Dirkx et al., 2019) spontaneously use generative learning techniques such as practicing problems (6% and 26.6% for highlighting).

The majority spontaneously uses either passive strategies such as rereading or strategies that rather consolidate the acquired knowledge such as summarizing, flashcards, and especially retrieval. However, generation is superior to retrieval the less learners have understood the contents because generation engages learners in knowledge construction and integration processes (Roelle & Nückles, 2019). Learners thus require a profound instruction on mending their metacognitive misbeliefs about the effectiveness of generation in the first place (McCabe, 2011).

For that reason, we consider a learning tool engaging learners in relational processing that is provided supplementary to the text less preferable to a one the applicability of which depends less on learners' proficiencies and willingness. We assume that the applicability of a learning tool that is incorporated directly into the text is at least less dependent on learners' metacognitive preferences and self-regulation strategies. Directly manipulating the text-characteristics – such as by means of cohesion gaps and interleaving – seems to be thus the most natural way of engaging learners in relational processing.

Furthermore, to handle the limitation regarding the implementation of generative instructions, in Experiments 2a and 2b, we incorporated our generation task into the text. We designed a *cohesion generation* task, which resembles a fill-in-the-blank task. Our generation task was implemented by removing causal connectives from the text, leaving behind explicit gaps. Learners were required to generate causal relations to complete the text.

## Promoting Relational Processing by Highlighting the Cohesion Gaps

To recapitulate, the *cohesion-coherence-mismatch* and the *ability-requirement-mismatch* illustrate the struggles of less skilled learners with learning from expository texts. That is, less skilled learners are not spontaneously engaged in relational processing while reading expository texts and therefore require learning tools compensating for the lack of spontaneous engagement such as cohesion gaps or generative instructions. However, less skilled learners lack also the ability to close cohesion gaps and perform well on a generation task. Over the above, it is reasonable to assume that less skilled learners are poorly motivated to follow supplementary generative instructions since generative instructions lead to disfluent processing, force readers to invest additional mental effort, and probably confront readers with their own knowledge gaps.

In view of that, for the present work, we pursued a different approach than cognitively engaging readers by a generative task as a supplement to reading a fully cohesive text (cf. Ainsworth & Burcham, 2007). We in contrast designed expository texts that were supposed to engage readers in spontaneous relational processing. For such a text design, we relied on the potential impact of cohesion gaps on relational processing (McNamara et al., 1996; McNamara & Kintsch, 1996) and thus removed the essential relations and inferential statements from the expository texts. We further incorporated learning aids into non-cohesive expository texts to compensate for learners' deficits in dealing with the demands imposed by the cohesion gaps.

We presumed two particular demands imposed by a non-cohesive text, that is, closing the cohesion gaps and detecting the cohesion gaps in the first place. Especially the demand to detect the gaps has been neglected in previous research on reading comprehension, presuming rather automatic repair processes across all readers who possess the necessary previous knowledge. However, different from an explicit generation prompt, a cohesion gap is more or less implicit and thus might be invisible to a reader. Studies indicating a *lazy* reader provide evidence in support of our assumption that many readers may overlook the relations among text ideas because they prevailingly focus on immediate context (cf. Cook & Mayer, 1988; Cozijn, 2000; Hyönä et al., 2002; McKoon & Ratcliff, 1992). If a gap remains invisible, the reader would not retrieve the corresponding content from the long-term memory (or attempt to link text ideas) in order to close the gap, resulting in no integration with the previous knowledge. We further assumed that learners differ in the ability to detect the cohesion gaps and this ability might be linked to a cross-domain proficiency, namely the reading skill. The ability to close the gaps in contrast might depend on a domain-specific proficiency, namely the previous knowledge.

To support especially poor readers in dealing with the invisibility of cohesion gaps and those who lack the necessary previous knowledge to close the gaps, we came up with two learning aids, *interleaving of information units* (Experiments 1a and 1b) and *cohesion generation* (Experiments 2a and 2b). These aids were supposed to highlight the cohesion gaps and provide means to close the gaps despite the lack of necessary previous knowledge.

By juxtaposing the comparable information units in text via interleaving, the information necessary for making comparisons was placed adjacently and hence enabled readers to make comparative inferences independently of their previous knowledge. We presumed that the discovery of similarities and differences across various categories would in turn raise readers' awareness of missing inferences regarding the underlying patterns. Because the expository text was enriched with basic information to make conclusions on underlying patterns, we expected less moderating impact of readers' previous knowledge on relational processing.

To implement the cohesion generation, the causal connectives were removed from the expository text, leaving behind explicit gaps, which had to be closed by readers. A reader thus was supposed to take the position of a co-creator of an unfinished text. To close the gaps, readers were required to select an appropriate causal connective from a drop-down list for each explicit gap. To conclude on the right connective, it was not necessary to retrieve contents from previous knowledge, but to bridge adjacent information units.

## Assessment of Relational Processing

Up to this point, we have addressed the independent variables. In this section, we consider our dependent variables, namely the indices of relational processing, learning outcomes and processes.

### *Via the Learning Outcomes*

Text-processing theories differentiate two types of mental representation, the *text-based representation* and the *situation model* (Kintsch, 1988; Kintsch & van Dijk, 1978). The former is a propositional representation embracing the semantic content of the text. Typically, the *retention* measures assess the text-based representation. Based on the *construction-integration model* of Kintsch (1988), Voss and Silfies (1996) consider the text-based representation a function of reading skills. The situation model in contrast refers to the integration of the text content with the previous knowledge and should be considered a function of previous knowledge. It represents the extent that the mental representation of text content is coherent. Typically, the measures capturing situation model are referred to as *comprehension* measures (but also *conceptual understanding* and *inference questions*). Situation model scores show a lower forgetting rate than the text-based representation, which indicates that sustainable learning can be achieved via deep processing over rote memorization (Kintsch et al., 1990). That is why we tested the learning success across Experiments 1b, 2a, and 2b not only immediately, but also with a delay.

Considering learning from expository texts in scientific domains such as physics and biology, situation model construction refers to understanding of scientific explanations, which is a substantial learning objective in science education (Britt et al., 2014). Particularly, a scientific explanation is characterized by a causal chain of causes and consequences. To give an example, the *greenhouse effect* (and *climate change*) – a topic, which we used in Experiments 2a and 2b – is characterized by a causal chain containing such explanatory steps as man's fossil fuel consumption, an increased proportion of carbon dioxide in the atmosphere, the heat trapped in the atmosphere, and increase of global temperatures. To understand the phenomenon of climate change, learners have to establish an interconnected mental representation of various events including common effects (that many factors contribute to one effect), common causes (that one factor has many consequences), and positive (and negative) feedback loops (cf. Goldwater & Gentner, 2015).

Britt et al. (2014) stated that researchers on reading comprehension use a too narrow definition of *comprehension*. The term *situation model* (as the wording alludes) was originally used in research on comprehension of narrative texts. The usage of the same definition for learning with expository texts seems underspecified since the term situation model does not make distinctions among different types of inferences, but serves rather as an umbrella term. To make relevant educational contributions that apply for real-world reading situations, researchers need thus to expand the meaning of *comprehension*.

For example, text comprehension research and research on category induction, which uses primarily visual categories, barely share any dependent measures. There is thus a research gap because category knowledge can also be acquired from a text (cf. Schnotz, 1982, 1984). The distinction between text-based representation and situation model is not very useful when applying to category learning. Category learning inherently requires comparison and distinction among categories. Thus, to bridge these branches of research, in Experiments 1a and 1b, we made an original distinction between two types of inferential questions in the final test – questions on *comparative* and *inductive reasoning* (rather than taking one comprehension measure). Especially the latter type of inference – inductive reasoning – refers to discovering of underlying patterns across natural categories (various types of *whales*) based on covariation data and should be considered a key learning objective in the real world learning (cf. Saffran et al., 2019).

### *Via the Learning Processes*

Numerous studies on learning from expository texts have not only payed attention to learning outcomes but additionally collected process data as indices of relational processing while reading. To name few examples: via think-aloud protocols (Coté et al., 1998; Kraal et al., 2017), self-explanations (Ainsworth & Burcham, 2007; Allen et al., 2015), self-generated questions (Bugg & McDaniel, 2012), eye-tracking patterns (Catrysse et al., 2016; Hyönä et al., 2002; Maier et al., 2018), and reading times on critical passages (Albrecht & O'Brien, 1993; van Silfhout, Evers-Vermeul, Mak et al., 2014). We also recorded learners' spontaneous text-box responses and self-generated questions in Experiment 1b. Experiments 2a and 2b extend this list by decisions on causal connectives and reaction times in a dual task paradigm. In the following, we will point out some important functions of process data in general and its functions across conducted experiments in particular.

Generally, studies differ regarding the subject of their research focus, e.g., inter-individual differences in being engaged in reading processes, the impact of reading processes on learning outcomes, or the effectivity of learning techniques in promoting reading processes that were shown to improve learning. Of particular interest to us are studies that investigated how the text-characteristics affect relational processing while reading. For example, van Silfhout, Evers-Vermeul, Mak et al. (2014) found shorter processing times for cohesive texts with a continuous layout. In the study of Meyer and Freedle (1984), participants organized their text-protocols in accordance with the macro-structure of the text (collection of descriptions, causation, and comparison), but not when the text was presented in the problem/solution manner. Further, Kraal et al. (2017) showed that the text genre affects the way readers process the text: While reading an expository text, learners created fewer text-based and knowledge-based inferences than reading narratives, but created more questions.

Collecting process data is theoretically and educationally high valuable because only the process data provides insights into how text-characteristics affect learning from reading.

For examples, in the study of Ainsworth and Burcham (2007), students who read a maximally cohesive text outperformed their counterparts who read a minimally cohesive text with respect to inference question scores. The analysis on self-explanations revealed, however, that readers of a minimally cohesive text generated a greater number of self-explanations (which included goal-driven explanations) compared to readers of a maximally cohesive text. Authors argued that depending on cohesion-level self-explaining serves different functions, either to compensate for cohesion gaps in a minimally cohesive text or to detect and repair the flaws in the mental representation while reading a fully cohesive text.

Regarding the effects of sequencing on relational processing, the interleaved presentation of belief-consistent and belief-inconsistent texts led to more lookbacks while reading, which in turn promoted the integration of belief-inconsistent information (Maier et al., 2018). Experiment 1b was supposed to extend the insights into the effects of interleaving vs. blocking on relational processing; particularly whether learners would notice cohesion gaps and close them by making high-level inferences. To address these questions, we recorded learners' spontaneous text-box responses and self-generated questions. Herein, we differentiated three cognitive levels of inferences: low-level (repetitions), comparative, and inductive inferences.

In Experiments 2a and 2b, we used advanced assessment tools of relational processing, decisions on causal connectives and reaction times in a dual task paradigm. The former assessment tool (connective choices) reflects learners‘ ability to close the cohesion gaps. Since cohesion gaps in the generation condition were explicitly marked (as opposed to a non-cohesive text without explicit marks), this measure was supposed to reflect the *pure* ability to close the gaps – disentangled from the requirements of detecting the gaps in the first place. We were interested in determining the impact of learners' proficiencies on the ability to close the gaps (=generation accuracy), and the impact of generation accuracy on learning outcomes.

In addition to the process data reflecting the ability to close cohesion gaps, we were interested in collecting process data reflecting whether learners recognize the gaps in the first place (while reading a non-cohesive text). We presumed that lower reaction times in the dual task while reading a non-cohesive as opposed to reading a fully cohesive text would indicate the recognition of cohesion gaps. No difference with respect to the reaction times was expected if learners overlook the cohesion gaps.

We used the latter assessment tool (reaction times in the dual task) in Experiments 2a and 2b to *objectively* measure cognitive load while reading. By using an objective process measure, we tried to overcome a frequent limitation of cognitive load research, that is, the assessment of cognitive load via subjective judgements subsequent to the study phase (Leppink et al., 2013). However, the process measure is disadvantageous due to one reason: The process measure indicates only the degree of cognitive load rather than the type of load. The retrospective measure in contrast tells apart three load types: Load due to the content complexity (intrinsic), its implementation (extraneous), and knowledge construction processes (germane). Therefore, we elicited both the process data and retrospective judgements.

## Overview of the Present Experiments

Altogether, the present work comprises four experimental studies (1a, 1b, 2a, and 2b). The respective articles presenting 1b and 2a are published (open access), 1a is submitted, and 2b is under preparation. The data of all studies is publically available.

Across four experiments, we investigated the effectiveness of two learning tools, interleaving of information units and cohesion generation. We designed our learning tools based on the presumption that cohesion gaps potentially trigger relational processing if readers succeed in both detecting and closing the cohesion gaps. We accordingly used expository texts lacking any relational information (i.e., cohesion), but consisting purely of factual statements in Experiments 1a and 1b, and also manipulated cohesion in Experiment 2b. Both tools – interleaving and cohesion generation – were supposed to highlight the cohesion gaps and support readers in bridging the information in text to close the gaps.

We largely aimed to overcome the common limitations of instructional science, that is, we administered delayed in addition to immediate tests, used real world and complex rather than artificial and simple learning materials, collected process data beyond retrospective judgements and outcomes, examined not only college students in a laboratory but also younger students in authentic educational settings (cf. Dunlosky et al., 2013).

Table 1 provides a brief categorization of experiments along important design criteria. Experiments 1a and 2a were carried out in classroom and Experiments 1b and 2b in laboratory. We collected data from 8th and 9th grade students (Experiment 1a), high school students (Experiment 2a), and college students (Experiments 1b and 2b). We have used real-world scientific texts in biology about the life of whales (Experiments 1a and 1b) and in physics about the climate change and the greenhouse effect (Experiments 2a and 2b). Across experiments, we manipulated the text sequence (Experiments 1a and 1b), generative instruction (Experiments 1b, 2a, and 2b), and the level of cohesion (Experiment 2b). We collected process data such as text-box responses and self-generated questions (Experiment

1b) as well as cognitive load via a dual task and generation accuracy (Experiments 2a and 2b). Considering the learning outcomes, we especially investigated the impact of our learning tools on sustainable learning (Experiment 1b, 2a, and 2b) and high-level inferences such as the discovery of underlying regularities (i.e., *inductive reasoning* in Experiments 1a and 1b) and causal relations (Experiments 2a and 2b). We also considered the potential interactions with learners' proficiencies in Experiments 2a and 2b.

Table 1

*Overview of the Experiments 1a, 1b, 2a, and 2b*

| | 1a | 1b | 2a | 2b |
|---|---|---|---|---|
| Stage of publishing | Submitted | Accepted | Published | Under preparation |
| Open Science (?) | Data available | Data available + open access | Data available + open access | Data available |
| Text-characteristics manipulation | Interleaving vs. blocking +<br>Fixed vs. shuffled | Interleaving vs. blocking | — | Fully cohesive vs. non-cohesive |
| Instruction manipulation | — | Spontaneous vs. prompted self-questioning | Cohesion generation | Cohesion generation |
| Setting | School | Laboratory | School | Laboratory |
| Sample | 8th and 9th grade students (*n* = 194) | College students (*n* = 114) | High school students (*n* = 199) | College students (*n* = 113) |
| Subject (topic) of expository texts | Biology<br>(life of whales) | Biology<br>(life of whales) | Physics<br>(climate change and greenhouse effect) | Physics<br>(climate change and greenhouse effect) |
| Process data | — | text-box responses + self-generated questions | Cognitive load *via dual task* + generation accuracy | Cognitive load *via dual task* + generation accuracy |
| Immediate vs. delayed testing (within-subjects) | Immediate | Immediate + one-week delay | Immediate + two-week delay | Immediate + one-week delay |
| Learning outcomes: low-level inferences | Object representation | Factual details | Text-based representation | Text-based representation |
| Learning outcomes: high-level inferences | Comparative + inductive reasoning | Comparative + inductive reasoning | Situation model (causal links) | Situation model (causal links) |
| Learning proficiencies | — | — | Previous knowledge<br>+ reading skill<br>+ word analogy | Previous knowledge<br>+ reading skill |

In the following, we will briefly outline the particular research questions respectively addressed by the four experiments and how these experiments are theoretically related. Beginning with Experiment 1a, we investigated the impact of a learning tool that is supposed to highlight the cohesion gaps and support learners in closing those gaps – namely *interleaving* – on learning in terms of object representation, comparative and inductive reasoning in unexperienced (less skilled) readers in the 8th and 9th grade. In this experiment, we orthogonally manipulated the text sequence by *category proximity* (interleaving vs. blocking) and *predictability of order* (fixed vs. shuffled). By manipulating the predictability of order, we manipulated the navigating demands while reading the text. The results revealed the superiority of reading an interleaved text with regard to both, comparative (when fixed) and inductive reasoning (i.e., identification of co-occurring patterns). We ascribe the benefit on inductive reasoning to a higher engagement in relational processing while reading an interleaved text, which we tested in the follow-up Experiment 1b.

The follow-up Experiment 1b investigated the underlying mechanisms behind the yielded interleaving effects in Experiment 1a with more advanced readers (college students). We were especially interested in whether the readers of an interleaved text are spontaneously engaged in relational processing. To do so, we combined the text sequence manipulation (interleaving vs. blocking) and the generative instruction manipulation (spontaneous vs. prompted self-questioning). Assuming that reading an interleaved text raises learners' awareness of cohesion gaps, prompted self-questioning should add no advantage when reading an interleaved text, but compensate for the disadvantages of a blocked sequence with respect to relational processing. To determine the extent of relational processing while reading, we recorded learners' text-box entries (spontaneous responses and self-generated questions).

In the next step, we investigated how learners' proficiencies interact with a learning tool that highlights the cohesion gaps, namely the cohesion generation task in comparison to reading a fully cohesive text (Experiment 2a). We also addressed the question to which extent learners' proficiencies attribute to the accuracy of closing the cohesion gaps. We assumed that learners who usually overlook the cohesion gaps would take advantage of such a tool, especially if they succeed to accurately close the gaps.

Experiment 2b incorporates one additional condition, namely reading a non-cohesive text. In Experiment 2b, we directly investigated 1) the impact of cohesion/gaps on relational processing depending on learners' proficiencies and 2) the impact of highlighting the cohesion gaps on relational processing depending on learners' proficiencies. In Experiment 2b, we finally tested our assumption of which particular proficiencies enable readers to recognize and close the cohesion gaps. We expected the reading skill to enable readers to recognize the cohesion gaps in the first place, and the previous knowledge to enable readers to close the cohesion gaps. Consequently, we hypothesized that only skilled readers with a high previous knowledge take advantage of reading a non-cohesive text, whereas poor readers should especially benefit from learning tools highlighting the cohesion gaps.

To recapitulate, in Experiment 1a, we investigated how text-characteristics affect learning. In Experiment 1b, we investigated how text-characteristics interact with a generative instruction. In Experiment 2a, we investigated how learners' proficiencies interact with a generative instruction. In Experiment 2b then, we extended our inquiry to the research question of how learners' proficiencies interact with text-characteristics.

# Experiment 1a

A version of this article is submitted as:

Abel, R., Mai, M., & Hänze, M. (submitted). Text sequence matters for category learning: Interleaving promotes comparisons and discovery of underlying regularities.

## Abstract

We address the question of how information in expository texts should be sequenced to support different goals for learning natural categories. We conducted a 2 x 2 between-subjects factorial experiment with 8th- and 9th-grade students (n = 194). We used a text about different types of whales and manipulated its sequence by *category proximity* (*blocked* = characteristics were grouped by whale type vs. *interleaved* = the various whales were juxtaposed on each characteristic) and *predictability of order* (*invariable* vs. *variable* order of presentation). Consistent with our hypotheses, learners in the interleaved/invariable condition performed better on *comparative reasoning* questions (e.g., concluding, which whale is larger or smaller) than in the other groups. Interleaving also supported *inductive reasoning*—the contrast via interleaving facilitated the discovery of underlying regularities among whale characteristics by enabling learners to link similarities and differences between whales with their co-occurring similarities and differences.

Keywords: interleaving effect; expository text; coherence; category learning; inductive reasoning.

# Text Sequence Matters for Category Learning: Interleaving Promotes Comparisons and Discovery of Underlying Regularities

## Introduction

Expository texts are suitable for learning of artificially and naturally occurring categories (e.g., animal species, scientific and historical phenomena, crime cases, psychological disorders and treatments, etc.). Such expository texts are usually composed of numerous facts, descriptions, and detailed information, but the main ideas are very often not explicitly stated in text. Hence, readers must infer them (van den Broek et al., 2015). Consequently, the rote memorization of factual details can result in a shallow, short-term representation of the learning content. To achieve deep comprehension, readers are required to bridge explicit ideas in the text, that is, to make *bridging inferences* (McNamara et al., 1996). Thus, learners are required to make comparisons between categories while reading the text to be able to sensitively discriminate among related and to generalize over seemingly unrelated categories (cf. Alfieri et al., 2013). For example, when reading a text describing different whales, readers need to consider information concerning their size from multiple sentences to be able to conclude on which whale is larger (i.e., *comparative reasoning*).

Moreover, similarities and differences between artificial and natural categories are not arbitrary but linked to underlying principles (e.g., functionality). Characteristics consequently do not occur in isolation but covary across various categories (e.g., large whales live in small groups of individuals but small whales live in large groups of individuals). To make inferences from covariation data is considered a key learning objective (cf. Saffran et al., 2019). Thus, the induction of how the characteristics co-vary across various categories (i.e., *inductive reasoning*) is an important goal of reading expository texts dealing with different categories. However, research on fostering learning from expository texts has not so far considered inductive reasoning a learning objective.

The requirements allowing the conclusion that the body size negatively correlates with the group size across different whales are manifold. Learners first need to consider information concerning the size and the group size from multiple sentences. They need to conclude on which body sizes and group sizes can be considered as relatively large and relatively small. Finally, to infer that the relatively high values in the size covariate with the relatively low values in the group size across whales, learners need to analyze the pattern of co-occurrence by considering information from all sentences tapping the body size and group size.

Given the importance of inferential processing while reading expository texts, investigating conditions that foster this ability holds educational value. Expository texts describing categories vary on the extent that readers are supported by text characteristics in bridging single information units. Among others, the degree to which learners are supported in inferential processing may depend upon the sequence in which information units (i.e., propositions) are presented (Schnotz, 1984). Given the constraints on learners' working memory capacity, adjacent information units are more likely to be processed simultaneously in active memory (Kintsch, 1988; Wiley & Myers, 2003). Consequently, processing comparisons may depend on the proximity of the to-be-linked information units. Until now, little is known about the impact of sequencing of information units in expository texts on different kinds of inferential processing while reading. The present article addresses this educational inquiry.

An expository text about various categories can be sequenced in multiple ways. *Blocking* categories means a coherent category presentation, that is, the complementary information units are adjacently placed (i.e., massing of complementary information), but comparable information units are spaced among the categories. For example, if a text is about marine mammals, all complementary characteristics of the blue whale, such as the body size, habitat, and group size, are adjacent within the same paragraph (see Table 1). Given the close succession of these object characteristics in the text, they can be simultaneously processed and integrated into a coherent and consistent representation of the object (*object representation,* cf. Schnotz, 1984). Thus, blocking maintains text coherence but impairs contrast. Blocking complies with the canonical way of sequencing expository texts.

In contrast, *interleaving* categories means a comparative presentation, that is, the comparable information units are adjacently placed in the text (i.e., massing of comparable information), but the complementary information is spaced across the paragraphs. For example, when whales are contrasted according to their size within the same paragraph (see Table 1), learners are more likely to compare the whales because they are presented adjacently (Birnbaum et al., 2013). Making comparisons in turn should benefit learning in terms of comparative reasoning (i.e., concluding on which whale is larger or smaller). Thus, interleaving promotes contrast but disrupts coherence. In a nutshell, both distributions of information units, blocking and interleaving, may result in a trade-off between what information is simultaneously processed while reading and which bridging inferences are consequently made.

Table 1

*Four Sequences of Information Units within the Expository Text.*

| | | Variability of order | |
|---|---|---|---|
| | | invariable | variable |
| Category proximity | blocked | 1a 1b 1c 1d 1e 1f | 1a 1b 1c 1d 1e 1f |
| | | 2a 2b 2c 2d 2e 2f | 2b 2d 2e 2a 2f 2c |
| | | 3a 3b 3c 3d 3e 3f | 3c 3e 3d 3f 3b 3a |
| | | 4a 4b 4c 4d 4e 4f | 4d 4a 4f 4b 4c 4e |
| | | 5a 5b 5c 5d 5e 5f | 5e 5f 5a 5c 5d 5b |
| | | 6a 6b 6c 6d 6e 6f | 6f 6c 6b 6e 6a 6d |
| | interleaved | 1a 2a 3a 4a 5a 6a | 1a 2a 3a 4a 5a 6a |
| | | 1b 2b 3b 4b 5b 6b | 2b 4b 5b 1b 6b 3b |
| | | 1c 2c 3c 4c 5c 6c | 3c 5c 4c 6c 2c 1c |
| | | 1d 2d 3d 4d 5d 6d | 4d 1d 6d 2d 3d 5d |
| | | 1e 2e 3e 4e 5e 6e | 5e 6e 1e 3e 4e 2e |
| | | 1f 2f 3f 4f 5f 6f | 6f 3f 2f 5f 1f 4f |

*Note.* Digits 1-6 indicate the six whale types. Characters a-f indicate the six whale characteristics. Each paragraph from the text is represented through a row and contains the description of all characteristics of one whale (blocked) or all whales on one characteristic (interleaved). Only in the invariable sequences, all information units of one characteristic (in the blocked condition) or one whale (in the interleaved condition) are arranged in columns. Information units in a constant position in turn can be easily located and linked by the reader. In contrast, variable sequences have whales differently arranged within paragraphs (in the interleaved condition) or have characteristics differently arranged within paragraphs (in the blocked condition).

### Interleaving of Expository Texts

Numerous studies on category learning have used categories that depended on visual classification, not semantic properties (Jones & Ross, 2011). The research on interleaving of written learning materials, especially expository texts, is therefore very limited. The few studies have reported inconsistent results.

Dobson (2011) used two semantically unrelated expository texts about immunology and reproductive physiology. Participants who read the texts in an interleaved order showed no benefits in recall. This null effect can be attributed to the lack of contrast between the texts despite their proximity. To induce contrast, units of information from different texts must be comparable. In line with this reasoning, Hausman and Kornell (2014) found no advantages from shuffling the flashcards containing textual information from two non-related topics—Indonesian-English word pairs and anatomy ($\eta_p^2 = .01$).

Semantically related expository texts were used in the studies on learning via different text sequences of Schnotz (1982, 1984). The texts about two forms of psychotherapy (psychoanalytical and behavioral) were comparable on five aspects: *theoretical foundation*, *principle of treatment*, *entity of neurosis*, *question of symptoms* and *methodological orientation*. In his terminology, the order of presentation was either *object*- or *aspect-oriented*. The object-oriented text was sequenced in a canonical way. Information that defined the object was presented adjacently. The complementary aspects of psychotherapy could be processed simultaneously in the object-oriented order. Thus, the likelihood of relating them to construct a coherent object representation increased. However, as the comparable aspects from different forms of psychotherapy were remote, the likelihood of contrasting them decreased. In contrast, the comparable aspects from two forms of psychotherapy were presented adjacently in the aspect-oriented text. These comparable aspects could be processed simultaneously. Thus, the two forms of psychotherapy were more likely to be compared. In line with this reasoning, participants who read the text in the aspect-oriented order discriminated the two forms of psychotherapy more reliably (see also Waller & Whalley, 1987). However, given that the complementary aspects of each form of psychotherapy were remote, learners were less likely to construct a coherent object representation and consequently recalled less propositions than their counterparts.

A learning advantage of interleaving over blocking was also shown with textual materials (Zulkiply, 2013; Zulkiply et al., 2012) ($\eta_p^2 = .52$). Cases of psychological disorders were presented either blocked (cases of the same disorder in succession) or interleaved (cases of different disorders in succession). The disorders shared numerous symptoms. Thus, learners were required to discern subtle differences to categorize novel cases without confusion. Helsdingen et al. (2011) provided further evidence for the benefits of randomizing crime case descriptions with and without critical thinking prompts (categorization advantage for cases differing in surface features ($\eta_p^2 = .04$) and structural features ($\eta_p^2 = .50$)).

The few findings on interleaving of textual materials are not consistent: Brunmair and Richter (accepted) found in their meta-analysis no significant effect of interleaving on learning with textual materials: Hedges' $g = 0.21$ [-0.06, 0.47], $p = .119$. As stated above, the effectiveness of interleaving may depend on semantic relatedness of the texts used in the study, that is, whether the texts are comparable or not. Studies that used semantically related textual materials found the interleaving effect (Helsdingen et al., 2011; Schnotz, 1982,1984; Zulkiply, 2013; Zulkiply et al., 2012), but studies using semantically unrelated textual materials found no advantage of interleaving (Dobson, 2011; Hausman & Kornell, 2014). Furthermore, the absence of an interleaving effect can be partially attributed to the dependent measure. Studies that found an interleaving effect measured how well participants could discriminate among categories (Helsdingen et al., 2011; Schnotz, 1984; Zulkiply, 2013; Zulkiply et al., 2012), but studies that found no benefit mostly measured text retention and comprehension (Dobson, 2011; Hausman & Kornell, 2014). A blocked presentation might have supported object representation to a higher degree than interleaving because of the increased chances to simultaneously process object's complementary characteristics (cf. Kintsch & van Dijk, 1978; Schnotz, 1982, 1984). In contrast, interleaving violates the text coherence and thus lowers the chances to simultaneously process object's complementary characteristics. Taking the previous arguments into account, we designed semantically related expository text passages about whales. By doing so, we expected the interleaved text to foster learning from contrast, that is, comparative reasoning, but to hamper coherence construction, that is, object representation.

The question of whether and how text sequence affects inductive reasoning is not only relevant from a practical but also from a theoretical point of view. However, this question has not been investigated to date. Identifying covariations across category characteristics requires learners to make comparisons between the categories and link their complementary characteristics. That is, information from multiple paragraphs must be included, regardless of whether the sequence is blocked or interleaved. As shown in the following, there are, however, good reasons to believe that the text sequence plays a key role for the promotion of inductive reasoning. The research on fostering inductive learning suggests that interleaving supports the pattern recognition while learning naturally occurring categories (Birnbaum et al., 2013; Eglington & Kang, 2017; Higgins & Ross, 2011; Wahlheim et al., 2011) and artificially occurring ones (Authors, under review; Kornell & Bjork, 2008; Lavis & Mitchell, 2006; Metcalfe & Xu, 2016; Yan et al., 2016). Similarly to inductive learning, inductive reasoning also taps pattern recognition across categories. We therefore assumed that interleaving textual materials also supports the discovery of underlying regularities among category characteristics (i.e., inductive reasoning).

## Predictability of Presentation Order

Critical information that is to be linked cannot always be placed adjacently in the text. Thus, their proximity cannot always be guaranteed. Deep understanding of the learning content, however, requires bridging inferences across distant idea units (Hannon & Daneman, 2001; van den Broek et al., 2015). Drawing conclusions on underlying regularities (inductive reasoning) may especially require learners to process multiple information units from different paragraphs.

Written texts provide a high degree of cognitive-processing control for the learner (Schnotz, 2014), thus, no matter whether the information is presented blocked or interleaved, learners can deliberately decide when to follow the linear structure of the printed text and when to deviate from it by self-determining the order of reprocessing the given information. Navigating in text by switching across the distant idea units can be regarded as a strategic behavior to overcome the lack of proximity while learning with demanding text (cf. Hyönä et al., 2002). Navigating might be especially helpful while reading expository texts, because expository texts are inherently complexly structured and learners often lack the previous knowledge to fully comprehend the content (Lorch, 2015; Meyer, 1975). Studies suggest, however, that students struggle to establish links between distant information units and merely focus on the immediate context in which the sentences are embedded (Cook & Mayer, 1988; Coté et al., 1998). Consequently, readers need additional support for improving the accessibility of remote information units. Thus, we faced the educationally relevant question of how to sequence expository texts to enable effortless navigation while reading and increase the opportunities of linking distant idea units.

A predictable order of presentation might enable effortless navigation. In the previous research on contextual interference in the motor-skill acquisition, the predictability was manipulated by a shuffled vs. fixed order of presentation (Lee & Magill, 1983). We thus manipulated the predictability of information units by changing the sequence of information units from paragraph to paragraph (variable sequence) or keeping it fixed (invariable sequence). An invariable sequence of information units across the paragraphs provides readers with the superstructural knowledge (van Dijk & Kintsch, 1983) of where to locate a critical information unit they may be tracking (see Table 1). Knowing this in advance, learners could easily deviate from the linear order by switching across the sentences. Suppose the expository text describes six whales according to six characteristics. In the invariable blocked paragraphs, the size of a whale could be presented first, followed by additional characteristics in a fixed order. In the invariable interleaved paragraphs, the blue whale can be described first, followed by other whales in a fixed order. It is reasonable to assume that an invariable sequence imposes less demands than variable sequences in navigating across the text to locate and link remote information units because an invariable sequence is highly predictable for learners.

The effects of category proximity and variability of sequence on learning might have been confounded in numerous previous studies that compared the effects of interleaving with blocking. In studies on category learning, an interleaved sequence is usually implemented either by shuffling the learning exemplars (e.g., ABC, BCA, CAB, ACB, BAC, CBA) or by holding the presentation order fixed (ABC, ABC, ABC, ABC, ABC, ABC). To control for this potential confound, we manipulated the category proximity (interleaving vs. blocking) and variability of order (variable vs. invariable) independently. We use the term *variable sequence* instead of a *shuffled sequence* to reduce confusion, because *shuffling* is often equated with *interleaving*.

## Present Study

In the present study, we investigated the extent that sequencing, particularly the category proximity and predictability of order, affects learning from expository texts. We designed an expository text describing six types of whales on six properties (see Appendix) and orthogonally manipulated the category proximity (blocking vs. interleaving) and the order of presentation (variable vs. invariable; see Table 1). The factual descriptions in our text were not isolated but inter-connected (e.g., whales can be compared based on factual descriptions; there are patterns of co-occurring characteristics). To achieve deep comprehension (inter-connected knowledge) that goes beyond the surface representation of facts, readers were therefore required to link multiple information units in text. Please note, the presented information in text was sufficient for making comparisons and concluding on co-occurring patterns. Thus, the text itself allowed such inferences to be made. Which is why readers did not have to rely on their previous knowledge.

We manipulated whether the whales were presented blocked or interleaved. In the blocked text, each whale was described according to all its characteristics in one paragraph, whereas in the interleaved condition, whales were juxtaposed on each characteristic in separate paragraphs. Accordingly, the blocked and the interleaved text differed in the distal spacing of complementary and comparable information units. Complementary characteristics of each whale were grouped in the blocked text but spaced apart in the interleaved text. In contrast, comparable information units from different whales were presented in close succession in the interleaved text but scattered across the paragraphs in the blocked text.

Predictability of order was manipulated by presenting information units in either a variable or invariable sequence across the paragraphs, that is, whether the order of presentation was the same or different in each paragraph. A variable sequence makes locating information units more difficult, whereas an invariable sequence allows readers to know in advance the location of information units in a paragraph, which facilitates the linking of remote information.

As dependent measures, we used three subsets of questions: items on object representation, comparative and inductive reasoning. Our hypotheses were based on the common assumption from the research on reading comprehension that the likelihood of making an inference on related information units A and B depends on the proximity of A and B in the text (Schnotz, 1984; Wiley & Myers, 2003). We expected the interleaved text to impair the object representation (H1) as was found in Schnotz's studies (1982, 1984) because complementary characteristics of a whale were grouped only in the blocked sequence. We expected the interleaved text to support comparative reasoning (e.g., deciding which whale is heavier and which is smaller) to a higher degree than blocking (H2), because characteristics of different whales were juxtaposed only in the interleaved sequence. Because interleaving was shown to improve pattern recognition in previous research, we also expected interleaving to promote inductive reasoning (i.e., drawing inferences about how characteristics are related; e.g., a large body size in whales is related to living in smaller groups, or only baleens migrate seasonally) to a higher degree than blocking (H3).

An additional research question explored the extent that variability of order (that is its predictability) affects the impact of category proximity on learning. No study has yet investigated this research question. In general, we assumed that when information units are placed adjacently or when the location of information units are known in advance, learners would be more likely to link them. In contrast, objects and characteristics that are remote or are hard to locate are less likely to be linked. Hence, we expected the variability of order to moderate the impact of category proximity on learning. Given that the invariable sequence is more predictable and thus makes it easier to relate the characteristics of the same whale across the paragraphs in the interleaved condition and to compare the whales in the blocked condition, the particular advantages of blocking (for object representation) and interleaving (for comparative and inductive reasoning) might be less pronounced in the invariable sequence. In contrast, in the variable sequence the advantage of blocking in terms of object representation and the advantage of interleaving in terms of comparative and inductive reasoning might be more pronounced (H4).

## Method

We conducted a 2 x 2 between-subjects factorial experiment in a school setting. We manipulated the sequence of the to-be-read expository text. The first factor was the category proximity: interleaved vs. blocked. The second factor was the variability of order: variable vs. invariable. Students were randomly assigned to one of the four learning conditions. The learning success was assessed immediately after reading by items assessing comparative reasoning, inductive reasoning and object representation.

## Sample

From middle schools in Hessen and Lower Saxony, 194 8th-grade (26.3%) and 9th-grade (73.7%) students participated in our study. The age of students ranged from 13 to 16 (*M* = 14.69 *SD* = .87). The sample consisted of 43.3% females and 54.1% males, and 2.6% did not respond.

Given the wide range of reported effect sizes in previous studies on interleaving of textual materials from .01 to .52 and the novelty of our study design, we oriented toward the effect size of .04. This effect size was found in the study of Helsdingen et al. (2011) for categorization of novel crime cases. Considering all studies that found an interleaving effect, it was the smallest effect size. The power calculation with test power of .8 and a small effect size of .04 showed that the number of participants needed to find an effect at α = .05 is 192.

The study took place during a regular lesson. We received written informed parental consent for all participants. The randomization was done at individual level.

## Learning Material

We developed a comparative expository text about six types of whales: *blue whale*, *fin whale*, *humpback whale*, *killer whale*, *sperm whale* and *narwhal*. Each whale was described according to six properties: *classification (baleen vs. toothed), size, yearly habitat, weight, group's size and behavior*, and *height and angle of the spout*. The text in each condition comprised 661 words and contained seven paragraphs. The introductory paragraph (life of whales) was the same in all conditions. Six additional paragraphs were differently sequenced according to the condition. In the blocked condition, each paragraph included the whole characterization of a particular whale, whereas in the interleaved condition, whales were contrasted with reference to a particular characteristic per paragraph. The sequence within the paragraphs was either invariable across the paragraphs or variable (all sequences are depicted in Table 1). The translation from German into English of the full text in the canonical sequence (blocked and invariable) can be found in the Appendix.

**Testing Material**

Three sets of multiple-choice testing items were developed to measure different facets of the learning success (with different number of questions per set). To correctly answer items on the test that assessed the object representation, learners were simply required to recall the propositions (cf. Schnotz, 1982, 1984). To answer items on the test that assessed comparative reasoning, students needed to make comparisons across the whales with reference to their characteristics. For the test that assessed inductive reasoning, students needed to discover regularities across characteristics.

The 12 items that assessed the object representation required choosing the correct absolute value that represents a given whale characteristic (e.g., *How heavy can a sperm whale become?*). These items capture the text-based representation of single sentences. Note, this measure doesn't necessarily reflect the inter-linked-ness of complementary characteristics. However, the recall of single propositions should be enhanced when single propositions are interlinked within the mental representation (cf. Schnotz, 1982, 1984).

The 8 items on comparative reasoning required students to choose between the whales on a given comparative statement (e.g., *Which whale lives in the smallest groups?*). Answering these items required learners to infer the correct answer by comparing absolute values reported in the text (e.g., group size). For example, killer whales live in groups of up to 70 whales, whereas fin whales mainly live alone or in groups of up to six. Thus, fin whales live in *smaller* groups (comparative inference). Readers should note that the distance across comparable information units strongly differs between the blocked and interleaved sequences. The range of distance is zero (two whales standing next to each other) to four sentences (first and last sentence in a paragraph) between the critical information units in the interleaved conditions because all comparable information units stay within the same paragraph. In contrast, the range of distance is five (when taking from two subsequent paragraphs) to 34 sentences (when taking from the first and last paragraph) in the blocked conditions.

The 9 items on inductive reasoning required learners either to choose the correct complementary characteristic of a whale in response to an item that does not provide the whale type (e.g., *A whale lives in groups of 8. What is the whale's approximate size?*) or to choose the incorrect characteristic (e.g., *This whale belongs to the class of baleens. Which statement is definitely wrong?*). Three distractors and the correct answer of this particular item were: a. *this whale is a solitary animal* b. *its spout emerges heart-shaped* c. *during the summer, it lives in sub-tropical areas* (correct answer) d. *it weighs as much as 100 tons*. Answering these questions required learners to draw conclusions of how whales' characteristics are related in general. For example, a whale's body size correlates negatively with its group's size. The smaller the whale, the bigger its group. However, this regularity is not directly reported in the text. Instead, it must be inferred by readers.

**Procedure**

Students received booklets with the expository text and subsequent questions. The students were told they have 8 min in total to read the text. Underlining and making brief notes was explicitly allowed. If they were finished in less than 8 min, they were required to reread the text, self-paced, in preparation for a test afterwards. Students were told to memorize and comprehend the content because both would be tested. Note that the given time was sufficient to read the text once and to additionally scroll through the text at the reader's own pace, which they were encouraged to do. We were interested in the spontaneous linking of distal information units in the text. Thus, we provided no instructions that they could deviate from the linearity of the text by switching across sentences. After reading, students answered the multiple-choice final test questions in a fixed order (the text was no longer in view). The examination took approximately 25 minutes in total.

# Results

We computed a 2 x 2 x 3 mixed ANOVA with category proximity (blocked vs. interleaved) and variability of order (variable vs. invariable) as between-subject factors, and the type of test (object representation, comparative reasoning, and inductive reasoning) as a within factor. Participants in the interleaved conditions outperformed their counterparts in the blocked conditions, $F(1, 190) = 5.95$, $p = .016$, $\eta_p^2 = .03$ ($M = .52$, $SE = .01$ vs. $M = .48$, $SD = .01$). Participants in the invariably ordered conditions outperformed their counterparts in the variably ordered conditions, $F(1, 190) = 3.99$, $p = .047$, $\eta_p^2 = .02$ ($M = .52$, $SE = .01$ vs. $M = .49$, $SD = .01$). Type of test had no main effect, $F < 1$. There was no two-way interaction of category proximity and variability of order, $F < 1$. Category proximity did not interact with the type of test, $F(2, 380) = 1.81$, $p = .165$, $\eta_p^2 = .01$. However, there was a significant two-way interaction of variability of order with the type of test, $F(2, 380) = 4.10$, $p = .017$, $\eta_p^2 = .02$, as well as a three-way interaction, $F(2, 380) = 3.75$, $p = .024$, $\eta_p^2 = .02$. To disentangle these interactions, we computed a 2 x 2 ANOVA for each type of test (object representation, comparative reasoning, and inductive reasoning) with proximity (blocked vs. interleaved) and variability of order (variable vs. invariable) as between-subject factors.

## Object Representation

Figure 1 shows the pattern of results for questions that required learners to recall propositions from single sentences. Sequencing showed no impact on object representation, neither by category proximity, $F < 1$, nor by variability of order, $F(1, 190) = 1.78$, $p = .183$, $\eta_p^2 = .01$, nor by their interaction, $F < 1$ (rejecting H1).

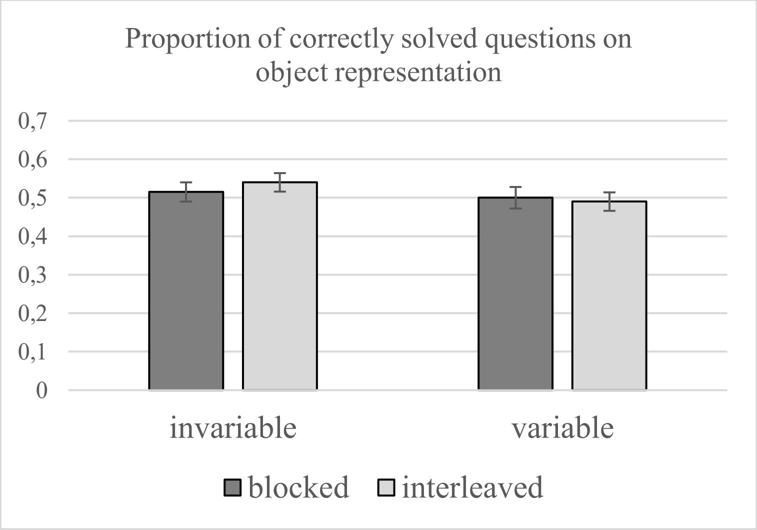


*Figure 1*. Proportion of correctly solved questions on object representation in the final test as a function of category proximity (blocked vs. interleaved) and variability of order (invariable vs. variable), beginning with the canonical sequence of expository texts—blocked (grouping by categories) and invariable. Estimated means and standard errors are depicted.

## Comparative Reasoning

Figure 2 shows the pattern of results for questions that required comparative reasoning. In line with H2, interleaving positively affected comparative reasoning as opposed to blocking, $F(1, 190) = 7.27$, $p = .008$, $\eta_p^2 = .04$. Invariable sequences led to a higher performance as opposed to variable sequences, $F(1, 190) = 10.17$, $p = .002$, $\eta_p^2 = .05$. We further revealed an interaction effect between the category proximity and variability of order, $F(1, 190) = 5.16$, $p = .024$, $\eta_p^2= .03$. The advantage of interleaving was significant in the invariable sequences, $p < .001$, 95% *CI* [.06, .19], *MD*[2] = .12, *SE* = .03, but no significant difference emerged between interleaving and blocking in the variable sequences, $p = .769$, 95% *CI* [-.06, .08], *MD* = .01, *SE* = .04. To check whether the main effect of variability of order would stand, we performed simple comparisons between variable and invariable for both the blocked and the interleaved sequences. The advantage of invariability of order was

[2] *MD* stays for mean difference

significant in the interleaved sequences, $p < .001$, 95% *CI* [.06, .21], *MD* = .14, *SE* = .04, but no significant difference emerged between the blocked sequences, $p = .495$, 95% *CI* [-.04, .09], *MD* = .02, *SE* = .03. Thus, both main effects could be attributed to the interaction effect.

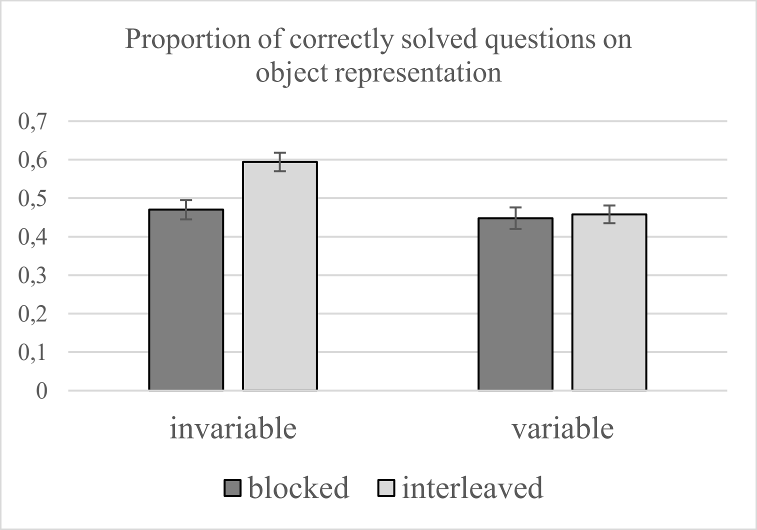


*Figure 2*. Proportion of correctly solved questions on comparative reasoning in the final test as a function of category proximity (blocked vs. interleaved) and variability of order (invariable vs. variable), beginning with the canonical sequence of expository texts—blocked (grouping by categories) and invariable. Estimated means and standard errors are depicted.

## Inductive Reasoning

Figure 3 shows the pattern of results for questions that required inductive reasoning. In line with H3, the interleaving effect could be demonstrated for inductive reasoning, $F(1, 190) = 4.26$, $p = .040$, $\eta_p^2 = .02$. Neither a main effect of the variability of order was found, $F < 1$, nor its interaction with the category proximity, $F(1, 190) = 1.23$, $p = .269$, $\eta_p^2 = .01$.

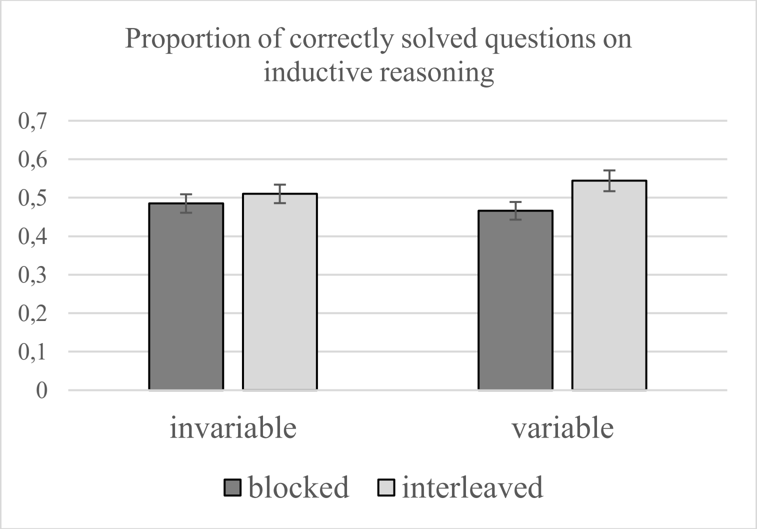


*Figure 3*. Proportion of correctly solved questions on inductive reasoning in the final test as a function of category proximity (blocked vs. interleaved) and variability of order (invariable vs. variable), beginning with the canonical sequence of expository texts—blocked (grouping by categories) and invariable. Estimated means and standard errors are depicted.

## Discussion

The present study expanded our understanding of how to sequence textual learning materials to optimize learning in a classroom setting. Results demonstrated the benefits of interleaving on learning natural categories with expository texts.

Contrary to our expectations based on findings from Schnotz's studies (1982, 1984), blocking did not improve the memory of single propositions, rejecting H1. Thus, the object representation did not depend on the proximity of complementary concepts. This result converges with studies that have found no difference between blocking and interleaving on retention and comprehension (cf. Dobson, 2011; Hausman & Kornell, 2014). However, our conclusion should be treated with caution. We expected interleaving to impair the memory of single propositions to the extent this measure reflects the coherence of object representation. Probably, our measure of the object representation did not tap the inter-linked-ness of characteristics. Thus, it is important to address the question of how sequencing influence the object representation using a more sensitive measure such as the free-recall and open-ended questions in future studies.

Interleaving was superior to blocking in terms of comparative reasoning (supporting H2) and inductive reasoning (supporting H3). Hence, the sequence affected inferential processing on multiple information units as opposed to simple recall. In the following discussion, we address the question of why interleaving augmented comparative *and* inductive reasoning.

**Contrast Accounts for Comparative Reasoning**

According to the insights from the research on reading comprehension, proximately placed information units are likely to be processed simultaneously (Schnotz, 1984; Wiley & Myers, 2003), resulting in more comparisons while reading an interleaved text. More specifically, the *discriminative-contrast-hypothesis* ascribes the benefits in category learning to the juxtaposition of different categories (Birnbaum et al., 2013; Kang & Pashler, 2012; Mitchell et al., 2008; Zulkiply & Burt, 2013). According to the discriminative contrast hypothesis, learners are more likely to detect subtle differences when different categories are juxtaposed. The advantage of interleaving over blocking increases with the material's inherent category similarity. The greater the difficulty to discriminate the categories, the greater the importance of a between-category comparison compared to the stand-alone categorization (Carvalho & Goldstone, 2014a, 2014b, 2015a, 2015b, 2017; Goldstone, 1996; Rohrer, 2012; Zulkiply & Burt, 2013). In line with this reasoning, placing comparable characteristics of different whales in the same paragraph increased the probability of comparing and contrasting whales. In contrast, comparable characteristics were distributed across six paragraphs in the blocked sequence. Making comparative inferences in this condition might have imposed higher demands on learners' working memory and strategic behavior, which in turn could have made inferencing less likely.

However, simple comparisons revealed the superiority of interleaving over blocking for comparative reasoning only when information units were invariably sequenced. Hence, learners could only make considerable comparative inferences when the comparable information units were both proximate and predictable. Our explanation for this finding assumes that a predictable order allows a better organization of the information and then serves as a good retrieval cue in the final test.

The finding that interleaving outcomes varied as a function of whether the sequence was variable or invariable emphasizes the potential of reconsidering the results from previous studies on category learning via interleaving. It might be particularly interesting to assess the effect size of the interleaving effect for shuffled categories (randomized order) as opposed to interleaved categories in a fixed order.

## Contrast Accounts for Inductive Reasoning

The superior inductive reasoning in the interleaved condition may be primarily grounded in the contrast between categories. The specific cognitive requirements of discovering regularities across characteristics should thus be considered. For example, to conclude the correct relation between body weight and group size, readers must relate *lower* body weights to *larger* group sizes and *higher* body weights to *smaller* group sizes. In other words, comparative inferences are more important than remembering single facts. Memorization of a single fact would result in recalling that narwhales weigh as much as 1.5 tons. In contrast, a comparative inference would result in the conclusion that narwhales' weight of 1.5 tons is a relatively low weight in comparison to especially baleens. A blocked sequence merely invited readers to link absolute characteristics of each whale individually (e.g., narwhales weigh as much as 1.5 tons and typically live in groups of 20) but not to identify the underlying regularities (e.g., the relative low weight of 1.5 tons is associated with the relatively large group size of 20). In sum, the comparative inferences may have been essential for the discovery of regularities. Thus, interleaving may have supported the discovery of regularities by highlighting the contrast.

It is reasonable to assume that the positive impact of interleaving on inductive reasoning generalizes beyond the scope of natural category learning but on artificially occurring categories. Diverse sets of categories, natural or artificial, might differ with regard to the particular characteristics, the pattern of their co-occurrence, and the underlying principles. However, so far we don't see any reason why the impact on inductive reasoning should be restricted to natural categories (e.g., whales) as long as the set of learning exemplars logically allows to make generalizations. Learners might, however, differ with regard to the previous knowledge of the principles underlying the co-occurrence of characteristics across particular categories. An important method procedure is thus to control for prior knowledge about the underlying regularities across category characteristics (and their underlying principles), that is, by assessing prior knowledge, varying the pre-instruction, or using artificial categories.

**Future Directions**

Reading should be seen as a hierarchical rather than a linear process (van Dijk & Kintsch, 1983). The research on reading comprehension, however, has not yet devoted sufficient attention toward the instructional support of self-regulated processes when linking distant information in text. A promising avenue for future research is to address the question of how to distribute information in expository texts to increase the accessibility of distal information units.

When reading an invariably sequenced text, learners can rely on fixed positions of critical information units across the paragraphs, which is why we expected the invariable sequence, as opposed to the variable sequence, to impose less demands on learners' navigational behavior in locating distant information units in text. As we could show, an invariable sequence appeared to be an appropriate method to support comparative reasoning when critical information units were juxtaposed *within* a paragraph (interleaved). However, an invariable order provided no additional support for linking distant information units that were spaced across the paragraphs, rejecting H4. Comparative reasoning was especially not enhanced after reading a blocked text. Thus, predictability did not compensate for a lack of category proximity as we had expected. Apparently, learners did not spontaneously use their superstructural knowledge of locating critical information and made no attempts to skip up and down between paragraphs. Converging evidence has shown that most readers make no attempts to navigate in text. For example, Hyönä et al. (2002) investigated different text reprocessing strategies in adult readers via their eye fixation patterns. They showed that many readers rigidly follow the linear text structure without looking back. Only a small number of readers demonstrated selectivity in their navigational behavior. In follow-up studies, variability of order could be manipulated in combination with prompts. Learners' attempts in linking proximate and distant information units could be then recorded via think-aloud protocols (Coté et al., 1998) and eye-tracking (Catrysse et al., 2016). The combination of an invariable sequence supporting readers in locating critical information units using prompts to encourage them in actively navigating through the text may increase the flexibility in navigational behavior and become observable in verbal protocols or eye fixation patterns. We expect the invariably sequenced text to produce greater effects when it is complemented by established techniques for attention guiding such as signaling (Naumann et al., 2007).

## Conclusions

The present study investigated the impact of sequencing in printed expository texts on learning educationally relevant content in school. The canonical sequence of expository texts—blocked (grouping by categories) and invariable—did not support the memory of single propositions. In contrast, interleaving (grouping by aspects) was shown to be the superior study sequence. Interleaving not only accentuated the contrast between categories but also supported learners in identifying underlying regularities across category characteristics.

Accordingly, we recommend designers of expository learning texts to implement interleaved sequencing instead of continuing to use canonical sequencing when the learning objective requires learners to discriminate naturally or artificially occurring categories and identify their underlying regularities. Future research may reveal whether the superiority of interleaving over blocking depends on the learning objective.

References

Authors (under review).

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392–402. https://doi.org/10.3758/s13421-012-0272-7

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

Carvalho, P. F., & Goldstone, R. L. (2014a). Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology*, *5*, 1–11. https://doi.org/10.3389/fpsyg.2014.00936

Carvalho, P. F., & Goldstone, R. L. (2014b). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481–495. https://doi.org/10.3758/s13421-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2015a). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, *22*(1), 281–288. https://doi.org/10.3758/s13423-014-0676-4

Carvalho, P. F., & Goldstone, R. L. (2015b). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*, 1–12. https://doi.org/10.3389/fpsyg.2015.00505

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of*

*Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1699–1719. https://doi.org/10.1037/xlm0000406

Catrysse, L., Gijbels, D., Donche, V., Maeyer, S. de, van den Bossche, P., & Gommers, L. (2016). Mapping processing strategies in learning from expository text: An exploratory eye tracking study followed by a cued recall. *Frontline Learning Research*, *4*(1), 1–16. https://doi.org/10.14786/flr.v4i1.192

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology*, *80*(4), 448–456.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*(1), 1–53. https://doi.org/10.1080/01638539809545019

Dobson, J. L. (2011). Effect of selected "desirable difficulty" learning strategies on the retention of physiology information. *Advances in Physiology Education*, *35*(4), 378–383. https://doi.org/10.1152/advan.00039.2011

Eglington, L. G., & Kang, S. H. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition*, *6*(4), 475–485. https://doi.org/10.1016/j.jarmac.2017.07.005

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*(5), 608–628. https://doi.org/10.3758/BF03201087

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*(1), 103–128. https://doi.org/10.1037/0022-0663.93.1.103

Hausman, H., & Kornell, N. (2014). Mixing topics while studying does not enhance learning. *Journal of Applied Research in Memory and Cognition*, *3*(3), 153–160. https://doi.org/10.1016/j.jarmac.2014.03.003

Helsdingen, A., van Gog, T., & van Merriënboer, J. (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgment task. *Journal of Educational Psychology*, *103*(2), 383–398. https://doi.org/10.1037/a0022370

Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for what. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*, 1388–1393.

Hyönä, J., Lorch, R. F., Jr., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*(1), 44–55. https://doi.org/10.1037//0022-0663.94.1.44

Jones, E. L., & Ross, B. H. (2011). Classification versus inference learning contrasted with real-world categories. *Memory & Cognition*, *39*(5), 764–777. https://doi.org/10.3758/s13421-010-0058-8

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103.

Kintsch, W. (1988). The role of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*(2), 163–182.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037//0033-295X.85.5.363

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*, 585–592.

Lavis, Y., & Mitchell, C. (2006). Effects of preexposure on stimulus discrimination: An investigation of the mechanisms responsible for human perceptual learning. *Quarterly Journal of Experimental Psychology*, *59*(12), 2083–2101. https://doi.org/10.1080/17470210600705198

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 730–746. https://doi.org/10.1037/0278-7393.9.4.730

Lorch, R. F., Jr. (2015). What about expository text? In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 348–361). Cambridge University Press.

McNamara, D. S., Kintsch, E., Butler Songer, N., & 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.

Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(6), 978–984. https://doi.org/10.1037/xlm0000216

Meyer, B. J. F. (1975). *The organization of prose and its effect on memory*. North-Holland.

Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 237–242.

Naumann, J., Richter, T., Flender, J., Christmann, U., & Groeben, N. (2007). Signaling in expository hypertexts compensates for deficits in reading skill. *Journal of Educational Psychology*, *99*(4), 791–807. https://doi.org/10.1037/0022-0663.99.4.791

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

Saffran, A., Barchfeld, P., Alibali, M. W., Reiss, K., & Sodian Beate (2019). Children's interpretations of covariation data: Explanations reveal understanding of relevant comparisons. *Learning and Instruction*, *59*, 13–20.

Schnotz, W. (1982). How do different readers learn with different text organizations. In A. Flammer & W. Kintsch (Eds.), *Discourse processing* (pp. 87-97). North-Holland.

Schnotz, W. (1984). Comparative instructional text organization. In H. Mandel, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 53-81). Erlbaum.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (2nd ed., pp. 72–103). Cambridge University Press.

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. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 94–121). Cambridge University Press.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*, 750–763.

Waller, R., & Whalley, P. (1987). Graphically organised prose. In E. de Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction. European research in an international context: Vol. 1* (pp. 369–381). Leuven University Press and Pergamon Press.

Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, *36*(2), 109–129. Retrieved from DOI: 10.1207/S15326950DP3602_2

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit.

*Journal of Experimental Psychology: General*, *145*(7), 918–933. https://doi.org/10.1037/xge0000177

Zulkiply, N. (2013). Effect of interleaving exemplars presented as auditory text on long-term retention in inductive learning. *Procedia - Social and Behavioral Sciences*, *97*, 238–245. https://doi.org/10.1016/j.sbspro.2013.10.228

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*(1), 16–27. https://doi.org/10.3758/s13421-012-0238-9

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215–221.

## Appendix

The full expository text used in the present study in the blocked and invariable sequence.

Translated from German into English.

Whales belong to the genus of marine mammals. These mammals are adapted to life in water, for example, like seals. All whales are classified into two major groups: toothed whales, which have teeth and mainly feed on fish, and baleen whales, which use baleen plates to filter krill out of water as their nourishment. Another characteristic of whales is the ejection of a spout, which means the forcefully expelled air coming through one or two blowholes when surfacing.

A blue whale attains a maximum body length of 33 meters. A blue whale belongs to the baleen whales. In summer, blue whales live in polar waters in the northern and southern hemisphere and migrate south to subtropical areas in winter. Adult blue whales weigh as much as 200 tons. Blue whales are solitary animals mainly living and traveling alone because they are too large to have natural predators; sometimes, they travel in small groups to protect their whale calves. The spout of a blue whale reaches nine meters in height and is ejected vertically.

A killer whale attains a maximum body length of 10 meters. A killer whale belongs to the toothed whales. Killer whales are widely distributed but are most common in the northern polar and coastal waters. However, because of global warming, killer whales have increasingly migrated to the northern regions. Adult killer whales weigh as much as 6.5 tons. Killer whales are social animals mainly living in groups of up to 70; because the littoral areas involve a higher chance of hunting success, these groups do not live near the coast. The spout of a killer whale reaches one to two meters in height but is not always visible.

A fin whale attains a maximum body length of 24 meters. A fin whale belongs to the baleen whales. In summer, fin whales live in polar waters for ingestion and migrate to subtropical areas in winter for mating and calving, always avoiding littoral areas; because of the reversed seasons in the northern and southern hemisphere, northern and southern populations of fin whales do not meet. Adult fin whales weigh as much as 70 tons. Fin whales are solitary animals mainly living alone or in groups of up to six groups; in large groups, it is more difficult to find huge amounts of krill, which is required as nourishment every day. The spout of a fin whale reaches six meters in height and can be recognized on a vertical narrow shape.

A sperm whale attains a maximum body length of 20 meters. A sperm whale belongs to the toothed whales. The sperm whale is widely distributed but is mainly found in deep waters; females and calves remain in tropical or subtropical waters throughout the year. Adult sperm whales weigh as much as 50 tons. Male sperm whales roam solo or in small groups, whereas females and their calves build their own groups; mothers form circular defensive positions to protect the calves against natural predators. The spout of a sperm whale reaches two meters in height and emerges in a 45-degree angle.

A humpback whale attains a maximum body length of 15 meters. A humpback whale belongs to the baleen whales. In summer, humpback whales live in polar waters for nourishment and migrate to subtropical regions in winter where they live off their fat stores; since they only find nourishment at a depth of under 50 meters, the animals stay near the coast. Adult humpback whales weigh as much as 30 tons. Humpback whales typically live alone or in groups up to nine animals in which the males protect the females and calves. The spout of a humpback whale reaches three meters in height and emerges heart-shaped.

A narwhale attains a maximum body length of five meters. A narwhale belongs to the toothed whales. Narwhales live in the Arctic Ocean and stay close to the ice pack, breaking the ice with their foreheads to create breathing holes. Adult narwhales weigh as much as 1.5 tons. Narwhales typically live in family associations of up to 20, although large herds of up to 1,000 members are formed when travelling for reasons which remain unclear. The spout of a narwhale is small in height and often inconspicuous.

## Experiment 1b

A version of this article is accepted as:

Abel, R., Niedling, L. M., & Hänze, M. (accepted). Spontaneous inferential processing while reading interleaved expository texts enables learners to discover the underlying regularities. *Applied Cognitive Psychology*. DOI: 10.1002/acp.3761

# Abstract

Recent studies on text sequencing found learning advantages of interleaving over blocking in terms of high-level inferences. We conducted a 2 x 2 x 2 mixed factorial experiment with college students (*n* = 117) by manipulating text sequence (interleaved vs. blocked) and self-questioning activity while reading (spontaneous vs. prompted) between subjects and testing delay (immediately vs. one-week delay) within subjects. Results revealed that students are spontaneously engaged in self-questioning and inferential processing while reading an interleaved text. Students who were spontaneously engaged while reading an interleaved text outperformed their counterparts in all other conditions in the immediate and delayed test on comparative reasoning, inductive reasoning, and memorization of factual details. The learning advantages were mediated by inductive inferences made while reading an interleaved text. Results support the discriminative contrast view that readers are encouraged to discover the underlying regularities when differences and similarities among categories are accentuated by their juxtaposition.

*Keywords:* interleaving effect; text comprehension; inductive learning; inductive reasoning; question generation

# Spontaneous inferential processing while reading interleaved expository texts enables learners to discover the underlying regularities

The sequence of presentation has a strong impact on how learning content is encoded, organized, and integrated. According to the discriminative contrast hypothesis, juxtaposition of examples of different categories via *interleaving* lead learners to make comparisons and identify category boundaries (Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012). Beginning with studies on categorization of paintings, a growing body of research has found evidence for the learning advantages of an interleaved study sequence (i.e., categories are presented mixed – ABC, ABC, ABC) over blocking (i.e., categories are presented uninterrupted – AAA, BBB, CCC) (Brunmair & Richter, 2019; Kornell & Bjork, 2008). Participants studying in an interleaved manner are more likely to correctly categorize the category examples in the final test. The interleaving effect has been well replicated with categorization based on visual characteristics such as with artificial (Abel, Brunmair, & Weissgerber, under review; Mitchell, Kadib, Nash, Lavis, & Hall, 2008) and natural categories (Eglington & Kang, 2017; Higgins & Ross, 2011; Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013; Wahlheim, Dunlosky, & Jacoby, 2011).

Research on interleaving is not limited to its impact on classification based on visual characteristics. Studies on interleaving expository texts have explored its impact also on classification based on semantic characteristics. For example, in Schnotz's experiments on reading texts about two forms of psychotherapy presented in different sequences (1982, 1984), students who read a text in which psychotherapies were juxtaposed on aspects were more likely to correctly discriminate the forms of psychotherapy than students who read the text in a canonical sequence. A growing body of evidence has shown that reading interleaved expository texts benefits category learning to a higher extent than reading texts sequenced in a canonical way (Helsdingen, van Gog, & van Merriënboer, 2011). For example, juxtaposing cases of psychological disorders via interleaving increased the likelihood of making comparisons during the study phase and consequently of correctly categorizing new cases during the immediate (Zulkiply, McLean, Burt, & Bath, 2012) and delayed final test (Zulkiply, 2013). Over and above, interleaved presentation of belief-consistent and belief-inconsistent texts fosters the processing and comprehension of belief-inconsistent information (Maier, Richter, & Britt, 2018).

Learning from expository texts, however, encompasses a high range of learning goals and demands. In addition to discriminating among categories, learners must establish a coherent mental representation because of a high inherent complexity of semantic relations (Britt, Richter, & Rouet, 2014; Zwaan & Radvansky, 1998). An expository text conveys principles, general patterns, and regularities. Moreover, learners are faced with demands that can be attributed to the expository text as an information medium, especially when essential ideas are not explicitly stated in the text and readers must infer its meaning (van den Broek, Beker, & Oudega, 2015). For example, to decode implicit relations among proximate sentences, readers are often required to close cohesion gaps by accessing and integrating information with previous knowledge (Hannon & Daneman, 2001; Kintsch, 1988) (i.e., making elaborative inferences). To link remotely placed idea units, readers are required to navigate among sentences and make bridging inferences (McNamara, Kintsch, Butler Songer, & Kintsch, 1996). However, cohesion gaps are less likely to be closed when critical sentences are spaced (Wiley & Myers, 2003). Studies indicate that most learners follow the text linearly, make no attempts to look back while reading (Hyönä, Lorch, & Kaakinen, 2002), make no use of their superstructural knowledge of the text (Abel, Mai, & Hänze, submitted), and merely focus on the immediate context (Cook & Mayer, 1988; Coté, Goldman, & Saul, 1998). Hence, readers usually fail to establish links between distant pieces of information.

In light of these learning goals and difficulties, Abel, Mai, and Hänze (submitted) investigated the impact of interleaving textual materials on a wide range of learning outcomes. In the study, the sequence of an expository text about the life of whales was manipulated. In the canonical text sequence (i.e., blocked), each whale was described on its characteristics in a separate paragraph. In the interleaved text, whales were contrasted on each characteristic in separate paragraphs. Accordingly, the blocked and interleaved text differed with reference to which information units were placed adjacently and which were placed apart. The study has revealed a learning advantage of interleaving for comparative reasoning (i.e., which whale is heavier, smaller). We explain this effect by referring to the constraints on learners' working memory capacity: adjacently placed information is more likely to be processed simultaneously (Kintsch, 1988; Wiley & Myers, 2003). A juxtaposed text structure allows readers to directly compare the categories (i.e., making local bridging—comparative—inferences). A blocked structure in contrast imposes higher demands on working memory because the comparable information units are spaced apart throughout the text (i.e., making global bridging inferences).

Over and above, the study has also revealed learning advantages of interleaving in terms of inductive reasoning. Participants who read an interleaved text were more likely to identify the underlying regularities among the whales' characteristics, e.g., that *lower* body weights are associated with *bigger* group sizes, and *larger* body weights with *smaller* group sizes (for categorization based on coherent covariations of properties across samples, see Rogers & McClelland, 2008).

Different from the interleaving effect in terms of comparative reasoning, the positive impact on inductive reasoning (i.e., identification of co-occurring patterns) cannot be explained by merely referring to the working memory constraints. Making inductive inferences requires learners to make global bridging inferences across multiple paragraphs even in the interleaved condition (e.g., based on solely one paragraph, learners are not able to identify that a large body size goes along with a small group size). This finding thus appears to be in need of explanation against the background of studies on text comprehension indicating a *lazy* reader.[3]

The main purpose of the present study was hence to investigate the underlying mechanisms behind the positive impact of interleaving on inductive reasoning. We attribute this finding to inferential processes entailed by the discriminative contrast. According to our interpretation, the discriminative contrast enables readers to make comparisons between the objects and in turn raises readers' awareness of the factors that contribute to the salient differences and similarities in objects. According to this explanation, learners spontaneously apply self-questioning and look for covarying differences in characteristics across objects (i.e., underlying regularities) when reading an interleaved text.

To test this assumption, it is essential to trace readers' spontaneous attempts to explain the differences in appearance and behavior in whales that lead to the discovery of how these differences covary. Thus, in the present study, we extended the previous research by addressing two additional research questions about inferential processes when reading an interleaved text. We investigated (1) whether readers of an interleaved text—as opposed to a blocked text—spontaneously question and generate explanations for the given differences and similarities between the whales and subsequently make conclusions on how different characteristics are related (inductive inferences) and (2) whether inductive inferences generated while reading mediate the learning advantages of interleaving. We recorded the process data on inferential processing while reading. We differentiated between three

[3] The pattern of results yielded by Abel, Mai, and Hänze (submitted) suggests that the interleaving effect in terms of inductive reasoning might be attributed to an increased cognitive engagement. Among others, Abel and colleagues addressed the question whether superstructural support in making global bridging inferences will help learners to overcome particular weaknesses of a blocked sequence. They manipulated—in addition to the sequence (blocked vs. interleaved)—the superstructural support via the predictability of text order (predictable by a fixed order vs. unpredictable by a shuffled order). A fixed order supported readers because it allowed to effortlessly locate critical information units such as a certain characteristic (e.g., *size* when blocking) throughout the printed text. However, readers of a blocked text performed equally in the final test, irrespective of the superstructural support. The lack of benefit by the superstructural support in the blocked conditions is compatible with the view of a *lazy* reader: readers of a blocked text were supported in making global inferences but not cognitively engaged. In contrast, readers of an interleaved text made use of superstructural support provided by a predictable text order.

cognitive levels of inferences based on reinstatements of single sentences (low-level inferences), comparisons between whales (comparative inferences), and linkages between complementary characteristics (inductive inferences on regularities). *Inductive inferences* refer to the discovery of how characteristics of whales are related in general (e.g., only baleens seasonally migrate; relatively small whales live in relatively large groups).

To further scrutinize our main research question, that is, whether learners spontaneously engage in self-questioning while reading an interleaved text, we manipulated whether readers received an instruction to generate questions and answers on each paragraph (i.e. prompted self-questioning) or no instruction (i.e., spontaneous self-questioning).

The advantage of prompted self-questioning (often referred to as *question generation*) over passive studying on learning from expository texts has been shown in numerous studies (Bugg & McDaniel, 2012; Foos, Mora, & Tkacz, 1994; Koch & Eckstein, 1991; van Blerkom, van Blerkom, & Bertsch, 2006; Weinstein, McDermott, & Roediger, 2010). The relative learning advantage of prompted self-questioning has also been demonstrated in comparison with techniques that engage learners in active processing. For example, in the study of Koch and Eckstein (1991), participants who generated questions outperformed their counterparts who were engaged in answering adjunct questions while reading. In the study of van Blerkom et al. (2006), prompted self-questioning yielded higher learning scores than highlighting while reading. Prompted self-questioning was not less effective than outlining (Foos et al., 1994), notetaking (van Blerkom et al., 2006) and testing (Weinstein et al., 2010). Koch and Eckstein (1991) explained that one of the reasons for the advantage of self-questioning on learning from expository texts is primarily the stimulation of curiosity generated from open answers on self-generated questions.

To explore how prompted self-questioning affects inferential processing while reading a text (interleaved vs. blocked), we compared the text box entries of participants in the prompted self-questioning conditions with those in the spontaneous activity conditions.

Assuming that learners spontaneously apply self-questioning and discover regularities while reading an interleaved text, instructional prompting to generate questions should add no learning advantage to spontaneous activity. In contrast, assuming that a blocked text fails to sufficiently stimulate inferential processes, learners are more likely to take advantage from instructional support of prompted self-questioning while reading.

## Present Study

We investigated the following three research questions: (1) whether the main finding of Abel, Mai, and Hänze (submitted), that is the immediate learning advantage of interleaving in terms of inductive reasoning, can be replicated with more advanced readers (and whether this advantage holds with a higher retention interval of one week) (2) the extent that readers of an interleaved text spontaneously apply self-questioning and generate inferences while reading, and (3) whether the learning advantage of interleaving is mediated by inferential processing while reading. We consider the second research question the main one. To address this question, we investigated the extent that readers spontaneously (=without self-questioning prompts) generate inductive inferences while reading an interleaved (vs. blocked) text and the impact of self-questioning prompts on learning processes and outcomes. We particularly expected self-questioning prompts to be redundant while reading an interleaved text.

Participants read an expository text describing six whales with regard to six characteristics (the blocked text is available on https://osf.io/mr6a4/ and the interleaved text is available on https://osf.io/gzu8c/). We orthogonally manipulated the sequence of the text (blocked vs. interleaved, see Appendix for comparison) and instruction for self-questioning (no prompting by which learners are spontaneously engaged in self-questioning while reading vs. prompting to generate questions). In the blocked text, characteristics were grouped by whales. In the interleaved text, the whales were juxtaposed on their characteristics. In the prompted self-questioning conditions, readers were required to generate a question at each paragraph and answer it. Prompting self-questioning was intended to stimulate learners to actively reprocess and rethink the learning content. In contrast, participants received no specific instructions in spontaneous activity conditions. Learners were merely told to type their thoughts about the text. We collected participants' responses (process data) to analyze the extent that they would make low-level inferences on factual details and comparative and inductive inferences. We also coded inferences that required the integration of world knowledge as elaborative inferences. Learning performance (outcome data) was assessed immediately after the study phase and after a one-week delay. We used three subsets of questions that elicited memorization of factual details, comparative reasoning, and inductive reasoning. The classification of items on learning performance (outcomes) corresponded to the three cognitive levels of inferences made while reading (processes), factual, comparative and inductive.

We expected to find differences in the process data between the interleaved and blocked conditions. Readers in the interleaved conditions should generate more comparative inferences because characteristics of different whales are juxtaposed only in the interleaved sequence. The discriminative contrast in the interleaved conditions should encourage readers to ask themselves about the differences between whales and appraise whether the differences co-occur. Thus, we expected the readers in the interleaved conditions to make more responses that reflect how characteristics are linked (inductive inferences) (*Learning Processes Hypothesis*). In contrast, the blocked sequence does not provide sufficient opportunities to compare between the whales. Thus, readers in the blocked conditions should produce more low-level inferences reflecting their attentional focus at factual details.

Moreover, we assumed that the discriminative contrast via interleaving would automatically trigger self-questioning in readers, and prompting self-questioning via the question generation instruction would be redundant while reading an interleaved text. As a result, the frequency of inductive inferences in the interleaved/spontaneous activity group should not be lower than in the interleaved/prompted self-questioning group. In contrast, we assumed the blocked sequence would not engage readers in spontaneous inferential processes. Hence, prompted self-questioning while reading a blocked text should compensate for the lack of spontaneous inductive processing by increasing the level of active processing. This effect might be indicated by a higher number of comparative and inductive inferences in the blocked/prompted self-questioning condition compared to the blocked/spontaneous condition. Accordingly, the difference in frequency of inductive inferences between interleaving and blocking should be clearly observable in spontaneous activity conditions and less observable in prompted self-questioning conditions (*Moderation Hypothesis for Learning Processes*).

We expected to replicate the results from the previous research conducted by Abel, Mai, and Hänze (submitted) with more advanced readers on the immediate final test. That is, comparative and inductive reasoning should be greater for interleaving than blocking in spontaneous activity conditions (*Learning Outcomes Hypothesis*). Furthermore, by incorporating additional performance assessment with one-week delay, we explored whether the expected immediate advantage of interleaving would hold in a long run. The superiority of interleaving over blocking should be more pronounced in spontaneous self-questioning conditions than in prompted self-questioning conditions for the same reason as previously stated (*Moderation Hypothesis for Learning Outcomes*), whereas the memorization of single sentences might not be affected by the sequence.

Finally, we investigated the extent that different levels of inferential processing while reading contribute to different kinds of learning outcomes in the final test immediately and after a one-week delay. We performed moderated mediation analyses to assess whether the impact of text-sequence on learning can be explained by inferential processes triggered while reading. We expected the inductive inferences while reading to mediate the benefits of an interleaved text sequence on inductive reasoning when participants are spontaneously engaged in self-questioning (*Moderated Mediation Hypothesis*).

## Method

We conducted a 2 x 2 between-subjects factorial experiment with sequence of the text (interleaved vs. blocked) and instruction for self-questioning (participants were spontaneously engaged in self-questioning vs. prompted to generate questions) as fixed factors. Students were randomly assigned between the four learning conditions. They read the text twice. The learning success was assessed immediately after reading and after a one-week delay. The research was conducted in compliance with the Declaration of Helsinki and ethical standards of the DGPS (German Society of Psychology).

### Sample

Our laboratory-experiment included 117 volunteer participants. Three participants, who did not attend the final test after one-week, were excluded from the data analysis. Out of 114 participants, 91% were college students. The age range was from 19 to 55 ($M$ = 24.6 $SD$ = 4.3) and 67.5% were female. Participants were randomly assigned to one of the four learning conditions and were tested individually ($n$ = 28 per condition, except $n$ = 30 in the interleaved/prompted self-questioning condition). As compensation, participants received either an academic credit or 10€. Participants were also entered into a raffle for a voucher valued at 20€, if they correctly answer at least 50% of the questions in the final test.

### Learning Material

We developed an expository text with descriptions of six whales (*humpback whale, fin whale, blue whale, sperm whale, narwhale,* and *killer whale*) on six characteristics: *classification (baleen vs. toothed), size and weight, annual habitat, group's size and behavior*, *lifespan,* and *sounds in communication*. The full text in the canonical (blocked) sequence is available on https://osf.io/mr6a4/ (and in the interleaved sequence on https://osf.io/gzu8c/).

The text comprised 1,060 words and contained seven paragraphs. The introductory paragraph (life of whales) was the same in all conditions. Six additional paragraphs were sequenced either blocked or interleaved. In the blocked sequence, each paragraph included the whole characterization of a particular whale. In the interleaved condition, whales were juxtaposed on a particular characteristic in each paragraph. See Appendix for a detailed depiction of the sequences.

The reader should note that whale characteristics generally covary. For example, baleens are larger and heavier than toothed whales, have a higher lifespan, live in smaller groups, are less likely to be located close to the shore, migrate along with seasonal changes, do not use echo location (only toothed whales do), and their females are larger than males (male toothed whales are larger than females). These characteristics are functionally linked to each other and to the needs set by the environment. For example, food and group size are related since krill can be consumed more efficiently living solitarily, which is the opposite for hunting fish.

Those functional links were not explicit in the text. Thus, the sentences were interconnected by their implicit—functional—links, resulting in a high element interactivity (Sweller, 2010), that is, a high necessity to establish the links by oneself. The co-occurring patterns across characteristics were also not directly reported. This does not make the text per se incoherent because the text allows such conclusions to be drawn based on the pattern of reported characteristics. Thus, learners could actively process the text and ask questions such as *Why does this whale have this migration pattern? Why does it live in groups of that size?* (i.e., making inductive inferences). Answering such questions would result in causal inferences contributing to a coherent mental representation of different whales. Furthermore, the text also lacked the comparative statements. Thus, readers could only conclude, for example, which whale is larger and heavier by directly comparing the whales on a given characteristic (i.e., making comparative inferences).

Omitting inferential statements was essential for our design because we investigated the effect of text structure on inferential processing. We hence have made every effort regarding the selection of the learning material to find a balance between a canonical expository text design (e.g., with regard to coherence) on the one hand and the aim of our study on the other.

**Assessment of Learning Processes**

Each text paragraph was displayed on a separate slide and accompanied by a text box for typing. In the prompted self-questioning conditions, readers were required to generate a question at each paragraph and answer it. In the spontaneous activity conditions, no specific instruction was given; learners were merely told to type their thoughts about the text. We classified the text box entries in all groups according to three hierarchical levels of inferences. Text box entries were coded as *low-level inferences on factual details* when participant comments merely stated explicit information (e.g., the precise weight of a whale). *Comparative inferences* were recorded when the responses referred to comparisons (e.g., which whale is heavier or lighter). A comparative inference thus inherently requires factual details to be compared. *Inductive inferences* were recorded when responses referred to the discovery of how characteristics of whales are related. An inductive inference thus inherently requires comparative inferences to be related (e.g., *small* body sizes with *larger* group sizes). Table 1 displays response examples of these three inference types depending on self-questioning instruction (spontaneous vs. prompted). Per text box entry, only the highest level of inference was coded, that is for example, when a text box entry was classified as an inductive inference, lower levels (low-level and comparative) were not coded.

We operationalized e*laborative inferences* as responses reflecting information that is not stated explicitly nor can be concluded based on the text, but instead requires integration with the world knowledge. We also recorded indistinct and missing responses. The first and the second author coded the text box entries of 20 participants (five per condition). Interrater reliability was .97, *p* < .001. Thus, only one rater (the second author) coded the remaining text box entries. Unclear responses were resolved by discussion.

**Testing Material**

Three subsets of questions were prepared to assess learning performance. All items had a multiple-choice format. The internal consistency as measured by Cronbach’s alpha coefficient ranged from .46 to .64.

*Items on comparative reasoning* (nine in total; Cronbach’s α for the immediate testing = .50; Cronbach’s α for the delayed testing = .46) required participants to choose the correct whale on a given comparative question (e.g., *Which whale has the longest lifespan?*). Thus, to correctly answer questions on comparative reasoning, participants were required to make comparisons among the whales and abstract from absolute values reported in the text (e.g., particular lifespan). For example, the life expectancy of humpback whales is estimated at 45 years. In contrast, the life expectancy of fin whales is estimated at 80 to over 100 years. Therefore, fin whales live *longer* (comparative inference).

*Items on inductive reasoning* (19 in total; Cronbach’s α for the immediate testing = .64; Cronbach’s α for the delayed testing = .63) assess the interconnectedness of mental representations. To correctly answer these questions, learners are required to identify the underlying regularities among whale characteristics. The reader should note that the regularities were not directly reported in the text. These items required participants either to assign a complementary characteristic to a given characteristic (e.g., *This whale uses echolocation. What is its approximate size?*) or to identify the incorrect characteristic without the relevant whale appearing in the text. For example, the item *“Whale watchers catch sight of a whale group with 20 members. Which statement is definitely wrong?”* has the following choices: a) *this whale uses echolocation*. b) *at the beginning of the warmer season, this whale migrates polarwards* (correct answer) c) *this whale weigh as much as 7.5 tons* d) *its average lifespan is between 30 and 50 years*.

*Item memorization of factual details* (13 in total; Cronbach’s α for the immediate testing = .54; Cronbach’s α for the delayed testing = .53) simply required participants to assign the correct characteristic to a given whale (e.g., *a killer whale is a representative of which subordination?*). These questions assess the memory of single sentences and do not require learners to make comparisons or linkages among characteristics.

### Procedure

Participants were tested in the laboratory individually or in groups of up to four individuals. They were instructed to memorize and comprehend the content because both aspects of learning would be tested. If they correctly answered at least 50% of the questions in the final test, they were entered into a raffle for a voucher valued at 20€. Participants read the text at their pace. Each paragraph was displayed on a separate slide and accompanied by a text box. Learners were asked to type their responses to the task in the text box. They were required to read the text twice. We were interested in whether learners would switch their attentional focus during the second reading, for example, from comparing to relating the characteristics. Letting students read the text one more time also provided them with an opportunity to evaluate their hypotheses on regularities based on a complete text-based representation. Students answered the multiple-choice final test questions immediately after reading and after a one-week delay.

## Results

The data are publicly available on https://osf.io/g4hxd/.

### Learning Processes

We computed separate ANOVAs with two between-subjects factors (sequence and self-questioning). The dependent measure for each ANOVA was the type of inferences reflected in the text box entries.

Figure 1 displays the distribution of average frequencies in generating inferences of different cognitive levels while reading as a function of sequence (blocked vs. interleaved) and self-questioning (spontaneous vs. prompted).

### *Comparative Inferences*

The main effect of sequence was significant, $F(1,110) = 112.27, p < .001, \eta_p^2 = .51$. Participants produced more comparative inferences while reading an interleaved text ($M = 2.54, SE = 0.15$) than a blocked text ($M = 0.25, SE = 0.15$). Self-questioning also had a significant impact, $F(1,110) = 5.07, p = .026, \eta_p^2 = .04$. Students who were instructed to generate questions produced more comparative inferences ($M = 1.64, SE = 0.15$) compared to students who were spontaneously engaged in self-questioning ($M = 1.15, SE = 0.15$). Both between-subjects factors interacted, $F(1,110) = 9.48, p = .003, \eta_p^2 = .08$. Participants who generated questions produced significantly more comparative inferences compared to participants who were spontaneously active while reading an interleaved text, $p < .001$, 95% *CI* [0.55, 1.75], $MD = 1.15, SE = 0.30$. In contrast, self-questioning had no impact while reading a blocked text, $p = .564$, 95% *CI* [-0.79, 0.43], $MD = -0.18, SE = 0.31$. Thus, the main effect of self-questioning can be ascribed to the interaction between sequence and self-questioning. The simple comparisons between the interleaved and blocked sequence for the spontaneous and prompted self-questioning conditions revealed a higher frequency of making comparative inferences while reading an interleaved as opposed to blocked text in combination with prompted self-questioning, $p < .001$, 95% *CI* [2.36, 3.56], $MD = 2.96, SE = 0.30$, as well as spontaneous activity, $p < .001$, 95% *CI* [1.01, 2.24], $MD = 1.63, SE = 0.31$. Thus, the main effect of sequence was present irrespective of whether participants were spontaneously engaged in self-questioning or prompted.

### *Inductive Inferences*

Sequence had a significant impact on making inductive inferences, $F(1,110) = 23.79, p < .001, \eta_p^2 = .18$. Participants who read an interleaved text produced more inductive inferences ($M = 0.77, SE = 0.11$) than their counterparts who read a blocked text ($M = 0.01, SE = 0.11$). Self-questioning had no significant impact, $F(1,110) = 3.55, p = .062, \eta_p^2 = .03$. However, the interaction term between sequence and self-questioning was significant,

$F(1,110) = 3.99$, $p = .048$, $\eta_p^2 = .04$. Simple comparisons between spontaneous and prompted self-questioning per sequence type revealed the following pattern. Students who were engaged in spontaneous self-questioning generated more inductive inferences than their counterparts who were prompted to generate questions while reading an interleaved text, $p = .007$, 95% *CI* [0.17, 1.04], $MD = 0.61$, $SE = 0.22$. In contrast, spontaneous and prompted self-questioning did not differ while reading a blocked text, $p = .936$, 95% *CI* [-0.46, 0.42], $MD = -0.02$, $SE = 0.22$. Simple comparisons between interleaving and blocking in spontaneous and prompted self-questioning conditions revealed the superiority of interleaving in students who were spontaneously engaged in self-questioning, $p < .001$, 95% *CI* [0.63, 1.51], $MD = 1.07$, $SE = 0.22$, as well as prompted to generate questions, $p = .042$, 95% *CI* [0.02, 0.88], $MD = 0.45$, $SE = 0.22$. Thus, the main effect of sequence was present irrespective of whether participants were spontaneously engaged in self-questioning or prompted.

### ***Low-Level Inferences on Factual Details***

Sequence had a significant impact on making low-level inferences, $F(1,110) = 171.62$, $p < .001$, $\eta_p^2 = .61$. Participants who read a blocked text payed more attention to factual details ($M = 5.02$, $SE = 0.18$) than their counterparts who read an interleaved text ($M = 1.75$, $SE = 0.18$). Self-questioning also had a significant impact, $F(1,110) = 26.50$, $p < .001$, $\eta_p^2 = .19$. Students who were prompted to generate questions payed more attention to factual details ($M = 4.03$, $SE = 0.18$) than students who were spontaneously engaged in self-questioning ($M = 2.74$, $SE = 0.18$). No interaction between sequence and self-questioning was found, $F < 1$.

### ***Elaborative Inferences***

We found no main effect of sequence, $F(1,110) = 1.74$, $p = .190$, $\eta_p^2 = .02$, but a significant impact of prompted self-questioning over spontaneous self-questioning on making elaborative inferences while reading, $F(1,110) = 7.75$, $p = .006$, $\eta_p^2 = .34$, 95% *CI* [.04, .23], *MD* = .14, *SE* = .05. There was no significant interaction of sequence and self-questioning, *F* < 1.

### *Missing Responses*

The analysis of missing responses revealed no main effect of sequence, *F* < 1, but a main effect of self-questioning, *F*(1,110) = 56.80, *p* < .001, $\eta_p^2$ = .34: Participants engaged in spontaneous self-questioning gave no responses to, on average, 3.56 (*SD* = 3.58) of twelve paragraphs (six per reading cycle). Thus, participants in the spontaneous activity conditions gave no responses to 29.61% of the paragraphs (for comparison, see Figure 1). In contrast, in prompted self-questioning conditions, participants responded to all of the paragraphs, resulting in no missing responses. The analysis revealed no interaction with the two between-subjects factors, *F* < 1.

## Learning Outcomes

We computed three separate repeated measures ANOVAs for the proportion of correctly solved items that assessed comparative reasoning, inductive reasoning, and memorization of factual details. We included the two between-subjects factors, sequence (interleaved vs. blocked) and self-questioning (spontaneous vs. prompted), and the within-subjects factor of testing delay (immediate, T1 vs. one week later, T2).

### *Comparative Reasoning*

Figure 2 (above) shows the pattern of results for the proportion of correctly solved questions on comparative reasoning. The analysis revealed a main effect of sequence, *F*(1,110) = 4.79, *p* = .031, $\eta_p^2$ = .04, indicating the superiority of reading an interleaved text (*M* = 0.46, *SE* = 0.02) over blocked text (*M* = 0.39, *SE* = 0.02). The effect of self-questioning was not significant, *F*(1,110) = 3.33, *p* = .071, $\eta_p^2$ = .03. Both between-subjects factors significantly interacted, *F*(1,110) = 15.91, *p* < .001, $\eta_p^2$ = .13. The simple comparisons revealed that self-questioning matters when reading an interleaved text. Spontaneous activity lead to a higher performance than prompted self-questioning, *p* < .001, 95% *CI* [.10, .27], *MD*

= 0.18, *SE* = 0.04. In contrast, self-questioning had no effect when reading a blocked text, *p* = .132, 95% *CI* [-.16, .02], *MD* = -0.07, *SE* = 0.04. Additionally, the benefits of interleaving over blocking were found only when participants were spontaneously engaged in inferential processing while reading, *p* < .001, 95% *CI* [.10, .28], *MD* = 0.19, *SE* = 0.04 whereas no benefits were found when participants were prompted to generate questions, *p* = .202, 95% *CI* [-.14, .03], *MD* = -0.06, *SE* = 0.04. Thus, the main effect of interleaving can be ascribed to its interaction with self-questioning.

The main effect of delay was significant, *F*(1,110) = 15.57, *p* < .001, $\eta_p^2$ = .12, indicating the decrease of performance over time. We found no interactions of delay with the between-subjects factors, neither with sequence, *F*(1,110) = 2.77, *p* = .099, $\eta_p^2$ = .03, nor with self-questioning, *F*(1,110) = 2.82, *p* = .096, $\eta_p^2$ = .03, and the three-way interaction was not significant, *F*(1,110) = 1.80, *p* = .183, $\eta_p^2$ = .02.

### ***Inductive Reasoning***

Figure 2 (middle) shows the pattern of results for the proportion of correctly solved questions on inductive reasoning. The analysis revealed a main effect of sequence, *F*(1,110) = 4.52, *p* = .036, $\eta_p^2$ = .04, indicating the superiority of reading an interleaved text (*M* = 0.64, *SE* = 0.02) over blocked text (*M* = 0.59, *SE* = 0.02). The effect of self-questioning was not significant, *F*(1,110) = 1.69, *p* = .197, $\eta_p^2$ = .02. Both between-subjects factors significantly interacted, *F*(1,110) = 5.89, *p* = .017, $\eta_p^2$ = .05. The simple comparisons revealed that self-questioning matters when reading an interleaved text. Spontaneous activity lead to a higher performance than prompted self-questioning, *p* = .009, 95% *CI* [.02, .16], *MD* = 0.09, *SE* = 0.04. In contrast, self-questioning had no effect when reading a blocked text, *p* = .430, 95% *CI* [-.10, .04], *MD* = -0.03, *SE* = 0.04. Additionally, the benefits of interleaving over blocking were found only when participants were spontaneously engaged in inferential processing while reading, *p* = .002, 95% *CI* [.04, .18], *MD* = 0.11, *SE* = 0.04, whereas no benefits were found when participants were prompted to generate questions, *p* = .830, 95% *CI* [-.08, .06], *MD* = -0.01, *SE* = 0.04. Thus, the main effect of interleaving can be ascribed to its interaction with self-questioning.

The main effect of delay was significant, *F*(1,110) = 3.99, *p* = .048, $\eta_p^2$ = .04, indicating the decrease of performance over time. We found no interactions of delay with the between-subjects factors: neither with sequence nor with self-questioning, and the three-way interaction was also not significant, *Fs* < 1.

### *Memorization of Factual Details*

Figure 2 (below) shows the pattern of results for the proportion of correctly solved questions on memorization of factual details. The analysis revealed a main effect of sequence, *F*(1,110) = 8.77, *p* = .004, $\eta_p^2$ = .07, indicating the superiority of reading an interleaved text (*M* = 0.61, *SE* = 0.02) over blocked text (*M* = 0.53, *SE* = 0.02). The effect of self-questioning was not significant, *F*(1,110) = 2.55, *p* = .113, $\eta_p^2$ = .02. Both between-subjects factors significantly interacted, *F*(1,110) = 10.13, *p* = .002, $\eta_p^2$ = .08. The simple comparisons revealed that self-questioning matters when reading an interleaved text. Spontaneous activity lead to a higher performance than prompted self-questioning, *p* = .001, 95% *CI* [.06, .21], *MD* = 0.13, *SE* = 0.04. In contrast, self-questioning had no effect when reading a blocked text, *p* = .269, 95% *CI* [-.12, .03], *MD* = -0.04, *SE* = 0.04. Additionally, the benefits of interleaving over blocking were found only when participants were spontaneously engaged in inferential processing while reading, *p* < .001, 95% *CI* [.09, .25], *MD* = 0.17, *SE* = 0.04, whereas no benefits were found when participants were prompted to generate questions, *p* = .875, 95% *CI* [-.08, .07], *MD* = -0.01, *SE* = 0.04. Thus, the main effect of interleaving can be ascribed to its interaction with self-questioning.

The main effect of delay was significant, *F*(1,110) = 20.24, *p* < .001, $\eta_p^2$ = .16, indicating a decrease in performance over time. We found no interactions of delay with the between-subjects factors: neither with sequence, *F* < 1, nor with self-questioning, *F*(1,110) = 2.30, *p* = .132, $\eta_p^2$ = .02, and the three-way interaction was not significant, *F* < 1.

**Learning Outcomes Mediated by Learning Processes**

In the next step, we analyzed whether inferential processing while reading was related to immediate and long-term learning. Table 2 displays the Pearson correlations across the indices of inferential processes and learning outcomes. The extent to which readers made comparative or elaborative inferences showed no effect on learning, $p$ values > .05. Inductive processing while reading positively affected immediate and delayed learning on all three dependent measures of comparative reasoning, inductive reasoning, and memorization of factual details; correlations ranged between .23 and .34, $p$ values < .05. In contrast, low-level inferences on single sentences showed no effect on the immediate comparative reasoning and the delayed inductive reasoning, $p$ values > .05, and a negative effect on the immediate inductive reasoning ($r$ = -.21), the immediate memorization of factual details ($r$ = -.24), the delayed comparative reasoning ($r$ = -.28), and the delayed memorization of factual details ($r$ = -.21), $p$ values < .05.

Given that solely inductive inferences while reading were positively linked to learning outcomes, we computed three moderated mediation analyses to test whether the effect of interleaving on learning (comparative reasoning, inductive reasoning and memorization of factual details) is mediated by inductive inferences and moderated by self-questioning.[4] The immediate and delayed performance on each type of questions were averaged because of the very similar pattern of results between the immediate and delayed testing. Figure 3 illustrates the components and relations of the moderated mediation model. Sequence was incorporated as the independent factor and self-questioning as the moderating factor. These dichotomous factors were dummy-coded with -.5 and .5 (blocked (-.5), interleaved (.5); spontaneous self-questioning (-.5), prompted self-questioning (.5)). We used Hayes' (2013) process tool to analyze our data via bootstrapping with $m$ = 5000.

[4] We additionally checked whether the negative correlational links between the low-level inferences and some of the learning outcomes would matter. Moderated mediation analyses showed no significant effect for path b with regard to comparative reasoning, $B$ = -.01, $p$ = .255, inductive reasoning, $B$ = -.00, $p$ = .708, and factual details, $B$ = -.00, $p$ = .513.

With regard to path a, we found a main effect of sequence on making inductive inferences while reading, $B = 1.52$, $p < .001$, no main effect of self-questioning, $B = -.59$, $p = .062$, and a significant interaction of sequence and self-questioning, $B = -1.25$, $p = .048$. The effect of interleaving was stronger in the spontaneous activity conditions, $B = 2.14$, $p < .001$, 95% *CI* [1.26, 3.02], than in the prompted self-questioning conditions, which was still significant, $B = .90$, $p = .042$, 95% *CI* [.03, 1.76].

In the following sections we report the findings regarding the effect of making inductive inferences while reading on learning when controlling for conditions (path b), the indirect effect of conditions on learning (path ab), and whether the direct effect of conditions on learning sustains when controlling for making inductive inferences (path c'). The sections are separated by type of questions.

The moderated mediation model is depicted in Figure 4 for the spontaneous self-questioning conditions and in Figure 5 for the prompted self-questioning conditions. The path models are shown only for inductive reasoning because the pattern of results was the same for all three learning outcomes (comparative reasoning, inductive reasoning, and memorization of factual details).

### *Comparative Reasoning*

Path b was significant when controlling for conditions, $B = .03$, $p = .004$, indicating the predictive impact of making inductive inferences while reading on answering comparative questions in the final test. The indirect effect of interleaving was significant in the spontaneous activity conditions, $B = .06$, and in the prompted self-questioning conditions, $B = .02$; that is, the bootstrapped 95% confidence interval of [.01, .11] and [.00, .06] excluded zero (Hayes, 2013). These regression coefficients of the indirect effect were not significantly different because zero was included, 95% *CI* [-.08, .00]. The direct effect of interleaving on comparative reasoning (path c') remained significant in the spontaneous activity conditions, $B = .13$, $p = .005$, 95% *CI* [.04, .23], but failed to reach significance in the prompted self-questioning conditions, *B* = -.08, *p* = .06, 95% *CI* [-.17, .00]. Thus, only in the prompted self-questioning conditions, the impact of interleaving on comparative reasoning was completely mediated by making inductive inferences. In spontaneous activity conditions, in contrast, the impact of interleaving was both direct and indirect.

### ***Inductive Reasoning***

Path b was significant when controlling for conditions, *B* = .02, *p* = .028, indicating the predictive impact of making inductive inferences while reading on answering inductive questions in the final test. The indirect effect of interleaving was significant in spontaneous activity conditions, *B* = .04, as well as in prompted self-questioning conditions, *B* = .02; that is, the bootstrapped 95% confidence interval of [.00, .07] and [.00, .04] excluded zero. These regression coefficients of the indirect effect were not significantly different because zero was included, 95% *CI* [-.05, .00]. The direct effect of interleaving on inductive reasoning (path c') remained significant in the spontaneous activity conditions, *B* = .08, *p* = .046, 95% *CI* [.00, .15], but failed to reach significance in the prompted self-questioning conditions, *B* = -.02, *p* = .522, 95% *CI* [-.09, .05]. Thus, only in the prompted self-questioning condition the impact of interleaving on inductive reasoning was completely mediated by making inductive inferences, whereas in the spontaneous activity conditions, the impact of interleaving was both direct and indirect.

### ***Memorization of Factual Details***

Path b failed to reach significance when controlling for conditions, *B* = .02, *p* = .066, indicating a smaller predictive impact of making inductive inferences while reading on answering questions on memorization of factual details in the final test. However, the indirect effect of interleaving was significant in the spontaneous activity conditions, *B* = .03, and in the prompted self-questioning conditions, *B* = .01; that is, the bootstrapped 95% confidence interval of [.00, .07] and [.00, .03] excluded zero. These regression coefficients of the indirect effect were not significantly different because zero was included, 95% *CI* [-.05, .00]. The direct effect of interleaving on memorization (path c') remained significant in the spontaneous activity conditions, *B* = .14, *p* = .002, 95% *CI* [.05, .22], but failed to reach significance in the prompted self-questioning conditions, *B* = -.02, *p* = .608, 95% *CI* [-.10, .05]. Thus, only in the prompted self-questioning condition the impact of interleaving on memorization was completely mediated by making inductive inferences. In the spontaneous activity conditions, in contrast, the impact of interleaving was both direct and indirect.

## Discussion

The present study served three purposes. First, we wanted to replicate the results from the previous research conducted by Abel, Mai, and Hänze (submitted), which showed immediate learning benefits of interleaving on comparative and inductive reasoning for secondary school pupils, but also to extend the results with more advanced readers and a higher retention-interval of one week. Second, we investigated whether readers of an interleaved text spontaneously apply self-questioning and look for regularities while reading by manipulating the learning instruction (prompted self-questioning vs. spontaneous activity) and eliciting readers' inferential processing. If reading an interleaved text engages readers in self-questioning, self-questioning prompts should not add any gain in terms of inferential processing and learning. Third, we aimed to extend our understanding of how an interleaved sequence supports inductive reasoning by exploring the link of inferential processing to learning performance.

We replicated the results from the previous research conducted by Abel, Mai, and Hänze (submitted). Participants involved in spontaneous activity while reading an interleaved text outperformed their peers who read a blocked text with regard to comparative and inductive reasoning in the immediate and delayed test, confirming the *Learning Outcomes Hypothesis*. Thus, these readers were more likely to identify the underlying regularities between whale characteristics. We additionally extend the findings from the previous research on learning with interleaved text materials by revealing the benefit of interleaving on memorization performance (cf. Dobson, 2011; Hausman & Kornell, 2014; Mandler & DeForest, 1979; Schnotz, 1982). Different from the previous study, which yielded no difference on memorization (cf. Abel, Mai, & Hänze, submitted), we were able to observe this advantage of interleaving probably by increasing the opportunities of reprocessing single sentences (e.g., the students read the text twice) and examining college students, who are more experienced with using reprocessing strategies while reading expository texts than 8th and 9th graders.

Also different from the previous research, the participants in the present study were extrinsically motivated to perform well in the final tests to enter into a raffle for a voucher. We yielded the interleaving effects despite these design differences, which might have worked against our hypotheses by stimulating and supporting learners to overcome the difficulty imposed by a poor text sequence (blocking).

The results clearly demonstrated that reading an interleaved text engages readers in spontaneous inferential processing. Participants in the interleaved conditions made significantly more comparative and inductive inferences while reading compared to participants in the blocked conditions, confirming the *Learning Processes Hypothesis*. In contrast, participants in the blocked condition predominantly payed attention to factual details. We conclude that reading a canonically structured text (blocked) does not stimulate integration processes but rarely extends further than stimulating shallow reading strategies (e.g., repetition). Thus, readers of a blocked text adopted a repetition strategy, whereas readers of an interleaved text were engaged in integration processes (cf. van Dijk & Kintsch, 1983).

Furthermore, if readers are spontaneously engaged in self-questioning while reading an interleaved text, as we have assumed, additional triggering of self-questioning via question generation prompts should have been redundant. In line with this reasoning, readers in the interleaved/spontaneous activity condition were *not* less engaged in inferential processing but made significantly more inductive inferences compared to readers in interleaved/prompted self-questioning condition. We assume that students being faced with the discriminative contrast (making comparisons) become inquisitive, apply self-questioning, and seek for characteristics of whales that covary with their differences in appearance and behavior (e.g., *Why do some whales travel up and down a hemisphere, and others do not? In which characteristics do baleens and toothed whales differ? Is there any link between the size of whales and different sounds they produce?*). The learning advantages of self-questioning while reading is well established in the research on elaborative interrogation (Navratil & Kühl, 2018; Ozgungor & Guthrie, 2004; Seifert, 1994; Smith, Holliday, & Austin, 2010). Corroborating evidence is also provided by Maier et al. (2018), who found more frequent lookbacks for belief-inconsistent information in the interleaved condition compared to the blocked condition, which can be interpreted in terms of a high cognitive engagement when readers face the discriminative contrast.

As further predicted by the *Moderation Hypothesis for Learning Outcomes*, interleaving achieved higher learning gains compared to blocking when readers were involved in spontaneous activity, but no difference emerged when readers were prompted to generate questions. However, both presumptions of this hypothesis could not be confirmed by the results. We predicted prompted self-questioning would trigger inferential processing and thus compensate for the lack of spontaneous inferential processing while reading a blocked text (first presumption by the *Moderation Hypothesis for Learning Processes*) and that self-questioning would be redundant while reading an interleaved text (second presumption). Yet prompted self-questioning did not elicit inferential processing in the blocked condition, neither while reading nor during the final tests, which fails to support the first presumption. No indices were observed in which both blocking conditions differed. Based on this pattern of results, we conclude that prompted self-questioning may be a vain strategy when the text sequence provides no opportunity to make comparisons between the described objects, and blocked sequencing does not. Thus, prompting self-questioning was futile in making use of absent chances.

We also found no support for the second presumption of no difference between both interleaving conditions. Readers in the interleaved/spontaneous activity condition showed a superior learning performance over all other groups. Thus, they also outperformed readers who were prompted to use the self-questioning technique while reading an interleaved text. The data pattern indicate that the prompts may have interfered with a spontaneous curiosity and thus narrowed the attentional focus to content presented within single paragraphs. In line with this reasoning, students in the interleaved/prompted self-questioning condition produced significantly more comparative inferences while reading than all other groups, and less inductive inferences than students in the interleaved/spontaneous self-questioning group. However, only making inductive inferences required readers to push the boundaries of single paragraphs and relate information units from different paragraphs.

Basing on assumptions of the transfer-appropriate-processing (TAP) account, one could have expected to reveal an overlap between mental procedures utilized while reading (process data) and required while testing (outcome data) (cf. McDaniel & Butler, 2011; McNamara & Healy, 2000; Morris, Bransford, & Franks, 1977). Accordingly, frequently making comparisons between the whales should have supported comparative reasoning in the final test. Analogously, focusing at factual details should have supported their memorization, resulting in memorization benefit for blocking. However, our results do not support the assumptions of TAP by showing discrepancies among the process and the outcome data pattern. It seems for example that making comparative inferences alone is neither sufficient to infer regularities nor to recall these inferences during the immediate or delayed test. Analogously, focusing at factual details did not support their memorization.

The results emphasize the importance of making inductive inferences while reading the text. The correlations between the process and the outcome data revealed the predictive impact of inductive inferences on answering questions of all subsets, whereas the comparative, low-level and elaborative inferences showed no impact on answering questions of any subset. The moderated mediation analysis revealed a significant indirect effect of interleaving on learning (when participants were spontaneously engaged in self-questioning), mediated by making inductive inferences, which confirms the *Moderated Mediation Hypothesis*[5]. This pattern of results converges with the finding of the link between coherence construction processes reflected in students' language responses while reading and learning outcomes (Abel & Hänze, 2019; Ainsworth & Burcham, 2007; Allen, McNamara, & McCrudden, 2015; Kurby et al., 2012; Magliano & Millis, 2003). Paraphrases (which can be considered low-level inferences) in contrast do not support the representation of factual details (McNamara, 2004).

We interpret the strong link of making inductive inferences and learning as the hierarchical nature of processes leading to the discovery of regularities. Low-level inferences on factual details may establish the basis for making comparative inferences, which in turn may prepare the reader to make inductive inferences. For example, the conclusion that the body and group size are related (inductive inference) requires readers to relate self-generated comparative inferences: *small* body sizes to *larger* group sizes and *larger* body sizes to *smaller* group sizes. Comparative inferences require readers to abstract the explicit factual details in text. Thus, inductive inferences may depend on more basic cognitive operations. As a result, factual details and comparative inferences may be integrated into a high-order representation of a regularity between two characteristics.[6] Merely paying attention to factual details without any construction and integration activity thus does not support memory.

[5] *Moderated Mediation Hypothesis* predicted the mediation only in the spontaneous self-questioning conditions, but the indirect effect was also significant in prompted self-questioning conditions. We do not consider this a counter evidence because interleaving and blocking were not different with respect to learning outcomes when participants were prompted to generate questions.

[6] Note that this interpretation is not supported by the correlational pattern across the types of inferences: Low-level inferences negatively correlate with comparative and inductive inferences. We do not consider this inconsistency a counter argument. Rather, we attribute this inconsistency to an *inherent* limitation of our assessment tool for inferential processing, which we will also discuss in the limitation section: The tool does not

**Limitations**

The learning success was completely mediated by inductive inferences only in the prompted self-questioning conditions. However, the direct effect of interleaving was larger than the indirect effect in the spontaneous activity conditions (path c', under control of making inductive inferences, in comparison to path ab). Thus, we were not able to fully uncover the mechanism underlying the interleaving effect on learning with expository texts. We ascribe this discrepancy to limitations of our assessment tool for inferential processing while reading (i.e., the distinction between factual, comparative, and inductive inferences). Theoretically, all three cognitive levels might be involved while reading a paragraph, although the text box entries mostly reveal solely the most ostensible type of inference (*either* factual, comparative, or inductive). Hence, the tool does not trace participants' implicit attempts of generating inferences on the next cognitive level. This lack is an important issue because participants could have hesitated to record their speculations on how whale characteristics covary. We presume that the direct effect of interleaving (path c') would decrease because of an increase in the indirect effect (path ab) coefficients when utilizing a more fine-grained assessment tool. Furthermore, learners' previous knowledge was not assessed. Previously, Schnotz (1982, 1984) found a stronger relation between the previous knowledge and recall when reading an aspect-oriented text compared to an object-oriented text. Thus, in the present study, previous knowledge could have interacted with text sequence, presumably favoring high-knowledge learners while learning with an interleaved text.

The Cronbach's alpha coefficients for the internal consistency of the three subsets of questions in the immediate and delayed tests (comparative reasoning, inductive reasoning, and memory of factual details) ranged from .46 to .64. Although the internal consistency of our

capture a particular inference independently of other inferences—but at their expense—because a participant's response is mostly coded *either* as low-level, comparative, *or* inductive inference. Due the hierarchical nature of cognitive processes (comparative inferences require factual details, but inductive inferences require comparative statements such as *smaller* and *larger*) we coded only the highest cognitive level of a response. Thus, frequencies of inferences were *inherently* negatively linked.

subsets of items is below .7, it can be considered satisfactory because of two reasons. First, we defined the subsets of items strictly by an item construction principle. For example, in items on memorization of factual details, it was required to assign the correct characteristic to a whale. The items on comparative reasoning were reversely constructed:  Learners were required to assign the correct whale to a given characteristic. In items on inductive reasoning, learners were required to assign the compatible characteristic based on a given one (without naming or requiring a particular whale). Second, domain specific conceptual knowledge is likely to involve a range of related but discrete aspects of understanding (Taber, 2018). The assessment of *learning* should therefore embrace the content in its diversity. A relatively high internal consistency would in contrast indicate that items cover more or less the same concept. From our point of view, it does not seem reasonable to presume that readers equally distribute their attention across the text passages and consistently make certain types of inferences (or consistently refrain from making certain types of inferences).

It is worth mentioning that while the immediate performance assessment was impeccable, the delayed performance assessment was probably contaminated by the former one due to *testing effect*. Although no feedback was given, it might have been the case that the long-term learning benefit of the interleaved/spontaneous self-questioning group was partially caused by consolidation processes in all groups. Accordingly, the long-term interleaving effects should be treated with caution.

Contrary to our expectations, readers in the interleaved/prompted self-questioning condition were outperformed by readers in the interleaved/spontaneous activity condition. Moreover, they performed equally to readers in the blocked conditions. We suppose that generating questions while reading an interleaved text may have hindered the learning advantage of interleaving. In the following discussion, we address an alternative explanation, referring to the theoretically possible confounds caused by the implementation of *spontaneous activity*, which challenges our key interpretation. The instruction to write down thoughts about the text in the spontaneous self-questioning conditions may have served as a prompt by advancing readers in the interleaved condition because notetaking is considered an effective strategy for fostering comprehension (McDaniel, Howard, & Einstein, 2009; Peper & Mayer, 1978). Nonetheless, several reasons speak against this interpretation. First, studies have successfully used uninstructed notetaking as a control condition to learning with prompts. For example, in the research of Roelle, Berthold, and Renkl (2014), participants in the conditions without prompts received the same text boxes as participants in the conditions with prompts. Participants in the no-prompts conditions received the instruction to use the text boxes to write down thoughts about the explanations, which is exactly what we did. Second, if uninstructed notetaking were an effective learning strategy, then spontaneous self-questioning would have been expected to yield a main effect in terms of learning processes or outcomes, indicating the advantage of spontaneous activity irrespective of the text sequence. However, the spontaneous activity was not different from prompted self-questioning while reading a blocked text. The text box entries in the blocked/spontaneous activity condition were predominantly verbatim because of the focus on factual details. This result is corroborated by findings from previous research that the effectiveness of uninstructed notetaking is low because the poor quality of the notes. Students tend to make verbatim notes (Bretzing & Kulhavy, 1979; Einstein, Morris, & Smith, 1985; Mueller & Oppenheimer, 2014). Third, readers in the spontaneous activity conditions made significantly less responses, indicating less perception of instructional restrictions and a lower commitment to perform the task. Readers wrote only what seemed important or interesting to them. Finally, the interleaving effect in terms of inductive reasoning was primarily demonstrated without the use of prompts while reading (Abel, Mai, & Hänze, submitted). In sum, the *spontaneous activity* label seems to be sufficiently justified despite the superior performance of the interleaved/spontaneous activity condition over the interleaved/prompted self-questioning condition.

The low learning performance in the interleaved/prompted self-questioning condition is not indicative of a poor implementation of prompted self-questioning as a learning strategy in the present study. We found indices that support the supposition of an adequate implementation of prompted self-questioning. For example, participants who were prompted to generate questions produced more elaborative inferences while reading. Furthermore, participants who read blocked text and were prompted to generate questions performed equally well on comparative reasoning questions as their counterparts in the interleaved/ spontaneous activity condition when immediately tested, and better than two other conditions (for comparison, see Figure 2 above).[7]

## Future Directions

In the present study, readers generated on average less than one inductive inference per reading cycle in three of the four conditions. Only readers who were spontaneously engaged in self-questioning while reading an interleaved text generated an inductive inference in one of six paragraphs (for comparison, see Figure 1). That is, readers established a link among merely two of the six characteristics (e.g., a negative correlation between the *body* and *group size*). As the moderated mediation analysis confirmed, simply *one* inductive inference per reading cycle was sufficient to increase the learning performance. Still, readers can perform better. Hence, exploring combinations of sequence (interleaved vs. blocked) with prompts that guide learners' attention to relations between the propositions within the text may be very fruitful for instructional research and valuable for educational praxis. Exploring the relative advantage of interleaving over blocking when readers are directly prompted to discover how differences and similarities in objects co-occur may be particularly fruitful.

[7] We did not previously report this particular finding because it was based on simple comparisons that we computed despite the lack of the three-way interaction of sequence, self-questioning, and delay on comparative reasoning. However, to avoid the beta error we explored the data in more detail. When immediately tested on comparative reasoning, blocked/prompted self-questioning outperformed interleaved/prompted self-questioning, *p* = .035, 95% *CI* [.01, .22], *MD* = .11, *SE* = .05, outperformed blocked/spontaneous activity, *p* = .022, 95% *CI* [-.23, -.02], *MD* = -.12, *SE* = .05, and was equal to interleaved/spontaneous activity, *p* > .05, 95% *CI* [-.21, .80], *MD* = -.06, *SE* = .05.

## Educational Implications

People in general erroneously believe that the blocked sequence is the effective one, whereas an interleaved sequence makes a mess of everything (Kornell & Bjork, 2008; McCabe, 2011; Tauber et al., 2013). Thus, learners are not aware of the benefits of juxtaposing categories on inferential processing. Not only is the majority erroneously convinced that blocking is the superior sequence, this misbelief is also relatively resistant against resolution (Yan et al., 2016). In light of this reasoning, many book designers might design expository texts and non-fiction books following the coherence principle *one category at a time* (blocked manner) due to this common misbelief and in anticipation of learners' expectations. This might apply across various subjects such as biology, chemistry, physics, history, and clinical psychology.

The present study demonstrates that reading a blocked—canonically sequenced—expository text prevents learners from making high-level inferences but engages them in shallow processes (i.e., repetitions), and hampers learning. Reading an interleaved text in contrast engages learners in making high-level inferences such as comparative (i.e., comparisons across categories) and inductive inferences (i.e., identifying co-occurring patterns), and consequently benefits long-term learning in terms of memorization of factual details, comparative and inductive reasoning. The pattern of results indicates that readers of an interleaved expository text spontaneously apply self-questioning and look for covarying similarities and differences across categories. In light of these insights, we suggest to textbook designers to adopt interleaved text structures. That is, to juxtapose the to-be-learned categories when the learning goal requires learners to discriminate categories and identify the underlying patterns.

References

Abel, R., Brunmair, M., & Weissgerber, S. C. (under review). Change one category at a time: Sequence effects beyond interleaving and blocking.

Abel, R., & Hänze, M. (2019). Generating causal relations in scientific texts: The long-term advantages of successful generation. *Front. Psychol. 10:199*. Retrieved from https://doi.org/10.3389/fpsyg.2019.00199

Abel, R., Mai, M., & Hänze, M. (submitted). Text sequence matters for category learning: Interleaving promotes comparisons and discovery of underlying regularities.

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction*, *17*(3), 286–303. https://doi.org/10.1016/j.learninstruc.2007.02.004

Allen, L. K., McNamara, D. S., & McCrudden, M. T. (2015). Change your mind: Investigating the effects of self-explanation in the resolution of misconceptions. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 78–83). Pasadena, CA.

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392–402. https://doi.org/10.3758/s13421-012-0272-7

Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *Contemporary Educational Psychology*, *4*, 145–153.

Britt, M. A., Richter, T., & Rouet, J.-F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist*, *49*(2), 104–122. https://doi.org/10.1080/00461520.2014.916217

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

Bugg, J. M., & McDaniel, M. A. (2012). Selective benefits of question self-generation and answering for remembering expository text. *Journal of Educational Psychology*, *104*(4), 922–931.

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology*, *80*(4), 448–456.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*(1), 1–53. https://doi.org/10.1080/01638539809545019

Dobson, J. L. (2011). Effect of selected "desirable difficulty" learning strategies on the retention of physiology information. *Advances in Physiology Education*, *35*(4), 378–383. https://doi.org/10.1152/advan.00039.2011

Eglington, L. G., & Kang, S. H. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition*, *6*(4), 475–485. https://doi.org/10.1016/j.jarmac.2017.07.005

Einstein, G. O., Morris, J., & Smith, S. (1985). Note-taking, individual differences, and memory for lecture information. *Journal of Educational Psychology*, *77*, 522–532.

Foos, P. W., Mora, J. J., & Tkacz, S. (1994). Student study techniques and the generation effect. *Journal of Educational Psychology*, *86*(4), 567–576. https://doi.org/10.1037//0022-0663.86.4.567

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*(1), 103–128. https://doi.org/10.1037/0022-0663.93.1.103

Hausman, H., & Kornell, N. (2014). Mixing topics while studying does not enhance learning. *Journal of Applied Research in Memory and Cognition*, *3*(3), 153–160. https://doi.org/10.1016/j.jarmac.2014.03.003

Hayes, A. F. (2013). *Mediation, moderation, and conditional process analysis: A regression-based approach*. New York: The Guilford Press.

Helsdingen, A., van Gog, T., & van Merriënboer, J. (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgment task. *Journal of Educational Psychology*, *103*(2), 383–398. https://doi.org/10.1037/a0022370

Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for what. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*, 1388–1393.

Hyönä, J., Lorch, R. F., Jr., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*(1), 44–55. https://doi.org/10.1037//0022-0663.94.1.44

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103.

Kintsch, W. (1988). The role of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*(2), 163–182.

Koch, A., & Eckstein, S. G. (1991). Improvement of reading comprehension of physics texts by students' question formulation. *International Journal of Science Education*, *13*(4), 473–485. Retrieved from DOI: 10.1080/0950069910130410

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, *19*, 585–592.

Kurby, C., Magliano, J. P., Dandotkar, S., Woehrle, J., Gilliam, S., & McNamara, D. S. (2012). Changing how students process and comprehend texts with computer-based self-explanation training. *Faculty Research and Creative Activity*, *25*, 1–48.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, *21*(3), 251–283.

Maier, J., Richter, T., & Britt, M. A. (2018). Cognitive processes underlying the text-belief consistency effect: An eye-movement study. *Applied Cognitive Psychology*, *32*, 171–185. Retrieved from DOI: 10.1002/acp.3391

Mandler, J. M., & DeForest, M. (1979). Is there more than one way to recall a story? *Child Development*, *50*, 886–889.

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

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–199). New York, NY: Taylor & Francis.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, *20*, 516–522.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1–30. DOI: 10.1207/s15326950dp3801_1

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language*, *43*, 652–679.

McNamara, D. S., Kintsch, E., Butler Songer, N., & 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.

Mitchell, C., Kadib, R., Nash, S., Lavis, Y., & Hall, G. (2008). Analysis of the role of associative inhibition in perceptual learning by means of the same-different task. *Journal of Experimental Psychology. Animal Behavior Processes*, *34*(4), 475–485. https://doi.org/10.1037/0097-7403.34.4.475

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science*, *25*, 1159–1168. Retrieved from http://dx.doi.org/10.1177/095679761452

Navratil, S. D., & Kühl, T. (2018). Learning with elaborative interrogations and the impact of learners' emotional states. *Journal of Computer Assisted Learning*. Retrieved from DOI: 10.1111/jcal.12324

Ozgungor, S., & Guthrie, J. T. (2004). Interactions among elaborative interrogation, knowledge, and interest in the process of constructing knowledge from text. *Journal of Educational Psychology*, *96*(3), 437–443.

Peper, R. J., & Mayer, R. E. (1978). Note taking as a generative activity. *Journal of Educational Psychology*, *70*(4), 514–522. Retrieved from http://dx.doi.org/10.1037/0022-0663.70.4.514

Roelle, J., Berthold, K., & Renkl, A. (2014). Two instructional aids to optimise processing and learning from instructional explanations. *Instructional Science*, *42*, 207–228. Retrieved from DOI 10.1007/s11251-013-9277-2

Rogers, T. T., & McClelland, J. L. (2008). Precis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, *31*, 689–749.

Schnotz, W. (1982). How do different readers learn with different text organizations. In A. Flammer & W. Kintsch (Eds.), *Discourse processing* (pp. 87-97). Amsterdam: North-Holland.

Schnotz, W. (1984). Comparative instructional text organization. In H. Mandel, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 53-81). Hillsdale, NJ: Erlbaum.

Seifert, T. L. (1994). Enhancing memory for main ideas using elaborative interrogation. *Contemporary Educational Psychology*, *19*, 360–366.

Smith, B. L., Holliday, W. G., & Austin, H. W. (2010). Students' comprehension of science textbooks using a question-based reading strategy. *Journal of Research in Science Teaching*, *47*(4), 363–379.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ*, *48*, 1273–1296. Retrieved from DOI 10.1007/s11165-016-9602-2

Tauber, S. K., Dunlosky, J., Rawson, R. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review*, *20*, 356–363.

Van Blerkom, D. L., van Blerkom, M. L., & Bertsch, S. (2006). Study strategies and generative learning: what works. *Journal of College Reading and Learning*, *37*(1).

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. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 94–121). Cambridge: Cambridge University Press.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. San Diego, CA: Academic Press.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*, 750–763.

Weinstein, Y., McDermott, K. B., & Roediger, H. L. (2010). A comparison of study strategies for passages: rereading, answering questions, and generating questions. *Journal of Experimental Psychology: Applied*, *16*(3), 308–316. https://doi.org/10.1037/a0020992

Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, *36*(2), 109–129. Retrieved from DOI: 10.1207/S15326950DP3602_2

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918–933. https://doi.org/10.1037/xge0000177

Zulkiply, N. (2013). Effect of interleaving exemplars presented as auditory text on long-term retention in inductive learning. *Procedia - Social and Behavioral Sciences*, *97*, 238–245. https://doi.org/10.1016/j.sbspro.2013.10.228

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215–221.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185.

Table 1

*Sample Responses of the Three Inference Levels Depending on Self-Questioning Instruction*

| | Spontaneous self-questioning | Prompted self-questioning |
|---|---|---|
| Comparative inference | *Oldest whales: blue whale, then fin whale, sperm whale, humpback whale, killer whale, narwhale.* | *Which whale species has the highest life expectancy? Blue whales often reach an age of 90 years.* |
| Inductive inference | *Baleen whales are on average older than toothed whales.* | *Are baleen whales or toothed whales getting older on average? Baleen whales.* |
| Low-level inference | *Humpback whale: 45 years*<br>*Fin whale: 80-100*<br>*Blue whale: 90, single >200*<br>*Sperm whale: ca. 60*<br>*Narwhale: 30-40*<br>*Killer whale: 30-50, can also 90, mostly rather female* | *What is the life expectancy of sperm whales? Approx. 60 years.* |

*Note.* Selected excerpts from participants' text box entries on the lifespan of whales (the lifespan is one of six characteristics). Responses were given either as notes or questions/answers depended on whether participants were assigned to spontaneous or prompted self-questioning. Responses were finally coded either as comparative, inductive or simply low-level inferences. For example, the response in the left corner below was coded as a low-level inference because no further characteristic was related with the lifespan (i.e., no inductive inference was made) and no abstraction of given values was provided (i.e., no comparative inference was made). Instead, the response merely contains explicit information from the text. However, this response could have been coded as a comparative inference if the whales were ordered by size (see response in the upper left corner).

Table 2

*Pearson Correlations Between Dependent Measures*

| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|
| *Learning processes* | | | | | | | | | | |
| 1 Comparative inferences | .21* | -.62** | .03 | .08 | -.06 | .15 | .10 | .11 | .00 | .15 |
| 2 Inductive inferences | | -.47** | .13 | .03 | .32** | .34** | .28** | .34** | .23* | .28** |
| 3 Low-level inferences | | | -.07 | .05 | -.12 | -.21* | -.24** | -.28** | -.13 | -.21* |
| 4 Elaborative inferences | | | | .19* | .14 | -.03 | -.05 | -.01 | -.07 | -.06 |
| 5 Time-on task | | | | | .19* | .12 | .17 | .04 | .10 | .22* |
| *Learning outcomes T1* | | | | | | | | | | |
| 6 Comparative reasoning | | | | | | .50** | .65** | .53** | .53** | .48** |
| 7 Inductive reasoning | | | | | | | .53** | .44** | .72** | .45** |
| 8 Memory of factual details | | | | | | | | .56** | .57** | .60** |
| *Learning outcomes T2* | | | | | | | | | | |
| 9 Comparative reasoning | | | | | | | | | .43** | .49** |
| 10 Inductive reasoning | | | | | | | | | | .49** |
| 11 Memory of factual details | | | | | | | | | | |

*Note*. **p* < .05. ***p* < .01.

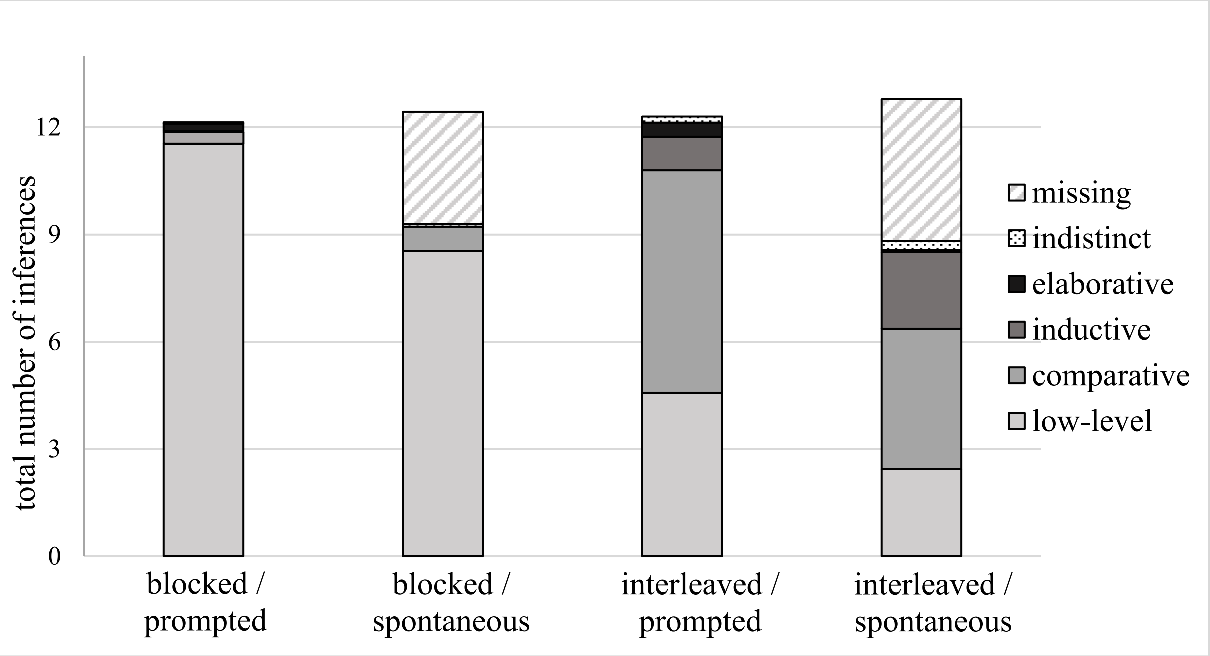


*Figure 1.* The distribution of averaged frequencies in generating inferences of different cognitive levels while reading (collapsed for both reading cycles) as a function of sequence (blocked vs. interleaved) and self-questioning (spontaneous vs. prompted). Participants' text box entries were coded as inferences of different cognitive levels: either low-level, comparative, inductive, or elaborative. Indistinct and missing responses were also recorded. Each bar consists of twelve inferences from the six paragraphs by two reading cycles. A higher number than twelve occurred when a text box entry was assigned to more than one cognitive level.

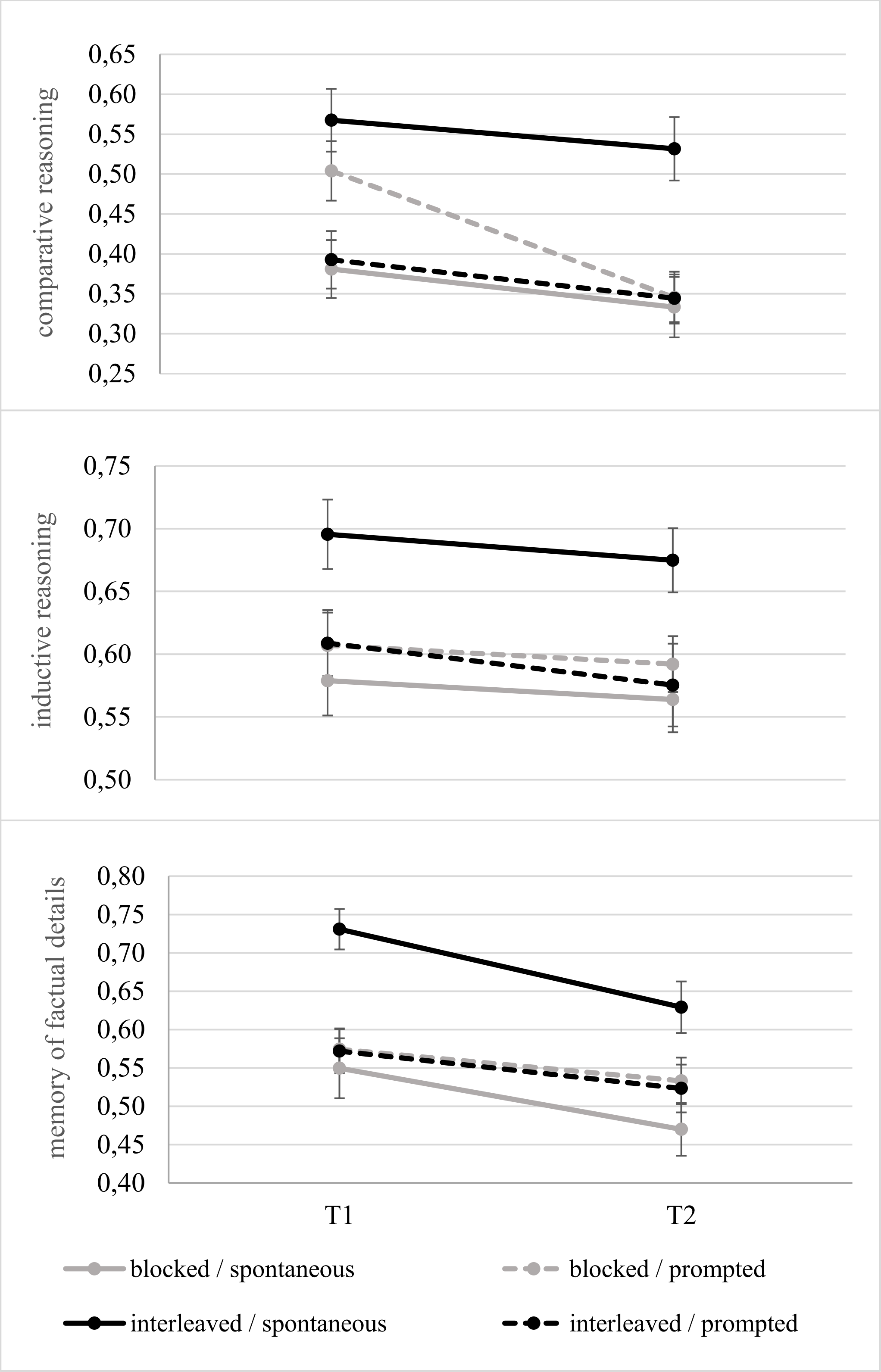


*Figure 2.* Proportion of correctly solved questions on comparative reasoning (above), inductive reasoning (middle), and memorization of factual details (below) in the final test as a function of sequence (interleaved vs. blocked), self-questioning (spontaneous vs. prompted) and testing delay (T1 vs. T2). Estimated means and standard errors are depicted.

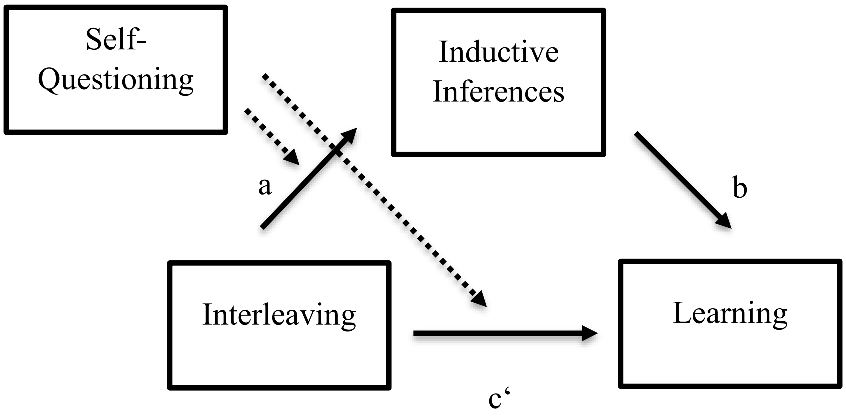
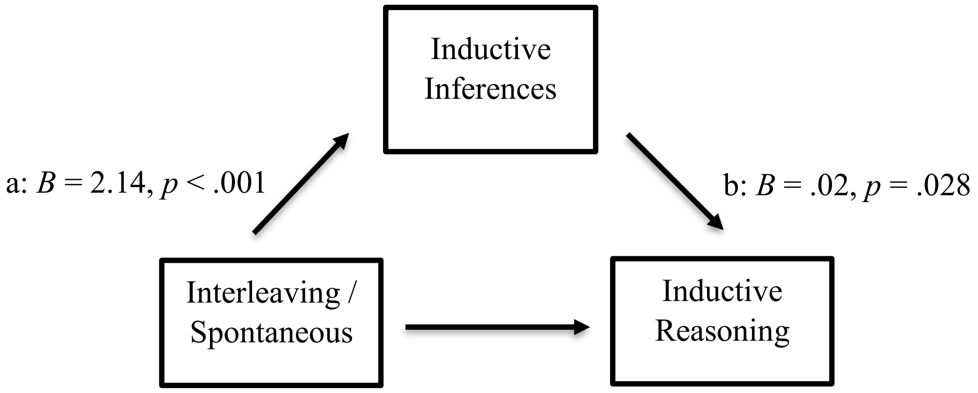


*Figure 3.* The moderated-mediation model. Effect of sequence (interleaving vs. blocking) on learning (comparative reasoning, inductive reasoning, and memorization of factual details; collapsed for T1 and T2) mediated by making inductive inferences and moderated by self-questioning (spontaneous vs. prompted).



Indirect effect ab: $B = .04$, 95% *CI* [.00, .07]

Direct effect c‘: $B = .08$, $p = .046$

*Figure 4*. Mediation model for spontaneous self-questioning conditions. Effect of sequence (interleaving vs. blocking) on inductive reasoning (collapsed for T1 and T2) mediated by making inductive inferences. Note that this pattern of results (a significant indirect effect and a significant direct effect) also applies to the effect of interleaving in the spontaneous self-questioning conditions on the final test questions that assessed the comparative reasoning and memorization of factual details mediated by inductive inferences.

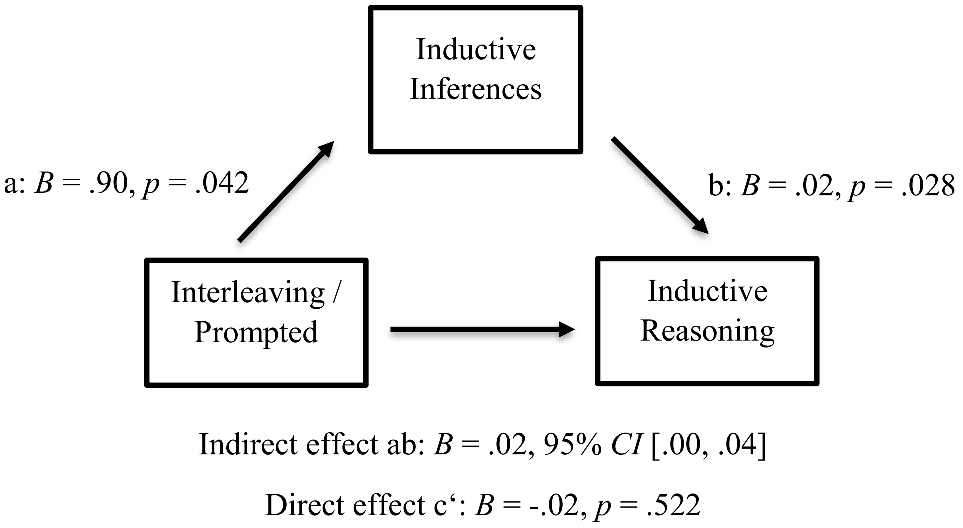


*Figure 5*. Mediation model for prompted self-questioning conditions. Effect of sequence (interleaving vs. blocking) on inductive reasoning (collapsed for T1 and T2) mediated by making inductive inferences. Note that this pattern of results (a significant indirect effect but no direct effect) also applies to the effect of interleaving in the prompted self-questioning conditions on the final test questions that assessed the comparative reasoning and memorization of factual details mediated by inductive inferences.

# Appendix

*Two Sequences of the Expository Text.*

| blocked | interleaved |
|---|---|
| 1a 1b 1c 1d 1e 1f | 1a 2a 3a 4a 5a 6a |
| 2a 2b 2c 2d 2e 2f | 1b 2b 3b 4b 5b 6b |
| 3a 3b 3c 3d 3e 3f | 1c 2c 3c 4c 5c 6c |
| 4a 4b 4c 4d 4e 4f | 1d 2d 3d 4d 5d 6d |
| 5a 5b 5c 5d 5e 5f | 1e 2e 3e 4e 5e 6e |
| 6a 6b 6c 6d 6e 6f | 1f 2f 3f 4f 5f 6f |

*Note.* Digits 1-6 represent the six whales: humpback whale (1), fin whale (2), blue whale (3), sperm whale (4), narwhale (5), and killer whale (6). Characters a-f represent the six characteristics: classification (baleen vs. toothed) (a), size and weight (b), habitat around the year (c), group's size and behavior (d), lifespan (e), and sounds in communication (f). Each combination of a digit and a character represents a sentence describing a particular characteristic of a particular whale. Paragraphs from the text are displayed by rows and contain six sentences each. In the blocked condition, all characteristics of a particular whale (a-f) are grouped. In the interleaved condition, all whales (1-6) are grouped by a particular characteristic.

## Experiment 2a

A version of this article is published as:

Abel, R., & Hänze, M. (2019). Generating causal relations in scientific texts: The long-term advantages of successful generation. *Front. Psychol. 10:199*. https://doi.org/10.3389/fpsyg.2019.00199

## Abstract

A high level of text comprehension can be achieved by engaging learners in processes of organization and integration while reading a cohesive text. In the present study, we investigated the impact of an innovative generative technique on learning with scientific texts. The *cohesion generation* was implemented by means of *explicit* cohesion gaps. High school students ($n$ = 199) were randomly assigned to either receive a fully cohesive scientific text (control condition) or a scientific text that required the selection of causal connectives, such as *because*, *although*, *therefore,* or *however* (generation condition). Learners in the generation condition were required to reflect on causal relations to complete the text. All students were tested immediately (T1) and two weeks after the learning phase (T2). Cognitive load was measured by a dual task and self-report measure. Contrary to our expectations, no differences were found in performance on inference questions (situation model). Learners in the generation condition performed worse on text-based questions at T1 but showed less forgetting from T1 to T2. The impact of condition on the situation model was moderated by reading skills. Remarkably, the generation success was highly predictive for learning outcomes even when controlling for learners' proficiencies. Consequently, learners who succeeded to employ effortful processes to overcome the difficulty showed a superior performance on both the text-base and situation-model questions compared to students reading the cohesive text. Moreover, in these learners, generative activity led to a sustainable learning performance two weeks later. Poor readers especially took advantage of generative activity, despite struggling to perform the cohesion task as indicated by the cognitive load measures. The results suggest that the activity of generating causal relations can augment inferential processing in learners who are not involved in inferential processing spontaneously. To successfully apply this generative learning technique, students require considerable instructional support.

*Keywords:* coherence, causal cohesion, expository text, reading comprehension, reading skill, generation effect, aptitude-treatment interaction, inferential processes, cognitive load, dual-task, desirable difficulties

# Generating causal relations in scientific texts:
# The long-term advantages of successful generation

Expository texts are a major source of scientific knowledge in educational settings. Unexperienced readers, however, struggle with expository texts, because the content in general and the macrostructures of the text are usually unfamiliar to them (Cook & Mayer, 1988; Lorch, 2015; Meyer, 1975). Apart from the complexity and informational density of scientific texts, the multi-causality of scientific phenomena appears to be especially challenging for readers (cf. Britt, Richter, & Rouet, 2014). Accordingly, learners have difficulties selecting the main ideas from the text, organizing them in a meaningful way, and integrating the content with previous knowledge. As reading of expository texts rarely goes beyond a shallow text-based representation, learners fail to construct a coherent representation (situation model) of the learning content. Poor readers may especially struggle to understand the content from scientific texts. As opposed to skilled readers, poor readers have difficulties in bridging inferences from distant idea units in the text and integrating novel content with previous knowledge (Hannon & Daneman, 2001), which are essential processes for the situation-model construction (Kintsch, 1988). Thus, one very important aim of instructional science in general and of this study in particular is to provide recommendations on how to increase the readability of expository texts and to facilitate the processes of knowledge construction during reading.

## The Gap between Cohesion and Coherence

In short, there are two ways to promote learning from expository texts. The first way is to provide learners with a well-written text. The research on reading comprehension has identified several text characteristics that make the text easier to understand (Graesser, McNamara, & Kulikowich, 2011). Among other characteristics, causal cohesion is considered an essential characteristic for supporting the coherence formation (Louwerse, 2001;

Noordman & Vonk, 1997; Sanders & Noordman, 2000). A text can be regarded as causally cohesive if the causal relations between propositions, clauses, and sentences are explicitly marked by connectives, such as *because*, *therefore*, *however*, and *although*. These linguistic markers provide readers with explicit instructions for organizing adjacent and distant concepts from the text into a network of relations (Gernsbacher, 1990; Zwaan & Radvansky, 1998). Moreover, to validate the causal relations encountered in the text, readers make world knowledge inferences by retrieving general premises (Noordman, Vonk, & Kempf, 1992). Thus, connectives support the integration of new content with previous knowledge. When learners lack the necessary knowledge, general premises can be inferred and assimilated into their knowledge base (Cozijn, Noordman, & Vonk, 2011). Numerous studies have demonstrated the positive impact of cohesion devices on the memory of causally connected sentences compared to isolated sentences (Fletcher & Bloom, 1988; cf. Myers, Shinjo, & Duffy, 1987; Trabasso & van den Broek, 1985) and on reading comprehension (Degand, Lefevre, & Bestgen, 1999; Degand & Sanders, 2002; Linderholm et al., 2000; Maury & Teisserenc, 2005; Sanders, Land, & Mulder, 2007; van Silfhout, Evers-Vermeul, Mak, & Sanders, 2014; van Silfhout, Evers-Vermeul, & Sanders, 2014, 2015).

The second way to promote learning from expository texts is by directly engaging learners in active knowledge construction. For example, encouraging learners to self-explain while reading prompts them to draw inferences, monitor their own understanding, and detect and repair the flaws in their mental representation (for a review, see Wylie & Chi, 2014).

Engaging students in active knowledge construction with poorly written texts or engaging them with a cohesive text deprived of active processing provides an insufficient basis for establishing deep comprehension. Apparently, incorporating both cohesion and active processing is necessary to optimize learning. Ainsworth and Burcham (2007) showed that self-explanation from a maximally cohesive text leads to superior comprehension compared to self-explanation from minimally cohesive text. Thus, the function of self-explanation seems to change depending on information provided by the text structure. In minimally cohesive texts, self-explanation serves to compensate for the cohesion gaps, whereas in fully cohesive texts, self-explanation supports the coherence formation based on explicit relations. This finding supports the view that cohesion and generative learning address different aspects of knowledge construction. Both processes appear to be necessary for coherence formation. Active processing should be promoted to establish a congruent relation, whereas linguistic markers should be used to provide the instruction of *how* to relate information. Following this reasoning, active processing of minimally cohesive texts may result in efforts unconnected to schema construction.

Correspondingly, a fully cohesive text itself does not sufficiently initiate coherence formation and often leads to shallow processing. For example, Millis, Graesser, and Haberlandt (1993) found no retention benefit for causally connected statements. Noordman et al. (1992) found that readers did not spontaneously construct inferences of unfamiliar causally related clauses. Instead, the level of active processing depended on how the reader made use of the information. Only those readers who were prompted to judge for inconsistencies or to respond to questions about a causal relation in the text generated inferences. Thus, reading processes heavily depend on learners' goals and the nature of the reading task (Graesser, Haiying, & Feng, 2015).

According to the minimalist hypothesis, reading a locally cohesive text does not result in the generation of global inferences. In contrast, inconsistencies and disruptions on the local level compel readers to draw inferences to fill the gaps (McKoon & Ratcliff, 1992). Reading a well-written text can even result in a decrease of coherence formation in high prior-knowledge learners because well-written texts do not require readers to make inferences (McNamara & Kintsch, 1996; McNamara, Kintsch, Butler Songer, & Kintsch, 1996). In line with this finding, Schworm and Renkl (2006) reported a decrease in quality of self-generated explanations when instructional explanations were provided for learners.

In the present article we address the following problem: a minimally cohesive text promotes processes of coherence formation but does not provide the necessary instructions for how to establish coherence, whereas a fully cohesive text provides the instructions for how to establish coherence but lowers the necessity to do so. These considerations underscore an open gap between cohesion as a text characteristic and coherence as the situation model of text content. Consequently, this article addresses the research question of how to close the gap between cohesion and coherence construction when reading expository texts. For this purpose, we designed a cohesion generation task that was intended to engage learners in coherence construction while reading.

## Benefits and Costs of Generative Learning

A learning advantage of reading strategies that require active processing, compared to passive approaches such as restudying, is called the *generation effect*. According to Wittrock's (1989) generative model of learning, the generation effect is due to the *internal* connections learners build between the information units of the to-be-learned materials and the *external* connections learners build between new content and previous knowledge. The internal and external connections as specified in Wittrock's (1989) model of generative learning can be compared to the central ingredients of further prominent models of meaningful learning, such as the processes of *construction* and *integration* within the *CI* framework (Kintsch, 1988) or the processes of organization and integration within the *select-organize-integrate* (SOI) framework (Mayer, 2014).

The classic experiments on the generation effect (Jacoby, 1978; Slamecka & Graf, 1978) entailed a large body of research on the generation of simple word associates (for a meta-analysis, see Bertsch, Pesta, Wiscott, & McDaniel, 2007; for a review, see McNamara, 1992). In these and similar experiments, learners in the generation condition were presented with incomplete words that needed to be completed according to specific rules. The generative activity of learners engaged them in more effortful processing compared to simply reading, and therefore increased long-term retention. Thus, challenging learners may be regarded as a desirable difficulty (cf. Bjork & Bjork, 2014). However, the insights from studies on generative learning that have employed only word associates in their design are not applicable for educational practice for numerous reasons. A word-completion task does not necessarily involve learners in relational processing nor lead to deep comprehension (McDaniel & Butler, 2011). According to cognitive load theory (CLT), element "interactivity", as defined by CLT, is very low in the case of word lists because the elements can be processed in isolation (Sweller, 2010). Consequently, demands of processing such learning materials are very low. Given the low-complexity of learning materials used in studies on generative learning, the examination of learning outcomes was limited to simple retention. Thus, whether generative activity while studying complex and coherent materials benefits learning remains controversial (Chen, Kalyuga, & Sweller, 2015, 2016). The gains from generative learning in terms of promoting relational processing might be outweighed by the costs of overwhelming learners.

Research on generative learning with complex and coherent materials, such as expository texts, widened the range of learning outcomes toward deep comprehension and transfer of novel knowledge. Additionally, generation activity diverged to particular generation targets (i.e., *what to generate*) and the kind of implementation by the type of task (i.e., *how to generate*). A few prominent generative learning strategies emerged from this research, such as the generation of concept maps (Nesbit & Adesope, 2006), drawings (Leutner & Schmeck, 2014), text structure via sentence scrambling (McDaniel & Butler, 2011), questions (Song, 2016), elaborative interrogations (Seifert, 1994), and self-explanations (Wylie & Chi, 2014). All generation approaches have in common that the to-be-generated units of information should be inferred based on the text rather than retrieved directly from the text (Fiorella & Mayer, 2016). For example, during the drawing activity, learners are required to transform the textual information into a visual representation (Leutner & Schmeck, 2014). In the case of self-explanation, the explanations should elaborate beyond the explicitly provided textual information (Wylie & Chi, 2014). Generation prompts serve the function of either supplementing the elaboration on complete learning materials (e.g., self-explanation during reading, or concept mapping after reading) or completing the initial learning material (e.g., word-completion task or scrambling sentences).

Along with the increased focus on learning material complexity, the consideration of generation success also became important. Successful learning is assumed to be contingent on the accuracy of generation task performance. Thus, learners must be able to perform the generation task accurately to unfold the potential of generative learning. However, most students are barely instructed to use generative learning strategies in educational settings. Given the lack of opportunities to practice generative learning during education, it may not be surprising that students usually process learning content passively and use learning strategies that target only rote learning (cf. Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). The advantages of generative learning may even be reversed, because learners gain a considerably higher expertise in passive learning strategies, such as restudying. According to the *randomness as genesis principle*, unsupported generation imposes a high level of extraneous cognitive load on learners' working memory, which consumes cognitive resources that as a consequence are no longer available for schema construction (Chen et al., 2015, 2016; Paas & Sweller, 2014). Furthermore, learners are not accustomed to performing generative strategies, thus their proficiencies, such as reading skill, previous knowledge, or general intelligence, may substantially contribute to generation success in particular and learning in general. Several studies on generative learning have shown greater advantages of generation on learning when subjects received merely a short-term training on how to perform the generation task (e.g., for drawing, see Leopold, 2009; for summarization, see Friend, 2001; for concept mapping, see Holley, Dansereau, McDonald, Garland, & Collins, 1979; for self-explanation, see McNamara, 2004 and McNamara, O'Reilly, Best, & Ozuru, 2006). Hence, successful learning depends on promoting and supporting active processing.

## Generation of Causal Relations

A growing body of evidence from research on generative learning suggests that elaborating on causal relations supports coherence formation. For example, Allen, McNamara, and McCrudden (2015) found a link between the learning performance and the extent of causal cohesion in students' language responses during self-explanation and think-aloud activities. Kurby et al. (2012) also found that local and distal inferences during self-explanation predicted comprehension. Similarly, Magliano and Millis (2003) demonstrated that readers whose verbal protocols overlapped with causally important sentences from the text achieved higher scores on the comprehension test.

The importance of reflecting on causal relations is broadly acknowledged in the research on generative learning. Generative learning strategies, such as elaborative interrogation, question generation, or concept mapping, likely entail deep processing because of the reflection on factual statements in terms of causes and consequences or reasons and claims. For example, the studies on elaborative interrogation showed that why-prompts promote learners to reflect on reasons, conditions, and causes of certain facts (cf. McDaniel & Donnelly, 1996; Ozgungor & Guthrie, 2004; Smith, Holliday, & Austin, 2010). Similarly, generating high-level questions, which target conceptual and causal relations in text, supports comprehension (Bugg & McDaniel, 2012). Engaging students in learning with concept maps triggers them to analyze the learning content in terms of causes and consequences

(McCrudden, Schraw, Lehman, & Poliquin, 2007). Moreover, theoretical underpinnings and a large body of empirical evidence exists for considering *deep comprehension* as a highly interconnected representation (Zwaan & Radvansky, 1998). Experts, as opposed to novices, possess a sophisticated network-representation of causes and consequences in their knowledge domain (Noordman, Vonk, & Simons, 2000). Accordingly, a powerful generative learning strategy ought to direct learners' focus on causal relations among factual statements in the learning content.

Participants in our study were required to generate causal relations between factual statements in a text in which causal connectives were removed, leaving behind visible gaps. Arguably, the absence of linguistic markers do not automatically promote the processes of organization and integration. Readers need to be aware of the cohesion gaps to close them (Glaser, 1989), but they often miss the *implicit* cohesion gaps in texts. Numerous studies have attributed the inferiority of poorly written texts to learners' inability to close cohesion gaps (Kamalski, Sanders, & Lentz, 2008; McNamara & Kintsch, 1996; McNamara et al., 1996). However, the lack of ability to detect cohesion gaps has yet to be explored as an alternative explanation. Thus, the demands imposed by reading a minimally cohesive text may be additionally attributed to the detection of cohesion gaps. In light of this view, the superiority of self-explaining while reading fully cohesive compared to minimally cohesive texts in the study of Ainsworth and Burcham (2007) can be partially attributed to additional demands that were imposed by *implicit* gaps. In contrast, the cohesion gaps in our study were *explicitly* marked as gaps in the text, and the generation activity was explicitly required for these gaps. We investigated the extent that a cohesion generation task during reading can facilitate construction and integration processes.

## Present Study

The generative learning technique we used extends the existing variety of generation techniques. Learners in the generation condition read text in which conjunction gaps were placed, and they were instructed to establish a causal relation for each gap by choosing the appropriate connective between four alternatives, *because*, *although*, *therefore*, or *however*. These connectives indicated causal relations between clauses, and varied systematically in *polarity*—positive vs. negative—and *direction*—backward vs. forward (cf. taxonomy reported in Louwerse, 2001; Sanders, Spooren, & Noordman, 1992). *Positive* (*because, therefore*) vs. *negative* (*although, however*) refers to confirming vs. violating expectations (Lagerwerf, 1998). The expectation is explicitly conveyed in positive-polarity sentences, whereas negative causal relations add a contrastive meaning to the given causal link. *Backward* (*because*, *although*) vs. *forward* (*therefore*, *however*) refers to the direction of cause and consequence. A backward connective heads the cause, whereas a forward connective is followed by the consequence. Thus, to choose the correct connective, learners were required to indicate the direction (*What is the cause and what is the consequence?*) and polarity (*Are the cause and consequence intuitive or counterintuitive?*). In contrast, without the need to evaluate causal relations while reading a fully cohesive text, the clauses within a sentence might simply be accepted by the readers as being causally related (Cozijn et al., 2011). Accordingly, the generation of causal relations was intended to bridge the gap from cohesion to coherence formation when reading an expository text.

The study was conducted in a German high school. Thus, the text was written in German using German counterparts of causal connectives: *weil* (because), *obwohl* (although), *deswegen* (therefore), and *dennoch* (however). One limitation of using the German language when employing a cohesion generation task should be noted. The direction of connectives and syntax are confounded. The German grammar rules of sentence construction change depending on the connective. The verb in the second clause must be placed next to the connectives *deswegen* and *dennoch* (forward direction), whereas the verb in the second clause must be placed at the end of the sentence when the connectives *weil* and *obwohl* (backward direction) are used. Consequently, the direction can be derived based on the position of the verb. Hence, generation choices could be partially made based on syntactically-driven conclusions.

In our study, each of the to-be-generated target words was embedded between two clauses within a sentence. The choice of target word was based on the meaning of contextual information. For example, the choice between *because* or *therefore* completely depends on the meaning of the preceding and subsequent clauses. Few published studies have used the word-generation task when expository statements are read (e.g., deWinstanley & Bjork, 2004; Peynircioglu & Mungan, 1993). However, in the study of deWinstanley and Bjork (2004), learners were only required to fill in the missing letters of target words. Consequently, the task could be performed nearly independently from the contextual information. In the study of Peynircioglu and Mungan (1993), participants were required to recall words during the final test that they had generated during the learning phase. In contrast, the cohesion generation task in the present study was intended to promote the learning of complex information in the surrounding text.

Given that the advantages of generative learning may be attributed to the processes of organization and integration, we were particularly interested in capturing indices of inferential processing during the generative activity. Thus, along with learning outcomes, we assessed processing measures such as time-on task, generation success, and cognitive load per self-report and via a dual task.

Generative learning—as claimed by the desirable difficulty framework—may lead to a subjective experience of a more effortful processing but also to long-term advantages in learning. Accordingly, participants in the generation condition were expected to experience a higher cognitive load caused by additional inferential processing and to achieve higher test scores after a two-week delay.

## Hypotheses

The generation task targeted the comprehension of relations between the concepts by requiring learners to infer the causal relations between the clauses. Based on the distinction of different levels of information integration in the *CI* framework (Kintsch, 1988), we expected the participants in the generation condition to benefit primarily in terms of the *situation model* assessed by high-level inference questions (H1). Answering such questions requires learners to relate multiple idea units from the text (organization) and to integrate the novel content into a coherent representation. Moreover, the *text-based* representation—assessed by low-level retention questions—also might be promoted (H2), because learners need to reprocess and reflect on the meaning of the previous and successive clauses to establish a causal relation. Information necessary to answer text-based questions can be simply recalled from the memory of single sentences. We also expected a long-term advantage of generative learning in terms of lower forgetting rates (H3).

The level of difficulty can hamper learning. We therefore considered several aspects and restrictions in generative learning. The generation effect on the situation model formation might depend on a successful inference of causal relations during the generation activity. A low accuracy reflects a failed attempt of constructing an appropriate representation of relationships, whereas a high generation accuracy indicates a coherent mental model. Therefore, we expected that only learners who perform accurately on the generation task could take advantage of the generation activity in terms of situation model construction (H4). In contrast, the text-based representation might depend less on generation success, because a low-level question targets the retention of isolated sentences rather than inferences. Thus, we expected to find a generation benefit even in students who showed a low performance during the generation task.

We further assumed that generation success would strongly depend on learners' proficiencies such as reading skills. Results from studies on reading comprehension have suggested that reading skill is an important factor in learning from complex expository texts. Its impact on learning is independent from previous knowledge (O'Reilly & McNamara, 2007; Ozuru, Dempsey, & McNamara, 2009; Voss & Silfies, 1996). The importance of strategic processing increases when learners lack the knowledge needed to bridge cohesion gaps in complex scientific texts (Lorch, 2015). Reading skills help learners to relate multiple ideas and various concepts throughout a text via effortful inferential processing and integrate textual information in a coherent mental representation (Hannon & Daneman, 2001). However, we expected that skilled readers would not benefit from the generation activity as much as the poor readers. The generation activity might be redundant in skilled readers, because a high level of reading skill is associated with spontaneous inferential processing during reading. In contrast, less skilled readers might lack the spontaneous use of inferential processing (McDaniel, Hines, & Guynn, 2002). Thus, an explicit instruction to generate causal relations might engage poor readers in organization and integration processes and in turn promote learning (H5).

## Method

### Design

Participants were randomly assigned to one of two learning conditions. In the control condition, participants read a cohesive text in which the clauses in the text were explicitly linked by means of causal connectives. In the generation condition, the text lacked the connectives. Learners were then instructed to choose between four alternatives—the German counterparts of *because*, *although*, *therefore*, or *however*—from a dropdown list for each missing link. See Table 1 for a direct comparison in which an exemplary paragraph from the control and the generation conditions are juxtaposed. The retrieval interval was manipulated as a within-participants factor, testing participants immediately and after a two-week delay. We tested students' text retention and comprehension.

## Sample

In total, 199 German high school students (grades 10-12) participated in the experiment of which 112 students were randomly assigned to the cohesive text condition and 87 students to the generation condition. Of the 199 students, 21 students were absent during the second examination (12 students in the cohesive text condition and 9 students in the generation condition). On average, students were 18 years old (*M* = 17.7; *SD* = 2.3), 44.7% were female, and 33.1% reported another native language instead of or besides German. The study was conducted during a regular class lesson. Students studied individually with notebooks. We received written informed parental consent for all participants under 18 in accordance with the Declaration of Helsinki.

## Learning materials

The study was programmed with *Inquisit 3* and presented on a notebook screen. Topics of the scientific text were climate change, global warming, and the greenhouse effect. The text was written in German and comprised 18 passages (124 sentences; 2,089 words in total). Each passage was presented on a slide with a headline above the text. Participants could click on *continue* to skip forward to the next passage, but no option was provided to skip back.

The scientific text was developed specifically for this study. The causal relations between the clauses were experimentally manipulated. In the control condition, participants read a fully cohesive text. In this text version, a total of 57 causal relations were made explicit by the four connectives, *weil* (*because*), *deswegen* (*therefore*), *obwohl* (*although*), and *dennoch* (*however*). The frequency of each connective in the text was different because of the constraints in creating text in which the variation in *polarity* throughout the text is more inflexible than the variation in *direction*. Negative-polarity connectives denote the negation of readers' expectations. Thus, a negative causal relation presumes the preexistence of such expectations that contradict the real phenomena (Lagerwerf, 1998). The connectives of negative polarity consequently appeared less frequently in the text (*although* = 7; *however* = 8), whereas the connectives of positive polarity appeared more frequently (*because* = 25; *therefore* = 17).

In the generation condition, the missing connective was indicated by a gap in a sentence. Students were instructed to choose one of the four connectives (*because*, *although*, *therefore*, and *however*) from a dropdown list, which could be activated by clicking on the gap. The choice required the participant to infer the connective based on the causal relation between two clauses. After choosing a connective, it was still possible to reconsider the decision and choose again. All gaps within the presented paragraph were required to be completed before advancing to the next page.

For a direct comparison, Table 1 shows sample text passages about the reflection of sunrays for the control and generation conditions. The text was translated from German into English. Note that the English version contains no syntactical hints on causal relations (see Present Study).

**Measures and scores**

*Learners' proficiencies*

Reading skill was assessed with the Reading-, Speed,- and Comprehension Test for grades 6–12 by Schneider, Schlagmüller, and Ennemoser (2007). According to the manual, students were given four minutes to proceed through the text as far as possible. The simultaneous task was to choose the correct word out of three alternatives for each encountered gap in the text. Given this context, participants were required to select the appropriate term. This measuring instrument was chosen because of the overlapping of cognitive demands with the cohesion generation task.

Previous knowledge on the topic was measured with 16 verification items and two open questions (e.g., *What is the natural greenhouse effect?*). The verbal component of general intelligence was assessed via the word-analogy subtest from the cognitive ability test by Heller and Perleth (2000). This test required that the participants analyze the relation between two presented word stimuli to choose the correct target word out of five alternatives, which is related to a new word stimulus in the same way.

No significant differences were found in the three proficiency measures between the two groups; reading skill $t(176) = 1.31, p = .191$, previous knowledge $t(176) = .75, p = .456$, and word analogy $t(176) = .64, p = .523$.

*Learning processes*

The responses on the cohesion generation task were recorded. The individual scores reflected the number of correctly constructed causal relations out of 57 relations in total. High scores indicate a high level of successful relational processing.

Cognitive involvement during reading was assessed by means of the dual task. Reaction time and accuracy of the responses were recorded. Quick and accurate reactions indicate a low load on working memory (Brünken, Steinbacher, Plass, & Leutner, 2002; cf. Park & Brünken, 2015). The dual task required a quick verification response to a trivial mathematical equation, which was either true (e.g., $5 + 1 = 6$) or false (e.g., $1 + 1 = 0$). A randomly chosen mathematical equation appeared once per slide in a randomly determined moment. Participants were instructed to hold their left hand on the keyboard and press *A* for *false* and *S* for *true* as fast as possible.

To differentiate cognitive load types, a questionnaire developed and evaluated by Leppink, Paas, Van der Vleuten, van Gog, and Van Merrienboer (2013) was used and adopted for the current learning material in German. The scale includes 10 items; three items for *intrinsic load* (e.g., *the topic covered in the activity was very complex*), three items for capturing the *extrinsic load* (e.g., *the explanations were, in terms of learning, very ineffective*), and another four items for *germane load* (e.g., *the activity really enhanced my understanding of the topic covered*). The response scale is between 0 (meaning *not the case at all*) and 10 (meaning *completely the case*).

*Learning outcomes*

The final test consisted of 59 sentence verification tasks, three matching tasks, and three open questions. These questions were designed to assess two different types of knowledge: the text-based representation and the situational model.

The text-based representation was assessed through low-level questions on isolated propositions. The necessary information to answer these questions could be found within single sentences. Text-based questions included 27 verification items and three matching tasks. Students responded to the verification task by choosing whether a statement was true or false. The statements could be recognized based on the explicit information that appeared in the text (e.g., *hot objects emit radiances with a short length* as a true statement; *hydrogen is a greenhouse gas* as a false statement). The matching task required participants to connect detailed information units that belonged together (e.g., *assign the following gases—oxygen, azote, carbon dioxide, noble gases—to the concentrations in the atmosphere—78%, 0.03%, 21%, 1%*). Cronbach's α for the text-based questions were acceptable (immediate testing = .79; delayed testing = .69).

The situation model was assessed through high-level questions, which required participants to draw inferences from multiple sentences in the presented content. Situation model questions included 32 verification items (e.g., *sun radiances can be reflected on sand* as a true statement; *it gets colder on Earth if the warmth gets absorbed* as a false statement) and three open questions. The open questions assessed conceptual understanding (e.g., "*Please explain how it gets warmer within the greenhouse compared to outside*"). The responses on open questions were scored by two student-assistants depending on the number of main ideas mentioned by the participant. The average interrater reliability was .91 for immediate and .95 for delayed testing. Discrepancies were resolved by discussion. Cronbach's α for the situation model questions were .76 for immediate testing and .73 for delayed testing.

**Procedure**

The study was conducted during a regular class lesson. Students studied individually with notebooks. The examination took place on two days with a two-week delay.

Following the test on previous knowledge, subjects received instructions combined with a training on the dual task. The participants were randomly assigned either to read a cohesive text or to generate the causal cohesion while reading an incomplete text. The students were instructed to read the text carefully to be able to answer the questions in the following test on memory and comprehension. Learners in the generation condition were further instructed on how to perform the generation task and to read carefully to be able to accurately choose the correct connective. While reading the text, a mathematical equation appeared once per text-slide. Students were required to quickly indicate whether the equation was true or false (dual task to objectively measure the cognitive load). When the participants finished reading, they answered questions about their experience of cognitive load. Participants then immediately worked on the final test. In most cases, the examination at T1 took no more than an hour.

The follow-up test was administered two weeks later. Participants were tested individually on the computer. They worked on the same questions as two weeks earlier. Then, reading skills and word analogy were assessed. The examination at T2 took approximately half an hour.

## Results

### Learning processes

The generation and control conditions were compared on measures recorded during the learning phase and afterwards by computing independent-samples *t* tests. Means and standard deviations in the time-on task and cognitive load measures are reported in Table 2.

#### *Time-on task*

Learners in the generation condition spent significantly more time reading the text, indicating a higher involvement because of the generation task, $t(197) = -5.85, p < .001$.

#### *Cognitive load via dual task*

The objective measure of cognitive load via a dual task revealed no differences between the two groups in reaction time, $t(197) = 0.47, p = .639$, and response accuracy, $t(197) = -1.74, p = .084$.

#### *Cognitive load via self-report*

The self-report measures of cognitive load were also analyzed. The groups did not differ in their perceived complexity of the text in terms of intrinsic load, $t(197) = -0.90, p = .368$, and germane load, $t(197) = 1.93, p = .055$. However, generation activity imposed a significantly higher extraneous load, $t(197) = -2.05, p = .043$.

### Generation success

Students, on average, chose the correct connective in three out of four sentences ($M_{\%\ \mathrm{accuracy}} = 73.97$; $SD = 15.81$).[8]

[8] We also analyzed whether certain connectives were chosen with a higher accuracy relative to other ones. If so, the second aim would be to determine the impact of the connectives' dimensions of causality *direction* and

*Correlations with learning processes*

We focused on three learning processes involved in the generation activity in or investigation of the impact of generation success on learning outcomes. As Table 3 shows, generation success increased the more time participants spent on reading ($r = .47$, $p < .001$) and the quicker they responded on the dual task ($r = -.41$, $p < .001$). The latter correlation indicated that learners who experienced less restriction on memory capacity could more efficiently employ their cognitive resources for establishing causal relations. This interpretation is supported by the finding that generation success was also associated with a higher level of germane load ($r = .32$, $p = .003$) and a lower level of extraneous load ($r = -.34$, $p = .001$).

*Dependency on learners' proficiencies*

We attribute the individual accuracy in generating causal relations to learners' ability to bridge inferences across isolated ideas in text and to integrate new content into previous knowledge. Thus, the relation between generation success and learners' proficiencies was particularly interesting. Generation success significantly correlated with reading skills ($r = .44$), prior knowledge ($r = .57$), and word analogy ($r = .61$), all $p$ values < .001 (see Table 3). We computed an OLS linear regression with reading skills, prior knowledge, and word analogy as predictor variables, and generation success as a criterion variable. Overall, the model was significant, $F(3,74) = 32.19$, $p < .001$, and explained 56.6% of the variance. All three proficiencies were significant predictors of generation success: reading skill ($\beta = .18$,

*polarity* on generation accuracy. We computed the accuracy rates of all four connectives. An ANOVA with two within-subject factors for direction (forward vs. backward) and polarity (positive vs. negative) was computed. The polarity of causal relations was found to have the highest impact on generation accuracy ($F(1,86) = 144.77$, $p < .001$, $\eta^2 = .63$). Namely, negative connectives *however* ($M = 59.20$, $SD = 26.72$), and *although* ($M = 53.69$, $SD = 21.34$) were more difficult to correctly identify compared to positive connectives *therefore* ($M = 80.19$, $SD = 17.76$) and *because* ($M = 80.14$, $SD = 17.60$), indicating a higher cognitive demand of encoding negative causal relations. We found no significant main effect for *direction*, $F(1,86) = 3.75$, $p = .056$, $\eta^2 = .04$, nor a significant interaction between *direction* and *polarity*, $F(1,86) = 2.90$, $p = .092$, $\eta^2 = .03$.

$t(74) = 2.11, p = .038$), prior knowledge ($\beta = .42$, $t(74) = 5.05, p < .001$), and word analogy ($\beta = .37$, $t(74) = 4.15, p < .001$).

*Impact on learning outcomes*

Generation success was significantly related to learning outcomes for text-based representation and the situation model at both measurement points (correlations ranged between .63 and .76, all *p* values < .001). Note that the correlations between learning outcomes and learners' proficiencies were also significant (correlations ranged between .41 and .64, all *p* values < .001). The question of interest is whether generation success predicts learning outcomes over and above learners' proficiencies. We computed a stepwise regression analysis separately for text-based representation on the immediate and delayed final test scores and the situation model on the immediate and delayed final test scores. We entered the three predictor variables, reading skill, prior knowledge, and word analogy in the first step and generation success in the next step. Generation success significantly predicted the learning outcomes over and above learners' proficiencies: text-based representation T1 ($\beta = .42$, $t(73) = 3.84, p < .001$, $R^2$ changed from .54 to .62, $F(1,73) = 14.78, p < .001$), and T2 ($\beta = .33$, $t(73) = 2.53, p = .014$, $R^2$ changed from .42 to .47, $F(1,73) = 6.38, p = .014$); and the situation model T1 ($\beta = .54$, $t(73) = 4.92, p < .001$, $R^2$ changed from .50 to .62, $F(1,73) = 24.25, p < .001$); and T2 ($\beta = .47$, $t(73) = 4.60, p < .001$, $R^2$ changed from .57 to .66, $F(1,73) = 21.17, p < .001$).

**Learning outcomes irrespective of the generation success**

A repeated measures ANCOVA with the condition (cohesive text vs. generation condition) as a between-subjects factor and the delay (immediate vs. two weeks delay) as a within-subjects factor was computed for text-based representation and the situation model separately. We included the z-standardized score for reading skills as a covariate in the analysis to control for the effect of learners' spontaneous relational processing.

*Text-based representation*

Figure 1 displays the means and standard errors for text-based questions in the final test as a function of condition and retention interval. No significant main effect of condition was found on retention performance collapsed across both tests, $F(1,174) = 3.73$, $p = .055$, $\eta^2 = .02$. Overall, learners performed worse in the delayed test, $F(1,174) = 34.91$, $p < .001$, $\eta^2 = .17$. An interaction between the condition and retention interval was found, $F(1,174) = 7.93$, $p = .005$, $\eta^2 = .04$. Less forgetting occurred over a two-week delay in the generation condition compared to students who read the cohesive text. The significant difference that was found between the conditions at T1 was not significant at T2 ($B = 2.08$, $t(174) = 2.74$, $p = .007$, 95% *CI* [.58, 3.57], $\eta^2 = .04$ vs. $B = .49$, $t(174) = 0.72$, $p = .471$, 95% *CI* [-.86, 1.84] $\eta^2 = .00$). No interaction between condition and reading skills was found, $F(1,174) = 0.81$, $p = .369$, $\eta^2 = .00$.

*Situation model*

Figure 2 displays the means and standard errors for the situation-model questions in the final test as a function of condition and retention interval. No main effect of condition could be found, $F(1,174) = 0.22$, $p = .641$, $\eta^2 = .00$, nor an interaction of condition and retention interval, $F(1,174) = 0.30$, $p = .585$, $\eta^2 = .00$. Again, students performed worse during the delayed test, $F(1,174) = 8.65$, $p = .004$, $\eta^2 = .05$.

Reading skills had a significant impact on comprehension, $F(1,174) = 64.25$, $p < .001$, $\eta^2 = .27$. More importantly, the impact of condition was moderated by learners' reading skill level, $F(1,174) = 4.27$, $p = .040$, $\eta^2 = .02$. Figure 3 displays the estimates for collapsed performance across T1 and T2 on the situation-model questions for learners with a high (+1 *SD*) and a low level of reading skills (-1 *SD*). Neither high-skilled readers scored significantly higher when reading the cohesive text, $p = .077$, nor low-skilled readers performed significantly better when generating cohesion, $p$ = .258. No further significant interactions with reading skills were found.

**Learning outcomes of successful generators**

Given the high impact of generation success on learning, the relative benefits for students who performed highly accurately on the generation task compared to students who simply read the text was further explored. We repeated the analysis on text-based representation and the situation model by means of a repeated-measures ANCOVA, with condition as a between-subjects factor, delay as a within-subjects factor, and reading skills as a covariate. Only students with a generation success of ≥ +1 *SD* ($n$ = 13) were included in the generation condition. These students generated ≥ 90% of causal relations correctly.

*Text-based representation analysis of successful generators*

The results of text-based representation analysis are depicted in Figure 4 (see Figure 1 for comparison with the entire generation group). Students who successfully performed on the generation task highly outperformed the students in the control condition, $F(1,109) = 25.60$, $p < .001$, $\eta^2 = .19$. The performance decreased after the delay, $F(1,109) = 4.14$, $p = .044$, $\eta^2 = .04$). The ANCOVA did not reveal a significant interaction of condition and delay, $F(1,109) = 3.71$, $p = .057$, $\eta^2 = .03$. However, students in the generation condition showed less forgetting. Although the performance in the control condition decreased significantly, performance in the generation condition did not differ between immediate and delayed testing ($p < .001$, 95% *CI* [1.70, 3.20] vs. $p = .955$, 95% *CI* [-2.27, 2.40]). We also found an interaction effect between condition and reading skills, $F(1,109) = 21.75$, $p < .001$, $\eta^2 = .17$. Figure 5 displays the estimates for collapsed performance across T1 and T2 for learners with a high (+1 *SD*) and low level of reading skill (-1 *SD*). Simple comparisons revealed no significant differences between condition for high-skilled readers ($p$ = .473). However, the low-skilled readers showed superior learning performance in the generation condition compared to the control condition ($p < .001$). Thus, poor readers who achieved a high generation accuracy were greatly advantaged by the generation activity, but for skilled readers, the condition did not matter.

*Situation model of successful generators*

The results of the situation model are presented in Figure 6 (see Figure 2 for comparison with the entire generation group). Students who successfully performed the generation task outperformed the students in the control condition, $F(1,109) = 14.88, p < .001, \eta^2 = .12$. The performance decreased after the delay, $F(1,109) = 8.35, p = .005, \eta^2 = .07$. No significant interaction of condition and delay was found, $F(1,109) = 1.53, p = .218, \eta^2 = .01$. We further found an interaction effect between condition and reading skills, $F(1,109) = 8.89$, $p = .004, \eta^2 = .07$. Figure 7 displays the estimates for collapsed performance across T1 and T2 for learners with a high (+1 *SD*) and a low level of reading skills (-1 *SD*; see Figure 3 for comparison with the entire generation group). Simple comparisons revealed no differences between the conditions in high-skilled readers ($p = .968$). In contrast, the low-skilled readers showed superior learning performance in the generation condition compared to the control condition ($p < .001$). Thus, poor readers who achieved a high generation accuracy were greatly advantaged by the generation activity, but for skilled readers, the condition did not matter.

## Discussion

The present study investigated the effect of causal-relation generation—as an innovative generative learning technique—on learning scientific content in high school. We compared students who generated cohesion connectives to students who read a fully cohesive text on learning processes and learning outcomes assessed by an immediate and a two-week delayed test. We could not confirm our assumptions about the effects of generation on learning. We found no main effect of condition on situation-model construction, which contradicted H1. We also found that the immediate text-based representation was inferior to reading a fully cohesive text, contradicting H2. However, we found support for the remaining hypotheses. Students in the generation condition showed less forgetting, confirming H3. Generation success was highly predictive of the situation model even when controlling for learners' proficiencies, confirming H4, but generation success was also a significant predictor of text-based representation. We predicted that text-based representation would be less dependent on generation success, resulting in learning benefits even for worse performers. Thus, the text-based results are not in line with H4. The effect of the learning condition on the situation model was moderated by reading skills, confirming H5. We discuss the results in terms of the conditions under which the generation of causal relations is an undesirable and when it is a desirable difficulty in learning.

In most learners, cohesion generation imposed extraneous cognitive load, resulting in inferior learning. However, the small group of learners who successfully performed well during the generation task took great advantage of generative activity. These advanced learners showed a superior performance on the situation model and text-based questions compared to learners who read a fully cohesive text. Their retention performance was shown to be more sustainable over time. Low-skilled readers especially gained an advantage from successful generation.

### *When is generation undesirable?*

The generation task was implemented by means of cohesion gaps within the sentences. Learners were required to choose the appropriate causal connective to complete a sentence. To establish a correct causal relation between two propositions, learners were required to reflect on the relations among concepts in the text. The impact of generation on learning was expected to be particularly apparent in terms of situation-model construction (H1). The situation model is usually assessed with questions requiring the reader to connect multiple sentences. Thus, learners were expected to apply mental procedures to answer questions on the situation model which overlapped with mental procedures necessary to establish an appropriate causal relation during reading (cf. McNamara & Healy, 2000). However, the ANCOVA revealed no main effect of condition on the situation model for the total sample, irrespective of learners' generation success.

The potential learning advantages of generation could have been reduced by the relatively low generation success.[9] Many learners in the generation condition were unsuccessful in establishing coherence. However, generation is likely to unfold its potential only if learners perform the generation task successfully. This interpretation is supported by a strong correlation between generation accuracy and performance in response to the situation-model questions in the immediate and delayed tests. The predictive power of generation success remained significant even when controlling for learners' proficiencies.

Apart from its impact on coherence formation, we also expected the generation condition to improve the text-based representation (H2). The generation task solely targeted the relation among factual statements. However, participants were required to reinstate the factual information and to check the adequacy of the generated solution to be able to conclude an appropriate relation among statements (Donaldson & Bass, 1980). Thus, the learning advantage of generation was assumed to involve learning content that also served as cues during the generation activity (Greenwald & Johnson, 1989), but participants in the present study who read the fully cohesive text outperformed participants who generated causal relations with respect to the text-based representation when immediately tested.

The text-based representation was expected to be less dependent on generation success than the situation model, because the text-based questions require only factual knowledge.

[9] The average rate of success was 74%. Given the unequal distribution of different connectives across the text and the presence of syntactical hints that indicated the direction of causal relations (see learning materials), the mean error rate of 26% can be regarded as relatively high.

Thus, the text-based representation was expected to be facilitated regardless of whether the sentences were correctly connected or not. Contrary to our expectation, generation success was significantly predictive of the text-based representation and explained a significant amount of variance when controlling for learners' proficiencies. Participants who correctly determined causal relations between factual statements were likely to recall the factual information, probably because a full comprehension of factual statements is necessary to determine the nature of causal relations.

We speculate from this pattern of results that many learners were overwhelmed by the requirements of the generation task and sought syntactical and semantical hints rather than focus on meaningful aspects. The possibility of making inferences based on syntactical structures could have averted learners' attention on meaning. According to the randomness as genesis principle, students without relevant schemas of how to perform the generation task rely on basic operations such as trial-and-error (Chen et al., 2015, 2016). The generation task was intended to widen attentional focus. However, many learners narrowed their attentional focus and processed the learning content fragmentally. Thus, learners who paid attention to irrelevant information units had not managed to construct a coherent or a basic representation of factual statements, resulting in verification of incorrect statements that include terms from the text and resulted in disaffirmation of correct statements that were slightly rephrased. The assumption about fragmented processing in many learners is supported by a higher extraneous cognitive load in the generation condition. A high difficulty of processing a generation task was indicated by a negative correlation between generation success and response reaction times during the dual task: Participants who experienced less cognitive load on working memory capacity could devote free cognitive resources to perform the generation task correctly. This interpretation is also supported by the positive correlation between generation success and germane load, and the negative correlation of generation success with extraneous load.

*Long-term retention*

Confirming H3, participants in the generation condition forgot less after a two-week delay with respect to the text-based representation, regardless of their generation success. The generation task may have reinforced the processing of single sentences. However, a lower forgetting rate in the generation condition did not result in higher text-based scores compared to the cohesive-text condition in the delayed test. The long-term advantage with text-based representations was especially clear in students who performed accurately during the generation task. The advantage of generation on text-based representations, compared to reading a cohesive text, was shown for the immediate testing. This advantage increased two weeks later because of the steeper forgetting rate in the cohesive-text group. These findings are consistent with the well-grounded generation effect in delayed tests (for a brief overview, see Chen et al., 2016). Furthermore, the meta-analysis by Bertsch et al. (2007) revealed an increase in effect sizes of generation benefits from immediate testing to more than a one-day delay.

In view of these findings and the current results, generation slows down forgetting. The effect was clearly pronounced in learners with high generation success, which suggests that the decreased rate of forgetting depends on deep processing. Elaboration of causal relations produces additional retrieval routes in memory, which in turn enhances retrieval (O’Brien & Myers, 1985). Numerous studies on learning techniques, which are considered to be desirable difficulties, revealed lower forgetting rates compared to conventional learning methods (e.g., rereading; for disfluency, see Weissgerber & Reinhard, 2017; for spacing and retrieval practice, see Delaney, Verkoeijen, & Spirgel, 2010). These learning techniques were shown to slow down initial learning but to advantage learning in the long run (Richland, Bjork, Finley, & Linn, 2005).

*For whom is generation desirable?*

High scores on generation accuracy led to a higher and more sustainable learning performance even when controlling for learners' proficiencies (confirming H4). Thus, only accurate performers in the generation task greatly outperformed the students in the control condition in terms of the situation model and text-based retention.

Although generation success depended on learners' proficiencies, such as reading skill, prior knowledge, and word analogy, high-skilled readers did not benefit from generative activity. In fact, skilled readers showed a more elaborated situation model after reading a fully cohesive text. No benefits of generation could be found even when only skilled readers who performed accurately on the generation task were analyzed. Skilled readers appear to spontaneously make use of explicitly marked links in text by generating world knowledge inferences (cf. Cozijn, 2000), and they exert more effort in achieving explanatory coherence (Magliano & Millis, 2003). Thus, generative activity might be redundant. In short, skilled learners are able to successfully generate cohesion, but they do not need it because of their ability to spontaneously engage in bridging inferences.

The impact of learning condition was different for high- and low-skilled readers (confirming H5). Remarkably, poor readers relied less than skilled readers on the instructional support provided by cohesion devices with reference to situation-model construction. When analyzing only students who performed accurately on the generation task, poor readers were greatly advantaged by the generation activity for both the text-based representation and situation model. This pattern of results can be attributed to the lack of spontaneous inferential processing in poor readers (McDaniel & Butler, 2011; McDaniel et al., 2002). In a complementary way, *explicitly* marked cohesion gaps engaged poor readers in inferential processing by minimizing the demands of detecting those gaps. In short, poor readers need stimulation provided by generative prompts, but they are less capable of performing accurately in the generation task. Consequently, poor readers require support on generating causal relations to unfold the full potential of generation.

### *Limitations*

One method limitation that needs to be addressed is our restricted selection of causal connectives that systematically varied along the two dimensions of direction and polarity. Other types of cohesion, such as the referential cohesion (Graesser et al., 2011), and other types of connectives, such as additive or temporal (Louwerse, 2001), and the specialization in either objective (consequence-cause) or subjective (claim-reason) causal relations (Canestrelli, Mak, & Sanders, 2013; Traxler, Sanford, Aked, & Moxey, 1997) were omitted. In follow-up studies, the generation task could be implemented either by forced choice between certain types of connectives or by using a free generation format in which participants could fill the gaps without any restrictions.

From another perspective, our restriction of using only causal connectives can be considered a strength of our method for three reasons. First, the research on how text characteristics affect learning can be differentiated by the broadness of the to-be-manipulated text characteristics and the length and complexity of texts (van Silfhout, Evers-Vermeul, Mak et al., 2014). Many studies have manipulated a very narrow text characteristic (e.g., whether *because* occurs or not) using isolated sentences or short texts. In contrast, other studies have defined cohesion broadly, varying many characteristics at once throughout full-length texts. We advanced the research by simultaneously manipulating just one text characteristic in full-length expository text. Second, deep understanding of scientific phenomena, such as the greenhouse effect and climate change, requires learners to understand causes and consequences in dynamic systems. Understanding causal relations should therefore be the major aim of studying such phenomena. Third, additional types of connectives, such as additive and temporal, are underspecified if they serve in causal relations (Louwerse, 2001; Sanders & Noordman, 2000) and less important for understanding (Noordman & Vonk, 1997). In other words, temporal and additive connectives provide no additional information that is not already addressed by causal connectives in causal relations. Instead, causal relations typically imply temporal and additive relations. In line with this reasoning, Goldman and Murray (1992) found that students overuse causal connectives compared to other types of connectives.

We used a dual task to objectively measure the cognitive load imposed by the generative activity. A dual task usually serves one of two possible functions by either interfering with the learning activity, which consumes necessary cognitive resources (time-on task and accuracy on the prior task would indicate the degree of interference), or the task is affected by the learning activity. An aim of the present study was to measure the impact of the generation task on cognitive processes. The dual task had a very low level of difficulty and thus did not resemble the requirements of text comprehension. We therefore expected the generative activity to be unaffected by the dual task. However, the possibility of posing additional load on learners' working memory and interfering with the generation task cannot be excluded (cf. Brünken et al., 2002).

We manipulated learning performance as a within-subjects factor (i.e., students were tested immediately and after a two-week delay), because we were particularly interested in how generative learning affects forgetting rates. This method poses a possible limitation of the effect. Generation effects from T1 to T2 could have been confounded by the testing effect (cf. Butler, 2010). Learners who read a fully cohesive text could especially gain an advantage from being required to retrieve learning content by retention-based items and to elaborate on the learning content by inference-based items (Roelle & Berthold, 2017).

*Issues for implementation and future directions*

In the present study, we attempted to promote relational processing by requiring students to generate causal relations during reading. To provide teachers and learners with an innovative learning technique, the implementation of the generative activity was intended to be easily applicable in educational settings (Dunlosky et al., 2013). Filling gaps in incomplete sentences is known as a conventional way to promote active processing in school. Thus, students are familiar with this type of task, commonly called *fill-in-the-blank*. Accordingly, choosing the appropriate term might be free from the extraneous load associated with unfamiliarity of the task type.

Although the type of task resembles the well-known fill-in-the-blank technique, the causal cohesion may have appeared to students as an unusual generative activity. Generally, students are inexperienced in reflecting causal relations in terms of direction and polarity. Students especially struggled to correctly determine a negative causal relation, reflecting higher cognitive demands to process adversative causal relations (Goldman & Murray, 1992; Knoepke et al., 2016). Students in our study were challenged by the cognitive demands imposed by an unusual generation target. A relatively low generation success and high scores in cognitive load measures support this view. A consequence of method unfamiliarity could have resulted in an overestimation of the positive impact when reading the fully cohesive text. The potential of the cohesion generation task may have been suppressed by the learners' inexperience with this method (cf. Rummer, Schweppe, Gerst, & Wagner, 2017). To allow for a *fair* comparison between the effectiveness of a fully cohesive text and generating cohesion during reading, familiarity with the activity should be similar between conditions. Familiarity of generative activity notwithstanding, an accurate performance during the generation task is crucial for learning, and students rely heavily on support to perform the generation task.

In follow-up studies on cohesion generation, instructional support in combination with a practice phase should compensate for the inexperience with a generative learning strategy. The instruction should steer learners' attentional focus to the dimensions of causality, namely direction and polarity. Identifying the correct connective in a given constellation of factual statements requires learners to address the following questions: *Which fact is the cause and which is the consequence? Have I expected that A follows from B, or does their relation contradict my expectation?* Selecting the connective *because*, *therefore*, *although*, or *however* directly depends on the answers to these two questions. That is, learners must systematically apply this knowledge to correctly determine the appropriate causal relation. Practicing cohesion generation with corrective feedback might therefore reinforce the autonomous use of this knowledge of causal cohesion during generative learning. Recent evidence points to the advantages of instructional support in increasing sensitivity to causal patterns (Goldwater & Gentner, 2015) or identifying structural components of arguments (von der Mühlen, Richter, Schmid, & Berthold, 2018). Short-term training might particularly increase the awareness in learners about the appropriateness of using cohesion devices.

One further possibility of improving learners' generation performance is to provide them with corrective feedback on their lexical decisions, which was not employed in the present study. For example, in the follow-up study, students could read the text two times. In the generation condition, students could perform the generation task during the first reading, then after receiving the fully cohesive text to be able to reflect on their lexical decisions.

### *Final Conclusions*

Generation is considered a desirable difficulty in learning (Bjork & Bjork, 2014). However, three conditions must be fulfilled to make a difficulty desirable. First, difficulty should promote the processes required to answer questions in the final test (McDaniel & Masson, 1985; McNamara & Healy, 2000). Second, difficulty should promote the processes of knowledge construction not spontaneously initiated by learners (McDaniel & Butler, 2011). Third, difficulty should be surmountable for learners (O'Brien & Myers, 1985). In this study, we proposed an innovative generative learning technique for educational practice. Generation of causal relations appears to be a promising learning tool, because it already fulfills two of the three conditions. First, to establish a coherent mental representation of the text, learners are required to infer the causal relations among the factual statements (process of organization) and to integrate the factual statements with previous knowledge by making world knowledge inferences (cf. Cozijn et al., 2011). As a result, a coherent mental representation supports learners' ability to make inferences required by the final test. Second, cohesion generation can benefit poor readers, because poor readers are usually not engaged in the spontaneous processes of organization and integration (cf. Fiorella & Mayer, 2016). The third condition, however, was not met in this study. The necessary support to overcome the difficulty imposed by the generation task was not provided. Nonetheless, learners who succeeded to employ effortful processing to overcome the difficulty, took great advantages of generative activity. Future research on cohesion generation should incorporate instructional support on the meaning of the two causality dimensions, direction and polarity (Louwerse, 2001; Sanders et al., 1992), including an opportunity to practice. We look forward to further discoveries in the effects of cohesion generation on long-term retention and coherence construction by boosting the generation success rates.

# References

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction*, *17*(3), 286–303. https://doi.org/10.1016/j.learninstruc.2007.02.004

Allen, L. K., McNamara, D. S., & McCrudden, M. T. (2015). Change your mind: Investigating the effects of self-explanation in the resolution of misconceptions. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 78–83). Pasadena, CA.

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

Bjork, E. L., & Bjork, R. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed., pp. 59–68). New York, NY: Worth.

Britt, M. A., Richter, T., & Rouet, J.-F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist*, *49*(2), 104–122. https://doi.org/10.1080/00461520.2014.916217

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, *49*(2), 109–119. https://doi.org/10.1027//1618-3169.49.2.109

Bugg, J. M., & McDaniel, M. A. (2012). Selective benefits of question self-generation and answering for remembering expository text. *Journal of Educational Psychology*, *104*(4), 922–931.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133.

Canestrelli, A. R., Mak, W. M., & Sanders, T. J. M. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, *28*(9), 1394–1413. https://doi.org/10.1080/01690965.2012.685885

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, *107*(3), 689–704. https://doi.org/10.1037/edu0000018

Chen, O., Kalyuga, S., & Sweller, J. (2016). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction*, *45*, 20–30. https://doi.org/10.1016/j.learninstruc.2016.06.007

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology*, *80*(4), 448–456.

Cozijn, R. (2000). *Integration and inference in understanding causal sentences*. Tilburg: Faculteit der Letteren, KUB.

Cozijn, R., Noordman, L. G. M., & Vonk, W. (2011). Propositional integration and world-knowledge inference: Processes in understanding because sentences. *Discourse Processes*, *48*, 475–500.

Degand, L., Lefevre, N., & Bestgen, Y. (1999). The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design*, *1*, 39–51.

Degand, L., & Sanders, T. J. M. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing*, *15*, 739–757.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 53, pp. 63–148). Burlington: Academic Press.

DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, *32*(6), 945–955.

Donaldson, W., & Bass, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning and Verbal Behavior*, *19*, 26–35.

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 and the Public Interest*, *14*, 4–58.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language*, *27*, 235–244.

Friend, R. (2001). Effects of strategy instruction on summary writing of college students. *Contemporary Educational Psychology*, *26*(1), 3–24. https://doi.org/10.1006/ceps.1999.1022

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum.

Glaser, R. (1989). Expertise and learning: How do we think about instructional processes now that we have discovered knowledge structures. In D. Klahr & K. Kotosky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 269–282). Hillsdale, NJ: Erlbaum.

Goldman, S. R., & Murray, J. D. (1992). Knowledge of connectors as cohesion devices in text: a comparative study of native-English and English-as-a-second-language speakers. *Journal of Educational Psychology*, *84*, 504–519.

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137–153. https://doi.org/10.1016/j.cognition.2014.12.001

Graesser, A. C., Haiying, L., & Feng, S. (2015). Constructing inferences in naturalistic reading contexts. In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 290–320). Cambridge: Cambridge University Press.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234.

Greenwald, A. G., & Johnson, M. M. S. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition*, *17*, 673–681.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*(1), 103–128. https://doi.org/10.1037/0022-0663.93.1.103

Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest KFT 4–12 + R (für 4. bis 12. Klassen, Revision)*. [Cognitive ability test for grades 4-12]. Göttingen: Beltz-Test GmbH.

Holley, C. D., Dansereau, D. F., McDonald, B. A., Garland, J. C., & Collins, K. W. (1979). Evaluation of a hierarchical mapping technique as an aid to prose processing. *Contemporary Educational Psychology*, *4*, 227–237.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.

Kamalski, J., Sanders, T. J. M., & Lentz, L. (2008). Coherence marking, prior knowledge, and comprehension of informative persuasive texts: Sorting things out. *Discourse Processes*, *45*, 323–345.

Kintsch, W. (1988). The role of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*(2), 163–182.

Knoepke, J., Richter, T., Isberner, M.-B., Naumann, J., Neeb, Y., & Weinert, S. (2016). Processing of positive-causal and negative-causal coherence relations in primary school children and adults: A test of the cumulative cognitive complexity approach in German. *Journal of Child Language*, 1–32. https://doi.org/10.1017/S0305000915000872

Kurby, C., Magliano, J. P., Dandotkar, S., Woehrle, J., Gilliam, S., & McNamara, D. S. (2012). Changing how students process and comprehend texts with computer-based self-explanation training. *Faculty Research and Creative Activity*, *25*, 1–48.

Lagerwerf, L. (1998). *Causal connectives have presuppositions: Effects on coherence and discourse structure*. The Hague: Holland Academic Graphics.

Leopold, C. (2009). *Lernstrategien und Textverstehen: Spontaner Einsatz und Förderung von Lernstrategien*. [Learning strategies and text comprehension: Spontaneous use and support of learning strategies]. Muenster: Waxmann.

Leppink, J., Paas, F., Van der Vleuten, C. P. M., van Gog, T., & Van Merrienboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Leutner, D., & Schmeck, A. (2014). The generative drawing principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (2nd ed., pp. 433–448). Cambridge: Cambridge University Press.

Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction*, *18*(4), 525–556.

Lorch, R. F., Jr. (2015). What about expository text? In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 348–361). Cambridge: Cambridge University Press.

Louwerse, M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, *12*(3), 291–315.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, *21*(3), 251–283.

Maury, P., & Teisserenc, A. (2005). The role of connectives in science text comprehension and memory. *Language and Cognitive Processes*, *20*(3), 489–512.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge: Cambridge University Press.

McCrudden, M. T., Schraw, G., Lehman, S., & Poliquin, A. (2007). The effect of causal diagrams on text learning. *Contemporary Educational Psychology*, *32*(3), 367–388. https://doi.org/10.1016/j.cedpsych.2005.11.002

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–199). New York, NY: Taylor & Francis.

McDaniel, M. A., & Donnelly, C. M. (1996). Learning with analogy and elaborative interrogation. *Journal of Educational Psychology*, *88*(3), 508–519. https://doi.org/10.1037//0022-0663.88.3.508

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

McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 371–385.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–466. https://doi.org/10.1037//0033-295X.99.3.440

McNamara, D. S. (1992). The generation effect: A detailed analysis of the role of semantic processing. *Technical Report*, *2*, 1–48.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*(1), 1–30. https://doi.org/10.1207/s15326950dp3801_1

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language*, *43*, 652–679.

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

McNamara, D. S., Kintsch, E., Butler Songer, N., & 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.

McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, *34*(2), 147–171. https://doi.org/10.2190/1RU5-HDTJ-A5C8-JVWE

Meyer, B. J. F. (1975). *The organization of prose and its effect on memory*. Amsterdem: North-Holland.

Millis, K. K., Graesser, A. C., & Haberlandt, K. (1993). The impact of connectives on the memory for expository texts. *Applied Cognitive Psychology*, *7*, 317–339.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*, 453–465.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*, 413–448.

Noordman, L. G. M., & Vonk, W. (1997). The different functions of a conjunction in constructing a representation of the discourse. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships: Studies in the production and comprehension of text* (pp. 75–93). Mahawah, NJ: Lawrence Erlbaum.

Noordman, L. G. M., Vonk, W., & Kempf, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, *31*, 573–590.

Noordman, L. G. M., Vonk, W., & Simons, W. H. G. (2000). Knowledge representation in the domain of economics. In L. Lundquist & R. J. Jarvella (Eds.), *Language, Text, and Knowledge: Mental Models of Expert Communication* (pp. 235–260). Berlin/New York: Mouton de Gruyter.

O'Brien, E. J., & Myers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 12–21.

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.

Ozgungor, S., & Guthrie, J. T. (2004). Interactions among elaborative interrogation, knowledge, and interest in the process of constructing knowledge from text. *Journal of Educational Psychology*, *96*(3), 437–443.

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*, 228–242.

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (2nd ed., pp. 27–42). Cambridge: Cambridge University Press.

Park, B., & Brünken, R. (2015). The rhythm method: A new method for measuring cognitive load - an experimental dual-task study. *Applied Cognitive Psychology*, *29*, 232–243.

Peynircioglu, Z. F., & Mungan, E. (1993). Familiarity, relative distinctiveness, and the generation effect. *Memory & Cognition*, *21*(3), 367–374.

Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: Generation and interleaving effects. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*, 1850–1855.

Roelle, J., & Berthold, K. (2017). Effects of incorporating retrieval into learning tasks: The complexity of the tasks matters. *Learning and Instruction*, *49*, 142–156.

Rummer, R., Schweppe, J., Gerst, K., & Wagner, S. (2017). Is testing a more effective learning strategy than note-taking? *Journal of Experimental Psychology: Applied*, *23*(3), 293–300. https://doi.org/10.1037/xap0000134

Sanders, T., Land, J., & Mulder, G. (2007). Linguistic markers of coherence improve text comprehension in functional contexts. *Information Design Journal*, *15*(3), 219–235.

Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in processing. *Discourse Processes*, *29*(1), 37–60.

Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*(1), 1–35. https://doi.org/10.1080/01638539209544800

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12*. [Reading-, Speed,- and Comprehension Test for grades 6–12]. Goettingen: Hogrefe.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, *46*, 426–445.

Seifert, T. L. (1994). Enhancing memory for main ideas using elaborative interrogation. *Contemporary Educational Psychology*, *19*, 360–366.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*(6), 592–604.

Smith, B. L., Holliday, W. G., & Austin, H. W. (2010). Students' comprehension of science textbooks using a question-based reading strategy. *Journal of Research in Science Teaching*, *47*(4), 363–379.

Song, D. (2016). Student-generated questioning and quality questions: A literature review. *Research Journal of Educational Studies and Review*, *2*(5), 58–70.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.

Traxler, M. J., Sanford, A. J., Aked, J. P., & Moxey, L. M. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 88–101.

Van Silfhout, G., Evers-Vermeul, J., Mak, W. M., & Sanders, T. J. M. (2014). Connectives and layout as processing signals: How textual features affect student's processing and text representation. *Journal of Educational Psychology*, *106*(4), 1036–1048.

Van Silfhout, G., Evers-Vermeul, J., & Sanders, T. J. M. (2014). Establishing coherence in schoolbook texts: How connectives and layout affect students' text comprehension. *Dutch Journal of Applied Linguistics*, *3*(1), 1–29.

Van Silfhout, G., Evers-Vermeul, J., & Sanders, T. J. M. (2015). Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, *52*(1), 47–76.

Von der Mühlen, S., Richter, T., Schmid, S., & Berthold, K. (2018). How to improve argumentation comprehension in university students: Experimental test of a training approach. *Instructional Science*, *22*(6), 25. https://doi.org/10.1007/s11251-018-9471-3

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.

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. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376.

Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (2nd ed., pp. 413–432). Cambridge: Cambridge University Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185.

Table 1

*A Sample Text Paragraph Taken from the Control and Generation Condition for Comparison*

| High causal cohesion (control) | Generation of causal cohesion |
|---|---|
| Solar radiation can be absorbed by the Earth's land surface and stored as heat, however some sunrays partially rebound. Reflection can happen on any surface, although some surfaces seem to be unsuitable. In certain cases, this phenomenon is called specular reflection, because the angle of reflection equals the angle of incidence. Diffuse reflection refers to the case that the incident ray is evenly reflected at many angles. If the incident ray is unevenly reflected at many angles, the phenomenon is called mixed reflection. Nature offers a variety of rough surfaces, therefore the mixed reflection is the most common case. A part of sunrays, which have been reflected, do not lose any energy, therefore its waves remain short. The reflected sunrays pass the atmosphere without being absorbed and escape into space because they retain short waves. | Solar radiation can be absorbed by the Earth's land surface and stored as heat, __________ some sunrays partially rebound. Reflection can happen on any surface, __________ some surfaces seem to be unsuitable. In certain cases, this phenomenon is called specular reflection, __________ the angle of reflection equals the angle of incidence. Diffuse reflection refers to the case that the incident ray is evenly reflected at many angles. If the incident ray is unevenly reflected at many angles, the phenomenon is called mixed reflection. Nature offers a variety of rough surfaces, __________ the mixed reflection is the most common case. A part of sunrays, which have been reflected, do not lose any energy, __________ its waves remain short. The reflected sunrays pass the atmosphere without being absorbed and escape into space, __________ they retain short waves. |

*Note*. The text was translated from German. The translation into English lacks the syntactical hints on the direction of causal relations that are necessary in German (see Present Study). In the generation condition, the text lacked the connectives. Learners were instructed to click on the missing links to choose the correct connective from a dropdown list.

Table 2

*Mean Scores and Standard Deviations of Learning Processes*

| | Control | | Generation | |
|---|---|---|---|---|
| Measure | *M* | *SD* | *M* | *SD* |
| Time-on task (in min.) | 15.44 | 5.44 | 20.16 | 5.91 |
| Dual-task reaction time (mean in ms) | 2414 | 1925 | 2311 | 754 |
| Dual-task accuracy (in %) | 94.17 | 7.83 | 95.94 | 6.10 |
| Intrinsic CL | 5.59 | 2.18 | 5.88 | 2.45 |
| Extraneous CL | 3.18 | 2.19 | 3.8 | 2.09 |
| Germane CL | 6.38 | 2.26 | 5.75 | 2.29 |

Table 3

*Pearson Correlations Between Dependent Measures*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Learning processes* | | | | | | | | | | | | | | |
| 1 Time-on task | | -.02 | .19$^{**}$ | .02 | .00 | .12 | .47$^{**}$ | .09 | .11 | .07 | .20$^{**}$ | -.07 | .10 | .02 |
| 2 Dual-task RT | | | .14 | -.08 | .09 | -.06 | -.41$^{**}$ | -.15$^{*}$ | -.05 | -.10 | -.18$^{*}$ | -.10 | -.11 | -.08 |
| 3 Dual-task accuracy | | | | .04 | -.19$^{**}$ | .07 | .07 | .05 | .06 | .16$^{*}$ | .14 | .11 | .03 | .09 |
| 4 Intrinsic CL | | | | | .34$^{**}$ | -.10 | -.03 | -.28$^{**}$ | -.34$^{**}$ | -.25$^{**}$ | -.17$^{*}$ | -.12 | -.27$^{**}$ | -.20$^{**}$ |
| 5 Extraneous CL | | | | | | -.34$^{**}$ | -.34$^{**}$ | -.36$^{**}$ | -.28$^{**}$ | -.36$^{**}$ | -.36$^{**}$ | -.34$^{**}$ | -.30$^{**}$ | -.22$^{**}$ |
| 6 Germane CL | | | | | | | .32$^{**}$ | .34$^{**}$ | .20$^{**}$ | .27$^{**}$ | .33$^{**}$ | .143 | .18$^{*}$ | .16$^{*}$ |
| 7 Generation success | | | | | | | | .73$^{**}$ | .63$^{**}$ | .76$^{**}$ | .76$^{**}$ | .44$^{**}$ | .57$^{**}$ | .61$^{**}$ |
| *Learning outcomes* | | | | | | | | | | | | | | |
| 8 Text base T1 | | | | | | | | | .75$^{**}$ | .74$^{**}$ | .71$^{**}$ | .46$^{**}$ | .61$^{**}$ | .55$^{**}$ |
| 9 Text base T2 | | | | | | | | | | .67$^{**}$ | .68$^{**}$ | .41$^{**}$ | .52$^{**}$ | .58$^{**}$ |
| 10 Situation model T1 | | | | | | | | | | | .74$^{**}$ | .51$^{**}$ | .64$^{**}$ | .60$^{**}$ |
| 11 Situation model T2 | | | | | | | | | | | | .46$^{**}$ | .60$^{**}$ | .59$^{**}$ |
| *Learning proficiencies* | | | | | | | | | | | | | | |
| 12 Reading skill | | | | | | | | | | | | | .36$^{**}$ | .43$^{**}$ |
| 13 Prior knowledge | | | | | | | | | | | | | | .49$^{**}$ |
| 14 Word analogy | | | | | | | | | | | | | | |

*Note*. *$p$ < .05. **$p$ < .01. The correlations with the generation success could only be computed in the generation condition. The correlations with the text-based representation on T2, the situation model on T2, reading skill and word analogy was computed only for participants who completed the examination on T1 and T2.

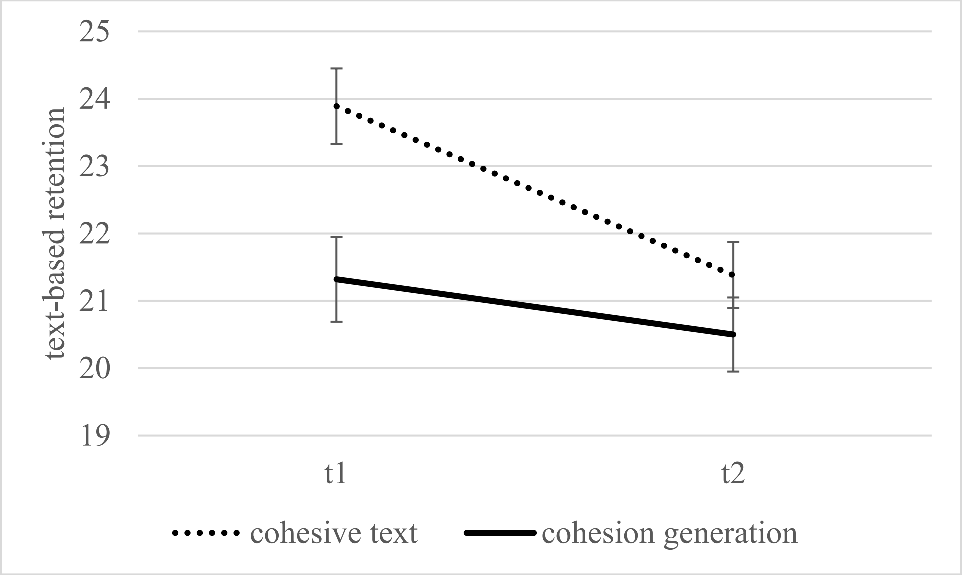


*Figure 1.* Text based representation as a function of condition and delay when controlling for reading skills (estimated means and standard errors). Max. performance was 38.

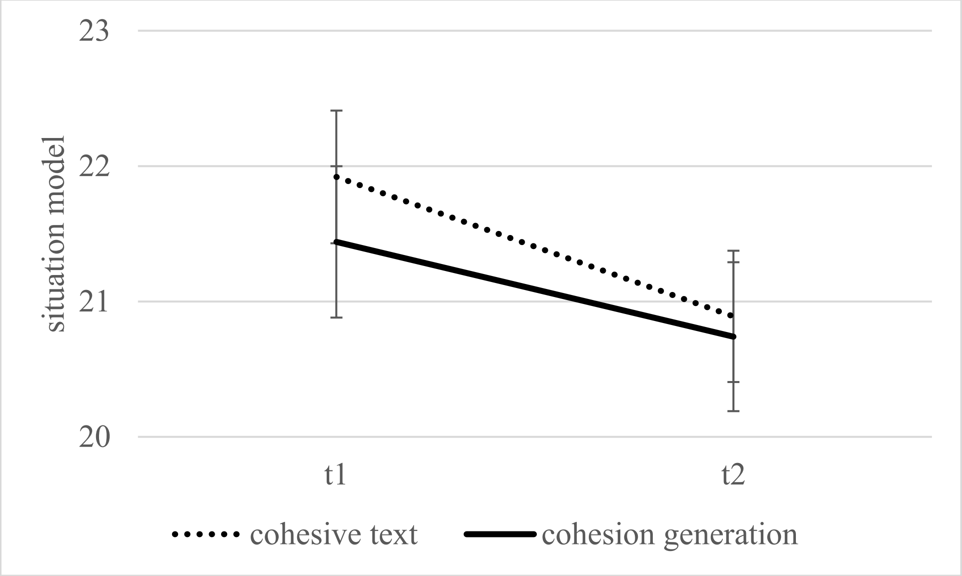


*Figure 2.* The situation model as a function of condition and delay when controlling for reading skills (estimated means and standard errors). Max. performance was 42.

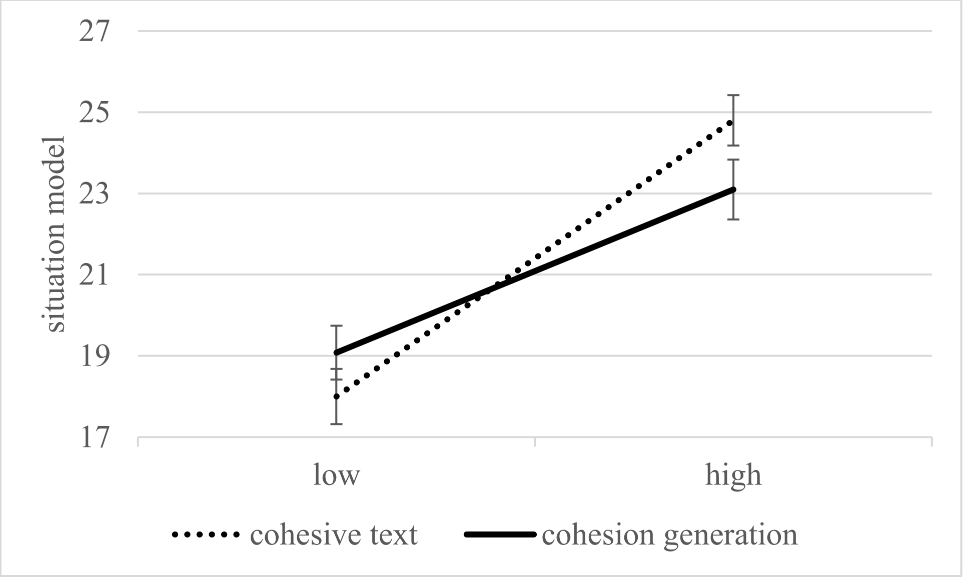


*Figure 3.* The situation model (collapsed across immediate and delayed testing) as a function of condition and the level of reading skills (estimated means and standard errors for -1 *SD* and +1 *SD*). Max. performance was 42.

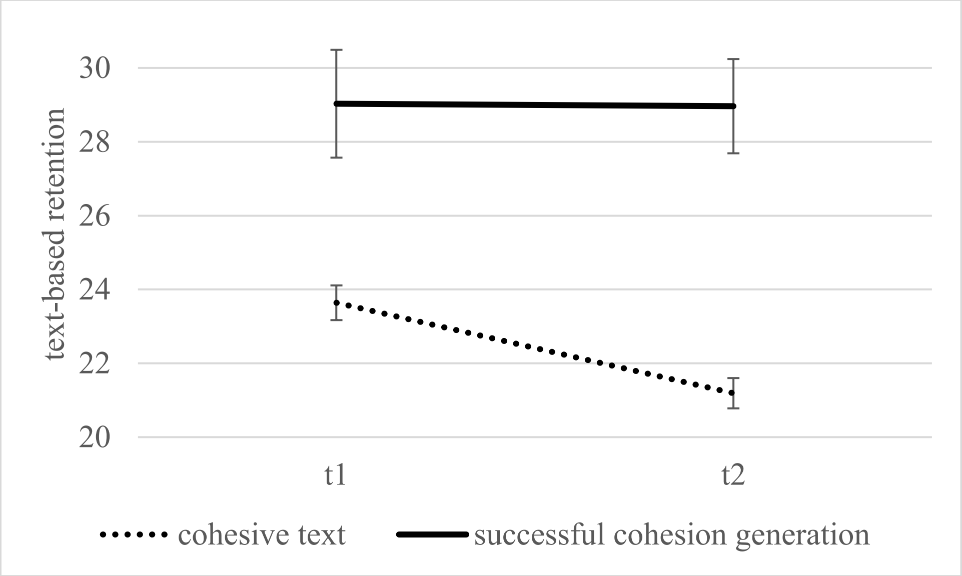


*Figure 4.* Text-based representation as a function of condition and delay when controlling for reading skills (estimated means and standard errors). In contrast to Figure 1, only learners who successfully generated (+1 *SD*) were analyzed. Max. performance was 38.

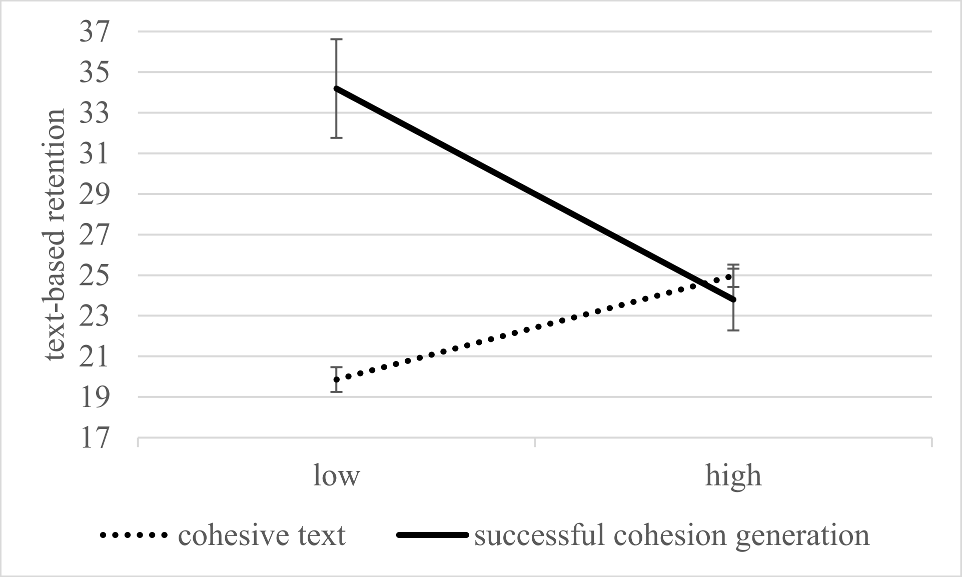


*Figure 5.* Text-based representation (collapsed across immediate and delayed testing) as a function of condition and the level of reading skill (estimated means and standard errors for -1 *SD* and +1 *SD*). Only learners who successfully generated (+1 *SD*) were analyzed. Max. performance was 38.

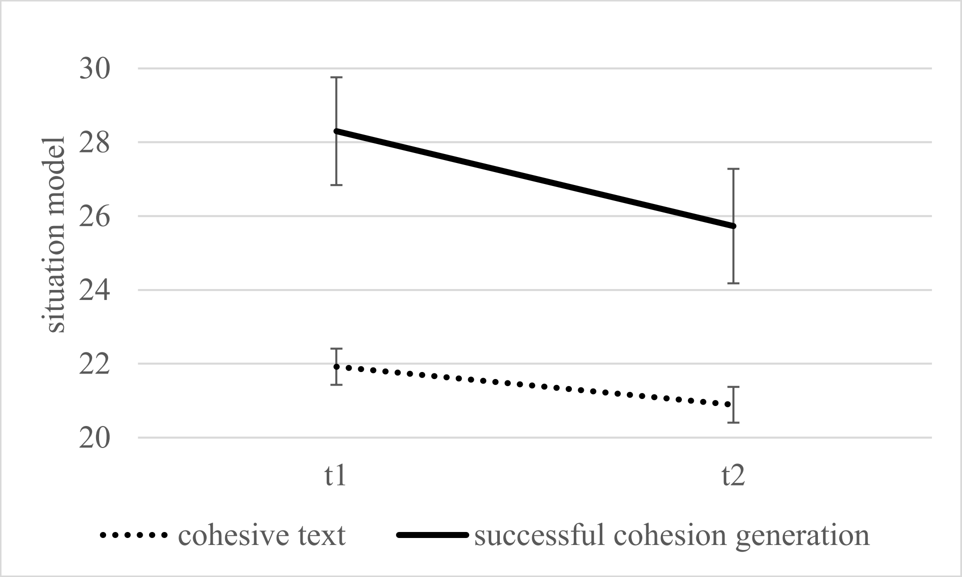


*Figure 6.* Situation model as a function of condition and delay when controlling for reading skills (estimated means and standard errors). In contrast to Figure 2, only learners who successfully generated (+1 *SD*) were analyzed. Max. performance was 42.

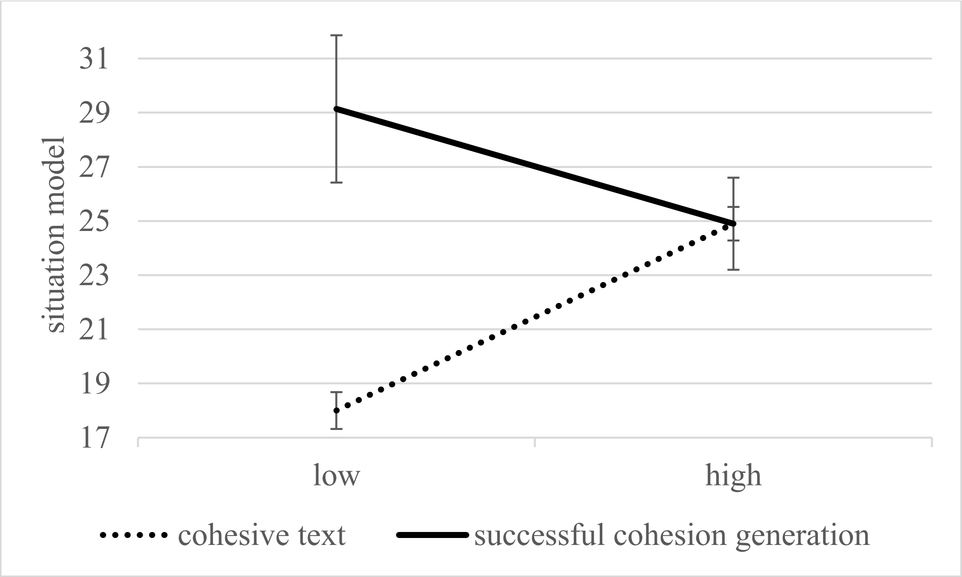


*Figure 7.* The situation model (collapsed across immediate and delayed testing) as a function of condition and the level of reading skill (estimated means and standard errors for -1 *SD* and +1 *SD*). In contrast to Figure 3, only learners who successfully generated (+1 *SD*) were analyzed. Max. performance was 42.

# Experiment 2b

A version of this article is in preparation as:

Abel, R., & Hänze, M. (in preparation). Who benefits from reading non-cohesive expository texts? The role of learners' proficiencies in detecting and bridging the cohesion gaps.

For the present experiment, we largely replicated the design of Experiment 2a, but incorporated a third condition, namely reading a non-cohesive text. The previous Experiment 2a investigated how a cohesion generation task vs. reading a fully cohesive text interact with learners' proficiencies. The present experiment was carried out to address two further research questions.

The first research question: How does *cohesion* (a non-cohesive text vs. a fully cohesive text) interact with learners' proficiencies, previous knowledge and reading skill? Previous research on text comprehension indicates that previous knowledge is necessary to close cohesion gaps, resulting in an interaction of cohesion and previous knowledge: Low knowledge readers require a fully cohesive text, whereas high knowledge readers benefit from reading a non-cohesive text (McNamara et al., 1996; McNamara & Kintsch, 1996). Thereby, reading skill is required to deal with higher processing demands imposed by a fully cohesive text due to a higher number of explicit relations (O'Reilly & McNamara, 2007; Ozuru et al., 2009; Voss & Silfies, 1996). Taken together, the prediction by previous research is the following: Reading a non-cohesive text should be beneficial for high knowledge readers, irrespective of their reading skill, whereas reading a fully cohesive text should be beneficial for low knowledge/high skilled readers.

Considering the pattern of results yielded in Experiment 2a, this prediction, however, appears controversial. Poor readers who accurately closed the gaps in the generation condition achieved higher learning scores and less forgetting than readers of a fully cohesive text. We interpret this pattern of results in the following way. Cohesion gaps promote coherence construction if readers are able to both recognize and close them. The cohesion generation implementation highlights the cohesion gaps and only requires learners to close them. Poor readers may struggle with detecting implicit cohesion gaps and therefore benefit from the cohesion generation tool if they succeed to close the gaps.

In view of that, we counter the prediction by previous research with our own (original) idea that reading skill might be helpful in detecting cohesion gaps. Assumed that reading skill supports learners in detecting the cohesion gaps and previous knowledge supports them in closing the gaps, we would expect that only high knowledge/high skilled readers benefit from reading a non-cohesive text. A fully cohesive text in contrast should support readers who lack at least one of these two proficiencies. One more note: Our prediction and the prediction by previous research overlap regarding the assumption that reading a fully cohesive text should be beneficial for low knowledge/high skilled readers.

We assume that poor readers largely do not notice the absence of causal relations in the non-cohesive text. Consequently, they are not expected to be engaged in processes of drawing inferences to close the cohesion gaps, which should be indicated by equal reaction times in the dual task (for objective measuring the cognitive load on poor readers' working memory) while reading a non-cohesive text compared to reading a fully cohesive text. Poor readers could experience even less cognitive load on working memory while reading a non-cohesive text because a non-cohesive text does not explicitly require readers to process interconnected relations between sentences. High skilled readers in contrast should notice the cohesion gaps and be more engaged than their counterparts who read a fully cohesive text, which should be indicated by lower reaction times in the dual task.

The second research question: How does *highlighting of cohesion gaps* (a non-cohesive text vs. a cohesion generation) interact with learners' proficiencies, previous knowledge and reading skill? By clarifying the second research question, we would be able to decompose the reasons for why some readers take no learning advantage of reading a non-cohesive text: Are they not capable of closing the gaps or do they not notice the cohesion gaps in the first place? The non-cohesive text condition shares with the cohesion generation condition the cohesion gaps, but differs only in whether the gaps are highlighted or subtle.

If poor readers struggle with noticing the cohesion gaps, they should benefit from generating cohesion, irrespective of their generation accuracy (because different from reading a fully cohesive text, reading a non-cohesive text provides no instructional support via cohesion devices anyway). For skilled readers there should be no difference or even a benefit from reading a non-cohesive text over cohesion generation because skilled readers might experience redundancy and interference due to interruptions in spontaneous activity.

The article presenting Experiment 2b is currently under preparation. We will briefly report the changes to the initial design of Experiment 2a and statistical analyses accompanied by a short discussion of the results.

## Method

### *Participants*

113 students (undergraduate psychologists and ongoing teachers) from the University of Kassel (Germany) were randomly distributed across three conditions: 42 participants read the fully cohesive text, 35 read a non-cohesive text, and 36 generated cohesion while reading. 87 participants were female. In average, the age of the students was 22.99 (*SD* = 3.45). 24 participants reported having another mother tongue besides German. Participants were rewarded with either 15€ or 1.5 credits.

### *Design and Materials*

In the present experiment, we largely replicated the design of Experiment 2a. We will thus list only the changes to Experiment 2a. We tested college students under controlled conditions in the laboratory (while Experiment 2a was carried out in a classroom with high school students). The retention interval for the follow-up testing was shortened to one week due to pragmatic reasons.

The main difference is a third condition, in which participants read a non-cohesive text. Similar to the text version in the cohesion generation condition, a non-cohesive text lacked any causal connectives, but different is that a non-cohesive text lacked also any indication of the absence of causal connectives – that is, the cohesion gaps in the non-cohesive text were implicit. The removal of 57 causal connectives inherently led to disconnection of the clauses that were connected within the fully cohesive text. For example, the fully cohesive text says, "*Nature offers a variety of rough surfaces, therefore the mixed reflection is the most common case*". While in the non-cohesive text, it says, "*Nature offers a variety of rough surfaces. The mixed reflection is the most common case*". Consequently, the non-cohesive text contained more sentences (174 vs. 124 in the fully cohesive text), but fewer words (2054 vs. 2089 in the fully cohesive text). The quantitative difference in length between the fully cohesive text and the non-cohesive text can be considered minor against the background of previous studies using full-length expository texts and manipulating cohesion (Ainsworth & Burcham, 2007; McNamara et al., 1996; Ozuru et al., 2009). Thus, our study overcomes a limitation of previous studies that confounded cohesion and text length.

The present experiment was also supposed to overcome a limitation due to the implementation of cohesion generation in Experiment 2a. In Experiment 2a, readers could conclude on the direction of connectives (*forward* vs. *backward*) simply based on syntax (i.e., the position of the verb in the second clause). According to German syntax, the verb in the second clause must be placed subsequent to a connective to indicate the forward direction – from the cause to the consequence – by using the connectives *deswegen* (*therefore*) or *dennoch* (*however*). To indicate the backward direction (from the consequence to the cause) by using the connectives *weil* (*because*) or *obwohl* (*although*), the verb in the second clause must be placed at the end of the sentence.

The implementation of the cohesion generation in the present experiment prevented participants from making syntactically driven conclusions. We placed the particular verb in the second clause of each critical sentence on both theoretically possible positions (subsequent to the connective *and* at the end of the sentence): "Der Anteil der Strahlen, der reflektiert wird, verliert keine Energie, __________ *behält* er die kurzwellige Form *behält*". The particular verb on both positions was italicized. Bevor reading the expository text, participants in the generation condition were instructed on the meaning of verb duplication.

## Results

The data is publically available at https://osf.io/2f8us/.

### *Learners' Proficiencies*

Separate ANOVAs with the condition (fully cohesive text vs. non-cohesive text vs. cohesion generation) as a between-subjects factor showed no difference regarding learners' previous knowledge, $F(2,110) = 1.06, p = .351, \eta^2 = .02$, and reading skill, $F < 1$. Both proficiencies correlated positively, $r = .37, p < .001$.

### *Learning Processes*

Means and standard deviations in the time-on task, cognitive load measures, and the generation accuracy are reported in Table 1.

**Time-on Task.** We computed an ANCOVA with the condition (fully cohesive text vs. non-cohesive text vs. cohesion generation) as a between-subjects factor. We included the z-standardized scores of the previous knowledge and reading skill as covariates in the analysis to control for the effects of learners' proficiencies in spontaneous detecting and closing the cohesion gaps.

Condition had a significant impact on time-on task, $F(2,108) = 56.06, p < .001, \eta^2 = .51$. Simple comparisons showed that cohesion generation took longer than reading a fully cohesive text, *MD*[10] = 9.07, *SE* = .91, $p < .001$, 95% *CI* [6.85, 11.29], and a non-cohesive text, *MD* = 7.97, *SE* = .97, $p < .001$, 95% *CI* [5.62, 10.32]. Reading a fully cohesive text and a non-cohesive text took the same amount of time, *MD* = -1.10, *SE* = .93, *p* = .709, 95% *CI* [-3.35, 1.15].

Considering the covariates, previous knowledge had no main effect, $F(1,108) = 2.50$, *p* = .117, $\eta^2 = .02$, whereas reading skill significantly reduced the time-on task, $F(1,108) = 4.57$, *p* = .035, $\eta^2 = .04$.

**Cognitive Load via Self-Report.** We computed a MANCOVA with the condition (fully cohesive text vs. non-cohesive text vs. cohesion generation) as a between-subjects factor for the retrospective cognitive load measures. We included the z-standardized scores of the previous knowledge and reading skill as covariates in the analysis to control for the effects of learners' proficiencies in spontaneous detecting and closing the cohesion gaps.

Condition had a significant impact on cognitive load measures, $F(3,107) = 3.67$, *p* = .015, $\eta^2 = .09$. Considering the covariates, previous knowledge had a main effect on cognitive load measures, $F(3,106) = 3.73$, *p* = .014, $\eta^2 = .10$, whereas reading skill had no effect, $F < 1$. In the following, we will thus report only the effects of condition and previous knowledge.

***Intrinsic Load.*** Condition had a significant impact on intrinsic cognitive load, $F(2,108) = 3.62$, *p* = .030, $\eta^2 = .06$. Simple comparisons showed no difference between reading a fully cohesive text and a non-cohesive text, *MD* = .70, *SE* = .52, *p* = .544, 95% *CI* [-.57, 1.98], and between reading a fully cohesive text and cohesion generation, *MD* = -.77, *SE* = .52, *p* = .422, 95% *CI* [-2.02, .49]. Readers who generated cohesion experienced a higher intrinsic cognitive load than readers of a non-cohesive text, *MD* = 1.47, *SE* = .55, *p* = .025, 95% *CI* [.14, 2.80]. Previous knowledge decreased the intrinsic cognitive load, $F(1,108) = 5.95$, *p* = .016, $\eta^2 = .05$.

[10] *MD* stays for mean difference

***Extraneous Load.*** Neither condition, $F < 1$, nor previous knowledge had any impact on extraneous cognitive load, $F(1,108) = 2.19$, $p = .142$, $\eta^2 = .02$.

***Germane Load.*** Condition had no impact on germane cognitive load, $F(2,108) = 2.57$, $p = .081$, $\eta^2 = .05$. Previous knowledge, however, increased the germane cognitive load, $F(1,108) = 6.65$, $p = .011$, $\eta^2 = .06$.

**Cognitive Load via Dual Task.** We computed an ANCOVA with the condition (fully cohesive text vs. non-cohesive text vs. cohesion generation) as a between-subjects factor for the reaction times and the response accuracy in the dual task respectively. We included the z-standardized scores of the previous knowledge and reading skill as covariates in the analyses to control for the effects of learners' proficiencies in spontaneous detecting and closing the cohesion gaps, and importantly, to investigate their moderating impact.

***Reaction Times.*** Condition had no impact on the reaction times, $F < 1$. Considering the covariates, previous knowledge had no main effect, $F < 1$, whereas reading skill significantly reduced the reaction times, $F(1,101) = 5.30$, $p = .023$, $\eta^2 = .05$. Neither previous knowledge significantly interacted with the condition, $F(2,101) = 1.15$, $p = .322$, $\eta^2 = .02$, nor reading skill does, $F < 1$. The latter result is inconsistent with our expectation that skilled readers should show lower reaction times when reading a non-cohesive text compared to reading a fully cohesive text because they may notice the cohesion gaps, whereas poor readers should show equal (or higher) reaction times because they may not. There was no interaction between both covariates, $F < 1$. Finally, the three-way interaction was also not significant, $F < 1$.

***Response Accuracy.*** Condition had no impact on the response accuracy in the dual task, $F(2,101) = 1.19$, $p = .309$, $\eta^2 = .02$. Both covariates had no main effect, *Fs* < 1. Neither previous knowledge significantly interacted with the condition, $F(2,101) = 1.17$, $p = .314$, $\eta^2$

= .02, nor reading skill does, $F < 1$. There was no interaction between both covariates, $F < 1$. Finally, the three-way interaction was also not significant, $F < 1$.

**Generation Accuracy.** Students, on average, chose the correct connective in seven of ten sentences ($M_{\%\ \text{accuracy}} = 69.49$; $SD = 12.62$). The generation accuracy differed depending on the connector type: *because* (76%), *therefore* (68.8%), *although* (63.1%), and *however* (56.1%). The repeated measures ANOVA with two within-subjects factors, *polarity* (positive vs. negative) and *direction* (backward vs. forward), revealed a higher difficulty of connectives with the negative polarity (*although* and *however*), $F(1,53) = 4.24$, $p = .044$, $\eta^2 = .07$, but neither a main effect of the direction, $F(1,53) = 1.29$, $p = .261$, $\eta^2 = .02$, nor an interaction effect, $F(1,53) = 0.00$, $p = .994$, $\eta^2 = .00$.

A direct comparison with the averaged generation accuracy in Experiment 2a ($M_{\%\ \text{accuracy}} = 73.97$; $SD = 15.81$) yielded a small effect of Cohen's $d = .31$. We explain the difference in the difficulty with lower generating demands in Experiment 2a due to the possibility to make syntactically driven conclusions, which was undermined in the present experiment.

In the next step, we computed correlations between the generation accuracy and learning outcomes to investigate the extent that the generation accuracy affects learning outcomes. Generation accuracy respectively correlated with the immediate text-based representation, $r = .73$, $p < .001$, delayed text-based representation, $r = .45$, $p = .006$, immediate situation model, $r = .57$, $p < .001$, and delayed situation model, $r = .64$, $p < .001$.

Finally, we analyzed the dependency of generation accuracy on learners' proficiencies. We computed an OLS linear regression with prior knowledge and reading skill as predictor variables. Generation accuracy was the criterion variable. The model explained 43.6% of the variance, $F(2,33) = 12.74$, $p < .001$. Prior knowledge, $\beta = .46$, $t(33) = 3.45$, $p = .002$, and reading skill, $\beta = .41$, $t(33) = 3.08$, $p = .004$, were both predictive for the generation accuracy.

We assume that the previous knowledge supports readers in making elaborative inferences, and the reading skill in making bridging inferences to close the cohesion gaps.

Table 1

*Mean Scores and Standard Deviations of Learning Processes*

| | Fully cohesive | | Non-cohesive | | Generation | |
|---|---|---|---|---|---|---|
| Measure | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Time-on task (in min.) | 12.57 | 3.58 | 13.44 | 2.93 | 21.81 | 5.33 |
| Intrinsic CL | 5.58 | 2.29 | 5.07 | 2.20 | 6.29 | 2.41 |
| Extraneous CL | 2.33 | 1.83 | 2.40 | 1.54 | 2.70 | 1.65 |
| Germane CL | 7.21 | 2.24 | 7.38 | 1.75 | 6.46 | 2.24 |
| Dual task reaction time (mean in ms) | 1906 | 459 | 1943 | 462 | 2039 | 468 |
| Dual task response accuracy (in %) | 94.96 | 66.97 | 93.95 | 62.98 | 96.57 | 47.43 |
| Generation accuracy (in %) | — | — | — | — | 69.49 | 12.62 |

### ***Learning Outcomes***

A repeated measures ANCOVA with the condition (fully cohesive text vs. non-cohesive text vs. cohesion generation) as a between-subjects factor and the retention interval (immediate vs. one-week delay) as a within-subjects factor was computed for the text-based representation and the situation model respectively. We included the z-standardized scores of the previous knowledge and reading skill as covariates in the analyses to control for the effects of learners' proficiencies in spontaneous detecting and closing the cohesion gaps, and to investigate their moderating impact.

**Text-based Representation.** Condition had a significant impact on the text-based representation, $F(2,101) = 4.07$, $p = .020$, $\eta^2 = .08$. Simple comparisons revealed the superiority of reading a fully cohesive text over generating cohesion, $MD = .07$, $SE = .02$, $p = .026$, 95% $CI$ [.01, .13], but neither a difference between reading a fully cohesive text and a non-cohesive text, *MD* = .01, *SE* = .02, *p* = 1.00, 95% *CI* [-.05, .08], nor between reading a non-cohesive text and cohesion generation, *MD* = .05, *SE* = .02, *p* = .113, 95% *CI* [-.01, .11].

Learners performed worse in the delayed test, *F*(2,101) = 34.06, *p* < .001, $\eta^2$ = .25, *MD* = .06, *SE* = .01. An interaction between the condition and retention interval was found, *F*(2,101) = 6.43, *p* = .002, $\eta^2$ = .11. Figure 1 displays the text-based representation as a function of condition and retention interval. Simple comparisons revealed significant forgetting rates for readers of a fully cohesive text, *MD* = .08, *SE* = .02, *p* < .001, 95% *CI* [.05, .12], and a non-cohesive text, *MD* = .08, *SE* = .02, *p* < .001, 95% *CI* [.05, .12]. However, there was no performance difference between the immediate and delayed testing in the generation condition, *MD* = .01, *SE* = .02, *p* = .578, 95% *CI* [-.02, .04]. Thus, in both reading conditions as compared to cohesion generation, learning was not sustainable. Different forgetting rates resulted in an inferiority of cohesion generation to reading a fully cohesive text, *MD* = -.10, *SE* = .03, *p* = .001, 95% *CI* [-.17, -.04], and a non-cohesive text, *MD* = -.09, *SE* = .03, *p* = .005, 95% *CI* [-.15, -.02], only at immediate testing. One week later, there was no performance difference to reading a fully cohesive text, *MD* = -.03, *SE* = .03, *p* = .921, 95% *CI* [-.1, .04], and a non-cohesive text, *MD* = -.02, *SE* = .03, *p* = 1.00, 95% *CI* [-.08, .05]. Reading a fully cohesive text and a non-cohesive text resulted in an equal performance at immediate, *MD* = .02, *SE* = .03, *p* = 1.00, 95% *CI* [-.05, .08], and delayed test, *MD* = .01, *SE* = .03, *p* = 1.00, 95% *CI* [-.06, .08]. Altogether, the main effects of condition and retention interval can be ascribed to this interaction pattern, which replicates the findings from Experiment 2a regarding the lower forgetting rate in the generation condition.

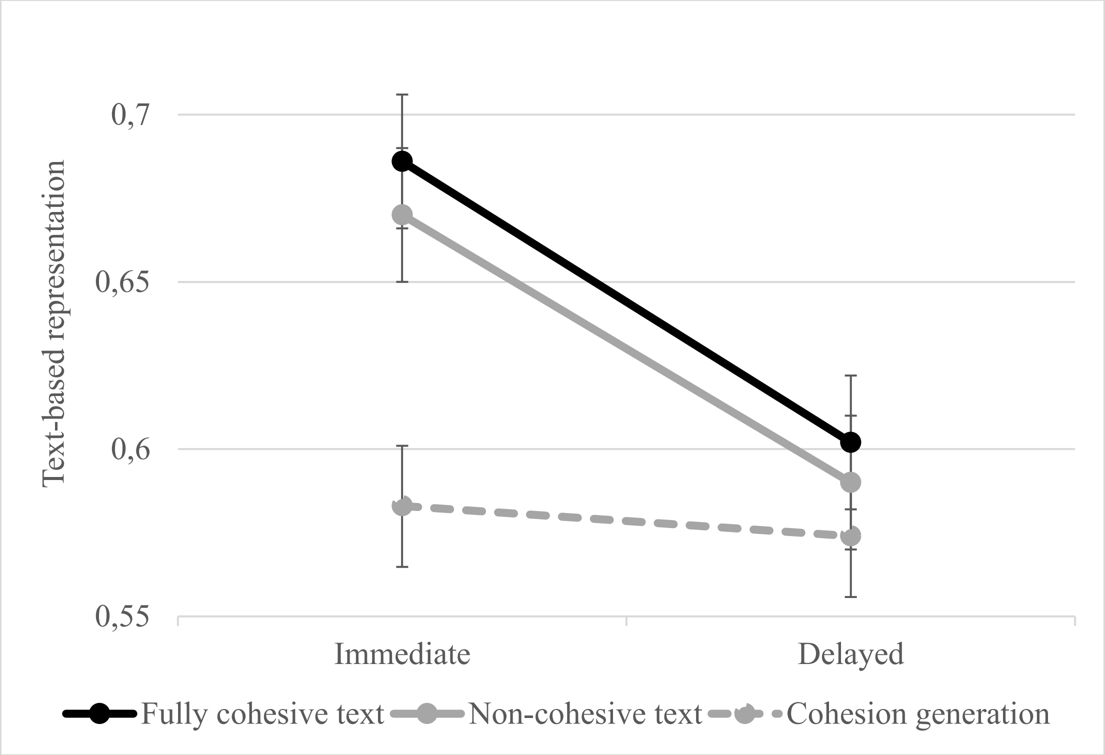


*Figure 1.* Text based representation (relative proportion between 0-1) as a function of condition and retention interval when controlling for previous knowledge and reading skill. Estimated means and standard errors are depicted.

Considering the covariates, previous knowledge significantly affected the text-based representation, $F(1,101) = 36.17$, $p < .001$, $\eta^2 = .26$, whereas the reading skill had no main effect, $F(1,101) = 1.40$, $p = .240$, $\eta^2 = .01$. Neither previous knowledge significantly interacted with the condition, $F(2,101) = 2.95$, $p = .057$, $\eta^2 = .06$, nor reading skill does, $F(2,101) = 2.01$, $p = .139$, $\eta^2 = .04$. Neither of both covariates interacted with the retention interval, $F < 1$. There was also no interaction between both covariates, $F(1,101) = 1.06$, $p = .306$, $\eta^2 = .01$.

We found no significant three-way interaction involving condition and retention interval, neither with previous knowledge, $F(2,101) = 2.86$, $p = .062$, $\eta^2 = .05$, nor with reading skill, $F < 1$. Previous knowledge, reading skill, and retention interval also did not interact, $F(1,101) = 1.80$, $p = .183$, $\eta^2 = .02$.

Importantly, we found a three-way interaction of previous knowledge, reading skill, and condition, *F*(2,101) = 4.53, *p* = .013, $\eta^2$ = .08. To decompose this three-way interaction, we computed second-order simple effects of condition and simple comparisons across the conditions at several level combinations of previous knowledge (-1 *SD* vs. +1 *SD*) and reading skill (-1 *SD* vs. +1 *SD*). Figure 2 displays the estimates for text-based questions (collapsed across immediate and delayed testing) for high vs. low skilled readers at the low level of previous knowledge and Figure 3 at the high level of previous knowledge. At the low level of previous knowledge (-1 *SD*), conditions did not differ irrespective of the level of reading skill, *Fs* < 1. At the high level of previous knowledge (+1 *SD*), however, conditions differed depending on the level of reading skill.

There was a second-order simple effect of condition at the level combination high previous knowledge (+1 *SD*) / low reading skill (-1 *SD*), *F*(2,101) = 8.36, *p* < .001, $\eta^2$ = .14. Simple comparisons showed that reading a fully cohesive text outperformed both reading a non-cohesive text, *MD* = .15, *SE* = .06, *p* = .041, 95% *CI* [.01, .30], and cohesion generation, *MD* = .22, *SE* = .05, *p* < .001, 95% *CI* [.09, .35], while reading a non-cohesive text and cohesion generation performed equally, *MD* = .07, *SE* = .06, *p* = .683, 95% *CI* [-.07, .21].

The second-order simple effect of condition was also significant at the level combination high previous knowledge (+1 *SD*) / high reading skill (+1 *SD*), *F*(2,101) = 3.35, *p* = .039, $\eta^2$ = .06. Simple comparisons failed to reveal significant differences between reading a fully cohesive text and a non-cohesive text, *MD* = -.09, *SE* = .04, *p* = .109, 95% *CI* [-.18, .01], reading a fully cohesive and cohesion generation, *MD* = .03, *SE* = .04, *p* = 1.00, 95% *CI* [-.07, .12], and also reading a non-cohesive text and cohesion generation, *MD* = .11, *SE* = .05, *p* = .050, 95% *CI* [-.00, .22]. However, to avoid the beta error, we explored the impact of condition at the same level combination – high previous knowledge / high reading skill – by using a higher *SD* for reading skill (+1.5 *SD*), yielding again a significant second-order simple effect of condition *F*(2,101) = 4.12, *p* = .019, $\eta^2$ = .08. Simple comparisons now showed that reading a fully cohesive text was inferior to reading a non-cohesive text, *MD* = -.15, *SE* = .05, *p* = .019, 95% *CI* [-.27, -.02], and equal to cohesion generation, *MD* = -.02, *SE* = .05, *p* = 1.00, 95% *CI* [-.15, .11], while reading a non-cohesive text and cohesion generation performed also equally, *MD* = .12, *SE* = .06, *p* = .119, 95% *CI* [-.02, .26].

Finally, we found no four-way interaction, *F* < 1.

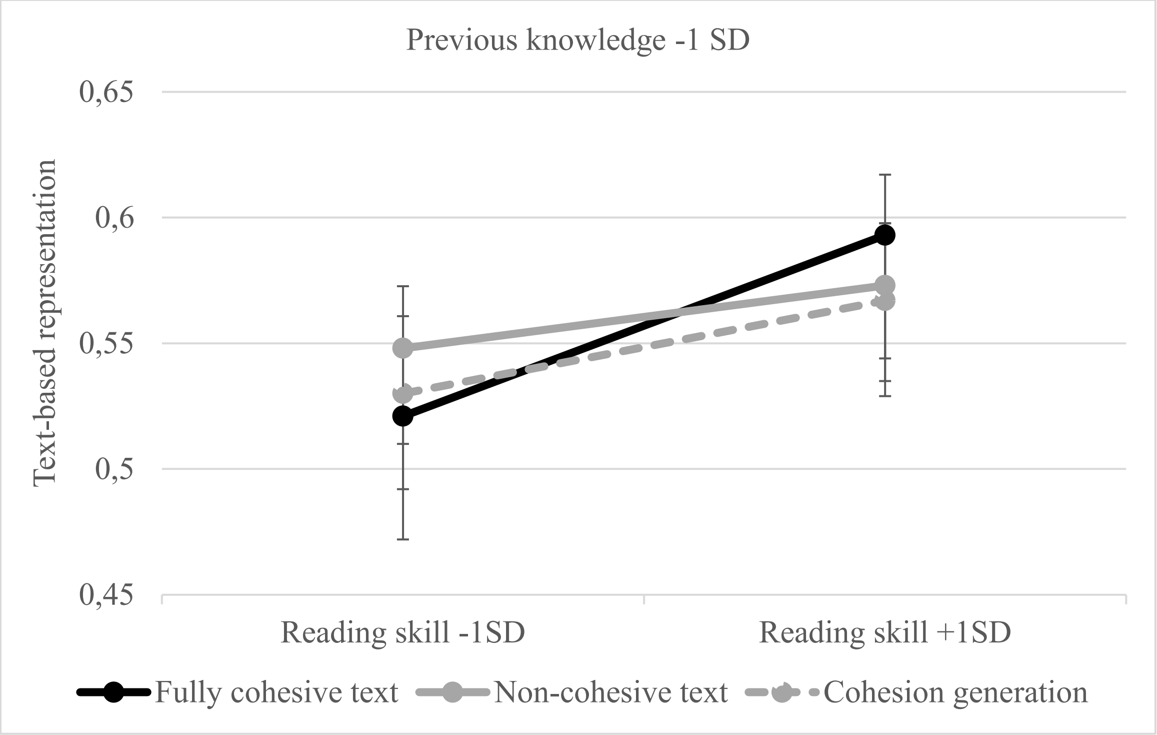


*Figure 2.* Text based representation (collapsed across immediate and delayed testing; relative proportion between 0-1) as a function of condition and reading skill for low knowledge learners (-1 *SD*). Estimated means and standard errors are depicted.

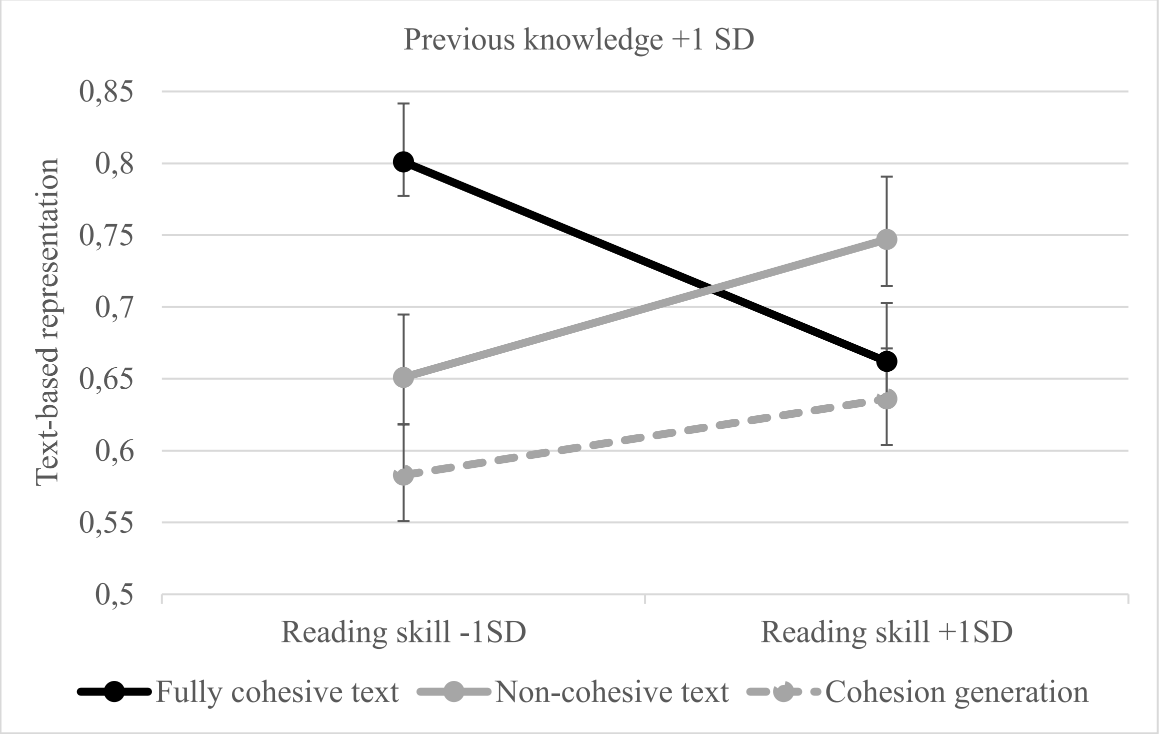


*Figure 3.* Text based representation (collapsed across immediate and delayed testing; relative proportion between 0-1) as a function of condition and reading skill for high knowledge learners (+1 *SD*). Estimated means and standard errors are depicted.

**Situation Model.** Condition had a significant impact on situation model construction, *F*(2,101) = 5.00, *p* = .009, η$^2$ = .09. Simple comparisons revealed the superiority of reading a non-cohesive text over generating cohesion, *MD* = .08, *SE* = .03, *p* = .011, 95% *CI* [.01, .14], but neither a difference between reading a fully cohesive text and a non-cohesive text, *MD* = -.02, *SE* = .03, *p* = 1.00, 95% *CI* [-.08, .05], nor between reading a fully cohesive text and cohesion generation, *MD* = .06, *SE* = .03, *p* = .067, 95% *CI* [-.00, .12].

Learners performed worse in the delayed test, *F*(1,101) = 9.64, *p* = .002, η$^2$ = .09, *MD* = .03, *SE* = .01. We found no interaction between condition and retention interval, *F* < 1. Figure 4 displays the situation model construction as a function of condition and retention interval.

Considering the covariates, previous knowledge significantly affected the situation model construction, *F*(1,101) = 67.50, *p* < .001, η$^2$ = .40, and also did the reading skill, *F*(1,101) = 5.86, *p* = .017, η$^2$ = .06. Neither previous knowledge significantly interacted with the condition, *F*(2,101) = 2.06, *p* = .133, η$^2$ = .04, nor reading skill did, *F*(2,101) = 2.77, *p* = .068, η$^2$ = .05. Neither of both covariates interacted with the retention interval, *F* < 1. There was also no interaction between both covariates, *F*(1,101) = 2.42, *p* = .123, η$^2$ = .02.

We found no significant three-way interaction involving condition and retention interval, neither with previous knowledge, *F*(2,101) = 1.20, *p* = .306, η$^2$ = .02, nor with reading skill, *F* < 1. Previous knowledge, reading skill, and retention interval did not interact, *F*(1,101) = 1.61, *p* = .208, η$^2$ = .02. The three-way interaction of previous knowledge, reading skill, and condition was also not significant, *F*(2,101) = 1.05, *p* = .354, η$^2$ = .02. Finally, we found no four-way interaction, *F*(2,101) = 1.35, *p* = .265, η$^2$ = .03.

Taken together, the results on situation model construction revealed the main effect of condition, retention interval, and both covariates, previous knowledge and reading skill, respectively. Importantly, especially readers of a non-cohesive text outperformed their counterparts who generated cohesion.

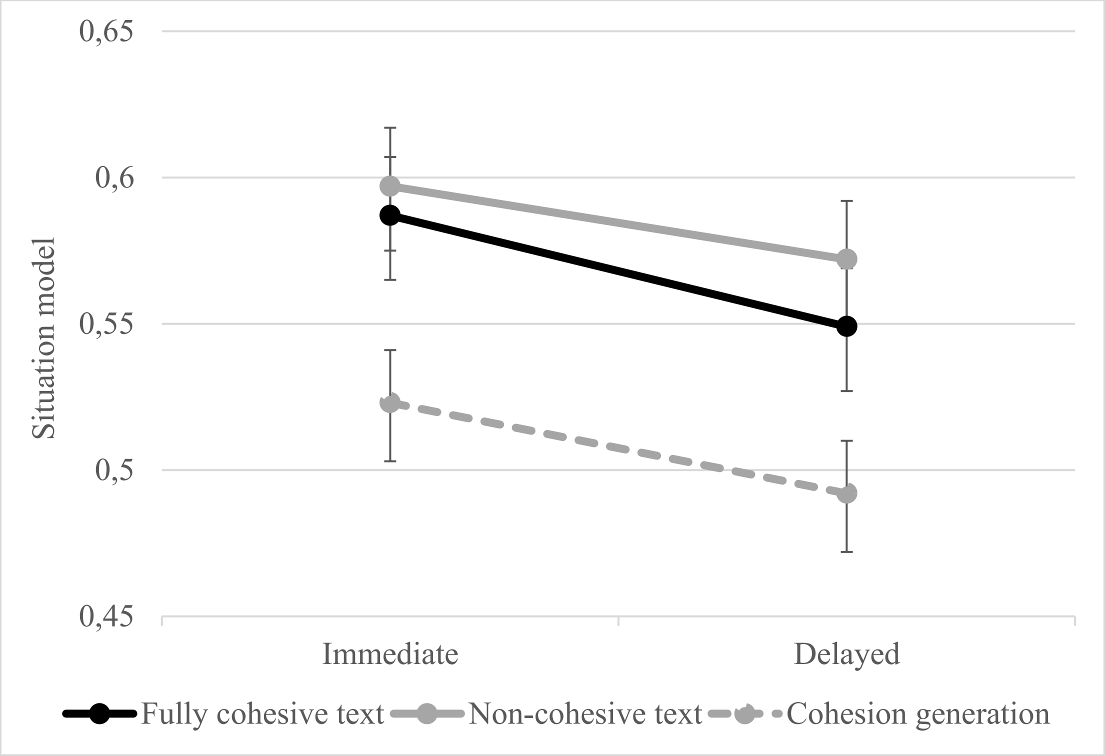


*Figure 4.* Situation model (relative proportion between 0-1) as a function of condition and retention interval when controlling for previous knowledge and reading skill. Estimated means and standard errors are depicted.

**Discussion**

The present experiment investigated the impact of reading a non-cohesive text as compared to both conditions used in Experiment 2a – reading a fully cohesive text and cohesion generation – depending on learners' proficiencies, previous knowledge and reading skill. Our hypotheses based on an original assumption that reading skill might be helpful in detecting the cohesion gaps while reading a non-cohesive text. Thus – different from skilled readers – poor readers largely might not notice the cohesion gaps. Accordingly, we expected poor readers to benefit from reading a fully cohesive text (which provides instructional support via cohesion devices) and cohesion generation (which highlights the cohesion gaps) relative to reading a non-cohesive text. Skilled readers in contrast should take advantage of reading a non-cohesive text (relative to reading a fully cohesive text) if they have enough previous knowledge to close the cohesion gaps. In the following, we will discuss the results of the present study with regard to two research questions.

### *How Does Cohesion (Non-cohesive Text vs. Fully Cohesive Text) Interact with Learners' Proficiencies?*

Taken together, our results partially confirmed our expectations, but there are also inconsistencies to report. Considering the learning outcomes in terms of the text-based representation, the present experiment demonstrated the superiority of reading a non-cohesive text over a fully cohesive text for high skilled readers with a high level of previous knowledge. This finding speaks in favor of our assumption that cohesion gaps engage readers in relational processing if they are able to detect (high reading skill) and close the cohesion gaps (high previous knowledge). Reading a fully cohesive text in contrast was superior to reading a non-cohesive text for poor readers with a high level of previous knowledge. We explain this finding with the reference to our assumption that poor readers fail to detect cohesion gaps, but a high level of previous knowledge might be supportive in processing an additional number of *explicit* relations across information units (Sweller, 2010). In accordance with the latter interpretation, we found that previous knowledge reduces the intrinsic cognitive load (and increases the germane cognitive load); and we found no impact of reading skill on intrinsic cognitive load as would have been expected by the assumption of previous research that reading skill helps in dealing with a high number of interconnections in text.

The pattern of results on text-based representation is at odds with the predictions by previous research on text comprehension, which predicts a benefit of reading a fully cohesive text only for skilled readers with a low level of previous knowledge. According to this view, cohesion is supposed to compensate for a low level of previous knowledge, whereas a high level of reading skill is considered necessary to process a high number of *explicit* relations across information units due to cohesion (O'Reilly & McNamara, 2007; Ozuru et al., 2009; Voss & Silfies, 1996). Reading a non-cohesive text in contrast should be advantageous for readers with a high level of previous knowledge, irrespective of their reading skill.

The pattern of results on text-based representation speaks thus in favor of our assumption that reading skill supports learners in detecting the *implicit* relations in text (=cohesion gaps) rather than dealing with processing demands imposed by *explicit* relations (which previous knowledge may do). However, the comparison of reading a non-cohesive text with reading a fully cohesive text regarding the learning outcomes further revealed less consistent results with either view.

We predicted that reading a non-cohesive text should benefit only skilled readers with a high level of previous knowledge because both is necessary, detecting and closing the gaps. Reading a fully cohesive text in contrast should benefit readers who lack at least one of these two proficiencies, particularly skilled readers with a low level of previous knowledge (which is predicted also by previous research on text comprehension, but due to another reasons). However, low knowledge learners (irrespective of their reading skill) took no advantage of reading a fully cohesive text, neither in terms of the text-based representation, nor in terms of the situation model – which is inconsistent with either prediction. The situation model measures revealed no differences between reading a fully cohesive and a non-cohesive text, irrespective of learners' proficiencies, indicating no impact of cohesion/gaps. This is an inconvenient result since the cohesion/gaps manipulation taps essential relations in text that are assessed by questions on situation model.

The comparison of reading a non-cohesive text with reading a fully cohesive text regarding the cognitive load assessment via a dual task also revealed no support for our assumption. Our hypothesis assumed that poor readers easily overlook the cohesion gaps while skilled readers are engaged in relational processes when facing cohesion gaps. This should be indicated by equal (or higher) reaction times for poor readers and lower reaction times for skilled readers while reading a non-cohesive text compared to reading a fully cohesive text. However, the objective cognitive load measure did not provide any insights into gaps detection processes while reading. In fact, the only significant effect was the main effect of reading skill (longer reaction times for poor readers), indicating processing struggles of poor readers compared to skilled readers. For future research, it is thus necessary to use further assessment tools to track the processes of detecting and closing the cohesion gaps. For example, tracking the eye fixation patterns such as lookbacks seems to be a promising way.

### *How Does Highlighting of Cohesion Gaps (Non-cohesive Text vs. Cohesion Generation) Interact with Learners' Proficiencies?*

Taken together, our results did not confirm our expectations. We expected a moderating effect of reading skill on learning from a non-cohesive text vs. cohesion generation in terms of text-based representation and situation model. Particularly, we expected that poor readers benefit from generation cohesion, which highlights the cohesion gaps, whereas skilled readers may spontaneously detect the cohesion gaps and thus do not require such an aid.

However, considering the text-based representation, we found no level combination of previous knowledge and reading skill, at which these two conditions significantly differed. Particularly, the inferiority of cohesion generation to reading a fully cohesive text for poor readers with a high level of previous knowledge is clearly inconsistent with our reasoning since the cohesion generation should compensate for a low reading skill and be surmountable due to a high level of previous knowledge. Considering the situation model, reading a non-cohesive text was superior to cohesion generation, irrespective of either learners' proficiencies.

The cohesion generation task in the main was probably too difficult for learners, which is indicated by the generation accuracy of 70% and longer reaction times in the dual task. Leaners also perceived the text in the cohesion generation condition as more difficult than the non-cohesive text, which is indicated by a higher intrinsic cognitive load in the generation condition: Learners erroneously attributed the presence of gaps to a higher number of relations in text (and analogously the absence of explicit gaps to a lower number of relations). It is also reasonable to assume that the explicit gaps in the cohesion generation condition narrowed learners' attentional focus to solely *local* relations, whereas the readers of a non-cohesive text may have had a broader attentional focus and detected *global* relations.

### ***Limitations***

One important limitation of the present study should be noted: By using a sample of 113 participants, the statistical power for the analysis of a three-way-interaction (of condition, previous knowledge, and reading skill) was low. Some comparable studies investigating the contribution of previous knowledge and reading skill in learning with fully cohesive vs. non-cohesive expository texts used a larger sample, which, however, should be also considered too small for analyzing three-way-interactions. For example, the sample of Ozuru et al. (2009) consisted of 170 participants, and O'Reilly and McNamara (2007) examined 143 participants. Voss and Silfies (1996) reported using substantially less participants (only 40). We actually found a medium effect of $\eta^2 = .08$ with respect to the text-based representation, which should be replicated in follow-up investigations with substantially larger samples. This is also necessary against the background of hypotheses that could not be confirmed (especially regarding the impact on situation model) due to null effects.

## General Discussion

The aim of the present dissertation was to investigate learning methods – interleaving of information units and cohesion generation – with respect to their effectiveness in promoting relational processing while reading expository texts. In the following, we will discuss the results across Experiments 1a, 1b, 2a, and 2b against the background of preceding assumptions such as the *cohesion-coherence-mismatch* and the *ability-requirement-mismatch*. Furthermore, we presumed that learners are less engaged in relational processing when they fail to recognize and close the cohesion gaps. Triggering learners to make inferences while reading could thus be achieved by making the cohesion gaps visible and supporting learners in bridging information to close the gaps independently of their previous knowledge. Both interleaving and generation task were tailored for making the cohesion gaps visible and supporting learners in bridging information.

After discussing whether the preceding assumptions could be supported, we will discuss how reading skill and previous knowledge help learners to overcome the particular struggles of detecting and closing the cohesion gaps. To be able to answer our main research question, that is, under which conditions and for whom are those learning methods beneficial, we will systematically compare interleaving and generation among each other with regard to their implementation, their impact depending on learners' proficiency levels (whether those learning methods are *desirable* for learners with a high vs. low level of reading skill and previous knowledge respectively), and learners' metacognitive judgments (whether learners perceive those learning methods *difficult*). We will compare interleaving and generation also against further learning methods that promote relational processing and particularly increase the awareness of *coherence* gaps. Finally, we will indicate the potential of the present research with regard to future directions and educational implications.

### Ability-Requirement-Mismatch

In accordance with previous literature, we presumed that especially less skilled learners need support and stimulation. Two mismatches represent the struggles of the less skilled learners, the *cohesion-coherence-mismatch*[11] and especially the *ability-requirement-mismatch*.

Engaging learners by a generative task helps in overcoming the cohesion-coherence-mismatch. However, applying generative learning instructions for all learners is also of little use. The ability-requirement-mismatch is a consequence of the interplay of processes triggered and required by a generative learning task and learners' proficiencies: High skilled learners are capable of performing accurately on the generation task, but do not require additional cognitive engagement, less skilled learners in contrast require additional cognitive engagement, but are less capable of performing accurately on the generation task.

Experiment 2a provides support for the preceding assumption of the ability-requirement-mismatch. A linear regression analysis revealed a significant impact of learners' proficiencies – previous knowledge, reading skill, and word analogy respectively – on generation accuracy. That means that less skilled learners are less able to accurately generate inferences; but the ability to generate inferences increases with higher scores on those proficiencies. At the same time, skilled readers do not take advantage of a generative learning instruction, but poor readers do, especially if they succeed to perform accurately on the task.

Altogether, the results speak in favor of the preceding assumption of the ability-requirement-mismatch. In view of that, a generative task sometimes may be an in vane

[11] The cohesion-coherence-mismatch is a consequence of opposite functions of a fully cohesive and non-cohesive text. A fully cohesive text supports learners in establishing a coherent mental representation, but lowers the necessity in relational processing. Cohesion gaps in contrast provide no instruction, but engage learners in relational processing. We designed our experiments based on this presumption rather than explicitly addressing it.

learning strategy for coherence construction, irrespective of the learners' proficiency levels (less skilled learners are not able and high skilled learners do not require).

To overcome the ability-requirement-mismatch, on the one hand, less skilled learners should be engaged in relational processing (to overcome the lack of spontaneity) and on the other hand, their generation accuracy should be supported (to overcome the lack of ability). Educational research suggests learning aids to compensate for both deficits respectively, lack of spontaneity and lack of ability: For example, self-explanation prompts for triggering relational processing to compensate for the lack of spontaneity (Roelle et al., 2014) and pre-training for enabling accurate performance to compensate for the lack of ability (Ainsworth & Burcham, 2007). However, as we pointed out in the introduction, there are reasons to assume that especially less skilled learners do not spontaneously follow generative learning instructions. That is why we considered learning tools that are incorporated into the expository text preferable to a generative learning instruction supplementary to the text.

## Overcoming the Cohesion-Coherence-Mismatch

Incorporating cohesion gaps engages learners in relational processing (McNamara et al., 1996). According to the previous line of reasoning, especially less skilled learners should benefit from cohesion gaps because they lack the spontaneity for relational processing. However, research showed no learning benefits – but disadvantages – for low knowledge learners due to the lack of ability to close the cohesion gaps (McNamara & Kintsch, 1996).

We traced our own interpretation of why some learners may not take advantage of cohesion gaps. Our interpretation is grounded in the assumption that cohesion gaps engage learners in spontaneous relational processing if learners notice the gaps. Within a non-cohesive text, the cohesion gaps are not explicit and might therefore remain invisible for readers. Irrespective of the level of previous knowledge that is necessary for closing a gap, knowledge integration cannot be initiated as long as a gap is not identified. Only after recognizing a gap, previous knowledge is of use for closing it. We designed our learning tools in accordance with this idea.

We enhanced the visibility of cohesion gaps by manipulating the expository text design. Across four experiments, we accentuated the cohesion gaps either via the juxtaposition of comparable information units (i.e., interleaving in Experiments 1a and 1b) or by making the gaps explicit (i.e., cohesion generation in Experiment 2a and 2b). Results speak in favor of the effectiveness of our learning tools in terms of relational processing, short-, and long-term learning (particularly for poor readers in Experiment 2a) and therefore indirectly in favor of the preceding assumption that the spontaneity deficit of less skilled learners can be partly attributed to their inability to detect the cohesion gaps. Importantly, readers of an interleaved text – which is supposed to highlight the cohesion gaps – were more likely to spontaneously generate inductive inferences than readers of a blocked text, as has been demonstrated in Experiment 1b, that is, without any specific instruction. It should be noted that our assumption that the spontaneity deficit of less skilled learners can be partly attributed to their inability to detect the cohesion gaps has not been addressed so far in previous research and requires further examination in future research.

It is also important to mention how we tried to support learners in overcoming the lack of ability to close the cohesion gaps. We provided the basic information across *all* conditions. We designed the expository texts in such a way that learners were able to make inferences independently of their previous knowledge (which they might lack), but based on basic information in text. Thus, theoretically, learners were able to bridge basic information in order to close the cohesion gaps if they lacked the previous knowledge. We furthermore manipulated the availability of critical information pieces via the sequence in Experiments 1a and 1b: Interleaving juxtaposed the critical pieces of information making bridging inferences more likely than presenting categories one by one via a conventional (blocked) sequence.

## What Matters, Continuity Disruptions or Visibility of Cohesion Gaps?

As next, we will try to exclude an alternative interpretation (which is *continuity disruptions*) of the interleaving effect in terms of inductive reasoning in Experiments 1a and 1b. Our interpretation refers to the visibility of cohesion gaps due to the juxtaposition of categories. As an additional inherent consequence of interleaving, whales' complementary characteristics were spaced across the paragraphs.[12] Thus, an interleaved presentation inherently led to *continuity disruptions*, whereas a blocked presentation maintained continuity. Consequently, interleaving set additional demands on self-regulated linking of complementary information (e.g., *size*, *lifespan*, *habitat* and further characteristics of a *fin whale*) across the paragraphs to maintain continuity. According to research findings on reading comprehension, interleaving as a disrupted presentation (as being compared to a blocked sequence) could be expected to hinder relational processing due to increased distances across complementary information units (cf. Schnotz, 1984). Reading an interleaved text results thus in fewer opportunities to simultaneously process complementary information than reading a blocked text (Wiley & Myers, 2003). We will call it the *continuity assumption*.

However, in Experiments 1a and 1b, relational processing did not depend on the continuity of presentation. On the contrary, the greater difficulty caused by continuity disruptions led to higher learning outcomes (cf. Bjork & Bjork, 2014). Based on the present findings and theoretical considerations, we can conclude that in defiance of continuity disruptions, interleaving engaged learners in relational processes. This finding contradicts the continuity assumption forwarded by the research on reading comprehension.[13] Hence, the

[12] Confounding of interleaving and spacing is not a specific limitation of our study design, but an inherent feature of interleaving. Thus, all previous studies on interleaving are limited in this sense. Birnbaum et al. (2013) and Kang and Pashler (2012) orthogonally manipulated interleaving (=juxtaposing of categories) and spacing (continuity disruptions). Their results speak in favor of interleaving over spacing.

[13] Text comprehension research also provides arguments that partially reconcile the continuity assumption and no harm result by continuity disruptions. Distant information that has been stored in episodic memory as part of comprehension could remain available during reading because the information in a discourse is hierarchically rather than simply linearly structured (van Dijk & Kintsch, 1983). Thus, the prior textual information that is no longer in attentional focus could be reinstated and connected with the newly read information (van den Broek et

results of the Experiments 1a and 1b may have valuable implications for understanding of how text-characteristics trigger learning processes.

We do not consider the continuity disruptions the valid explanation of the interleaving effects obtained in Experiments 1a and 1b due to two reasons. First, it is at odds with the continuity assumption, which assumes the continuity disruptions to cause learning *dis*advantages. The second reason refers to null findings in previous research when textual materials were discontinuously presented (without juxtaposition). Inconsistent findings across studies on sequencing textual learning materials suggest that the interleaving effect depends on, among other factors, semantical relations across textual sources: As a result, previous studies that have used textual materials of semantically non-related topics found no effect of interleaving (Dobson, 2011; Hausman & Kornell, 2014; Mandler & DeForest, 1979). If only continuity disruptions accounted for the effect, the semantical overlap of the texts would be insignificant. A reasonable assumption is that continuity disruptions, which inherently result from reading unrelated texts in a mixed manner, engage readers in distinctive rather than relational processing. Therefore, not the continuity disruptions but the juxtaposition of comparable information units may account for the interleaving effect.

Despite our reasoning, a seemingly near at hand explanation of the relational processing benefit by continuity disruptions may come from the construction-integration model. According to the construction-integration model, cohesion gaps engage readers in processes of knowledge integration (i.e., making *elaborative* inferences) (McNamara et al., 1996). Thus, if learners have a sufficient level of previous knowledge, cohesion gaps engage them in closing those gaps.

al., 2015). This assumption is even more plausible for the interpretation of our findings because we used printed learning materials in Experiment 1a – readers could easily look back in the text.

Such an argument though equates continuity disruptions and cohesion gaps, which is a wrong premise. *Continuity* refers to temporal spacing, whereas *cohesion* refers to the extent that the relations in text are explicit (e.g., number of relational statements). Continuity and cohesion can thus be manipulated orthogonally: A fully cohesive text can be presented discontinuously (spaced) and a non-cohesive text can be continuously presented. Referring back to sequences, an interleaved sequence is a discontinuous one because the categories are presented alternately, whereas a blocked sequence is a continuous one because the categories are presented one by one. Thus, interleaving inevitably leads to continuity disruptions. However, an interleaved text is not different from the blocked text with regard to the extent that the relations are explicit: The factual statements about the whales' characteristics (used in Experiments 1a and 1b) were identical for the interleaved vs. blocked conditions – only their sequence was manipulated. The text lacked any relational statements (e.g., comparative and inductive), resulting in a low cohesion level, irrespective of the sequence. Thus, the blocked text was not less cohesive, but equally non-cohesive as the interleaved text.

The interleaved and the blocked texts were not different with regard to cohesion gaps – both texts were non-cohesive. According to the text comprehension research, cohesion gaps promote relational processing. Hence, the question arises of why the cohesion gaps triggered the readers of an interleaved text to a higher extent than the readers of a blocked text. The answer might be that the interleaved and blocked texts were different with regard to the *visibility* of cohesion gaps. Based on the pattern of results by Experiments 1a and 1b, we conclude that learners were not aware of cohesion gaps while reading a blocked text: For example, learners focused solely on factual statements, irrespective of the support provided by a fixed order of factual statements (Experiment 1a) or by self-questioning prompts (Experiment 1b). Learners probably considered the text cohesive and did not miss any relational statements (as also indicated by equal experience of cognitive load while reading a non-cohesive text and a fully cohesive text in Experiment 2b). Consequently, readers of a blocked text made no efforts to compare or relate information units. A blocked text thus prevented readers from recognizing the cohesion gaps and establishing coherence. An interleaved sequence in contrast juxtaposes categories and consequently invites learners to make comparisons, which might increase learners' awareness of cohesion gaps concerning the underlying regularities.

### Cohesion Gaps – a Redundant Assumption?

In this paragraph, we will address possible objections concerning our explanation of the interleaving and generation effects in the present work. We namely explain an increased level of relational processing while reading an interleaved text (in Experiments 1a and 1b) and generating cohesion (in Experiments 2a and 2b) by referring to the increased visibility of cohesion gaps. In other words, we explain the generation of an inference by referring to its *noticeable* lack. Obviously, *making an inference* analytically implies the absence of this inference in the text. Our explanation might thus appear to be redundant. It might be questionable whether making an inference requires the awareness of a cohesion gap in the first place. Although we elicited indices of relational processing – such as text-box entries in Experiment 1b, cognitive load via a dual task and generation accuracy in Experiments 2a and 2b – we have no stringent measures of the awareness of cohesion gaps. In the following, we will provide some counterarguments against these theoretical objections by referring to previous research investigating the tools that are supposed to increase the awareness of gaps, and our studies in particular.

#### ***Comparing Tools Making Readers Aware of the Cohesion and Coherence Gaps***

First of all, the argument that cohesion gaps (that is, deleting of relational statements, background information, and cohesion devices) trigger relational processing is empirically supported and widely acknowledged across the researchers on text comprehension

(McNamara et al., 1996). That readers might easily overlook cohesion gaps (and inconsistencies in particular) is also empirically backed up (McKoon & Ratcliff, 1992).

Furthermore, there is a growing body of research demonstrating the impact of learning methods that are supposed to increase the awareness of *coherence* gaps (i.e., knowledge gaps) – that is the awareness of gaps in the own mental representation – on relational processing, study behavior, and learning outcomes. Herein, the basic line of argumentation is the following: The function of the expository text is to close learners' coherence gaps. However, learners are prone to the illusions of understanding (Koriat & Bjork, 2006a), which motivate learners to terminate study efforts too quickly and prevent them from deep processing of the contents. Therefore, successful knowledge integration requires learners to be aware of their coherence gaps.

Raising learners' awareness of their coherence gaps can be achieved either via a text design (e.g., *refutation texts*) or by a generative learning instruction triggering metacognitive monitoring. Examples of the latter are *problem solving prior to instruction – productive failure approach* (Loibl & Rummel, 2014), *self-testing* (McDaniel et al., 2009), and self-testing combined with *judgements of inference* (Nguyen & McDaniel, 2016). Over and above, approaches that prevent learners from self-overestimation and in turn from a quick termination of study efforts hold educational value (Roelle et al., 2017). Herein, e.g., Thiede et al. (2003) demonstrated a positive impact of delayed keyword generation on *monitoring accuracy*, which allowed learners to make appropriate study decisions and in turn improved learning.

To highlight the analogy between the explanatory mechanism proposed by us (visibility of cohesion gaps) and the explanatory mechanism suggesting the awareness of coherence gaps to be the missing link, we will explain one method – namely the refutation texts – more detailed. Interleaving alike, refutation text is a manipulation of text-characteristics. Different from a standard expository text, the refutation text addresses learners' common misconceptions on topic and label them being erroneous. Misconceptions are false beliefs on topic that are deeply rooted in learner's knowledge networks, which on the one hand creates the illusion of understanding and on the other hand interferes with correct knowledge (Kendeou et al., 2011; Kowalski & Kujawski Taylor, 2009). Refutation text is supposed to mend misconceptions, that is, not to update the knowledge network, but to outdate wrong assumptions (Asterhan & Resnick, 2020). In this sense, reconstructing of a knowledge network should not be called generally *learning*, but *conceptual change* (Prinz et al., 2019). Interleaving alike, refutation text *juxtaposes* correct contents and learners' common misconceptions. The *co-activation hypothesis* explains the positive impact of refutation texts (i.e., conceptual change) with reference to the co-activation of correct statements and wrong assumptions within the working memory (Allen et al., 2015). In other words, the impact of refutation texts is explained with reference to the discriminative contrast between correct statements and wrong assumptions. Thus, the explanatory mechanism behind the refutation text and interleaving effects is basically the same. In line with this reasoning, Maier et al. (2018) drew a parallel between the interleaved presentation of belief-consistent and belief-inconsistent texts on the one hand and refutation texts on the other hand. We also draw a parallel between refutation and interleaved texts: Whereas refutation texts help learners to recognize and close their coherence gaps, an interleaved presentation sequence may support learners in recognizing and closing the cohesion gaps and consequently in establishing coherence.

### *Awareness of Cohesion Gaps – the Necessary Link in Explaining Our Results?*

Referring back to our studies, reading interleaved texts promoted the generation of comparative and inductive inferences in Experiment 1b. We do *not* explain a greater number of *comparative* inferences with an increased awareness of cohesion gaps because a juxtaposed text structure allows readers to directly compare the categories. However, different from the impact on making comparative inferences, the explanation of the impact on making inductive inferences (i.e., identification of co-occurring patterns) is not that simple because making inductive inferences requires learners to make global bridging inferences across multiple paragraphs even in the interleaved condition. That is, based on solely one paragraph, learners are not able to identify that e.g., *a large body size goes along with a small group size*. Hence, the interleaving effect in terms of making inductive inferences while reading (and inductive reasoning as learning outcome) in Experiment 1b needs to be explained against the background of studies on text comprehension indicating a lazy reader (Coté et al., 1998; Hyönä et al., 2002).

We resolve this seeming inconsistency with the awareness of cohesion gaps while reading an interleaved text. Our particular explanation in the article referred to an increased level of curiosity: Learners might have been wondering of how to explain the similarities and differences across categories' characteristics. For example, they might have been asking themselves of: *Why does this whale migrate but the other one barely leaves the place? These two whales have similar group sizes, do they also resemble in other characteristics? The range of the size across whales is huge, how do whales differ in further characteristics?* In turn, such questions may have triggered learners to search for co-occurring patterns (the text was enriched with basic information necessary for a successful pattern recognition). In other words, juxtaposition of information units may have increased learners' awareness of absent relations across information units in text and engaged them in making global bridging inferences for closing those gaps.

## Specific Role of Learners' Proficiencies in Detecting and Closing Cohesion Gaps

### *Aptitude-Treatment-Interactions*

Learners differ with respect to their proficiencies and deficits. Learning materials and methods differ with respect to particular demands imposed on learners and the extent of compensating for learners' particular deficits. Accordingly, some combinations of learning methods and learners' proficiencies *mis*match (e.g., overwhelming demands or redundant support). The question of which learning method is suitable for which learner holds thus educational value.

Because of a broad interest, we propose that the so-called *aptitude-treatment interaction* is presently becoming an inflationary applied concept in educational science. The range of aptitudes in educational science embrace cognitive and motivational prerequisites. Studies reporting an aptitude-treatment interaction use various cognitive aptitudes such as previous knowledge (McNamara & Kintsch, 1996), reading skill (McDaniel et al., 2002; Naumann et al., 2007), working memory capacity (Lehmann et al., 2016; Sana et al., 2018), and even grades (Holley et al., 1979). Irrespective of the particular proficiency, there are review articles calling learners with high scores on either prerequisite *higher skilled* and their counterparts with low scores *less skilled* learners (Fiorella & Mayer, 2016). Motivational prerequisites are for example domain specific self-concept and anxiety (Fleischer et al., 2014), performance expectancies (Reinhard et al., 2019), need for cognition (Schindler et al., 2019), and prior beliefs (Maier et al., 2018).

Usually studies reporting an aptitude-treatment interaction refer solely to *one* prerequisite in their hypotheses and state having assessed only *this* prerequisite. Against the background of a plenty of cognitive and motivational prerequisites, the choice of a particular prerequisite is sometimes not even explicated. The reasons for a particular aptitude-treatment interaction assumption are often interchangeable, e.g., previous knowledge and memory capacity for many researchers seem to represent an aptitude dealing with the task demands (Kalyuga, 2006; Sweller, 2010). Since there are a plenty of theoretical possibilities, the decision *for* one prerequisite is a decision *against* further prerequisites. It may be thus favorable if researchers address und explain their particular choice by referring to specific task demands, how a particular proficiency matches those demands, and how particular learning aids compensate for those demands.

Not only the choice of a particular aptitude (or its particular assessment tool)[14], but also the direction of the interaction seems to be arbitrary in the educational research. That is, either the less skilled learners are supposed to fail to fulfill the task demands and thus require support whereas higher skilled learners require more challenging learning methods (Kalyuga et al., 2003; Lehman et al., 2014; McNamara et al., 1996; McNamara & Kintsch, 1996) or less skilled learners require stimulation whereas higher skilled learners are cognitively engaged anyway (McDaniel & Butler, 2011; Schindler et al., 2019). There is thus a plausible explanation for any possible direction of interaction, which makes the concept of aptitude-treatment interaction not only inflationary, but also *ad hoc*.

Due to the mentioned reasons, aptitude-treatment interactions may be prone to ad hoc interpretations. Generally, pre-registration may prevent the ad hoc interpretations of particular aptitude-treatment interactions in future research. The present work apparently cannot solve this problem. Our line of reasoning, however, is supposed to increase readers' awareness of ad hoc interpretations in educational research by highlighting the simplicity (and publishing advantage) of their usage.

[14] By assessing various prerequisites – but reporting solely the one that significantly interacted with condition – researchers incidentally boost the alpha error accumulation. Over and above, there are various measures for assessing several proficiencies that can be used within a study (or transversely – for students participating across multiple studies), which further increases the alpha error accumulation.

Furthermore, we encourage researchers to pay more attention to distinct processing components of various cognitive and motivational prerequisites. The insights in prerequisites' particular processing components might provide researchers with a priori decision aids and prevent them from making ad hoc interpretations. In the following, we will present some considerations on distinct processing components of previous knowledge and reading skill in learning from expository texts.

### ***Previous Knowledge – a Prerequisite of Closing Cohesion Gaps***

According to the research on text comprehension, previous knowledge is a necessary prerequisite to generate elaborative inferences and to close cohesion gaps (Kintsch, 1988). We found direct support for this presumption in Experiments 2a and 2b: Previous knowledge was predictive for generation accuracy (that is the accuracy of closing the cohesion gaps). Apparently, learners who lack the necessary knowledge to close the cohesion gaps cannot benefit from the learning potential of cohesion gaps and strongly rely on cohesion devices and background information in text (McNamara et al., 1996). High knowledge learners in contrast fulfill the necessary condition: They succeed in making elaborative inferences and closing the gaps while reading a non-cohesive text, and in turn benefit in terms of situation model construction (McNamara & Kintsch, 1996).

The following argumentation is in line with the pattern of results across a large number of studies on text comprehension. Noticeable cohesion gaps trigger relational processing, irrespective of the level of previous knowledge (cf. McKoon & Ratcliff, 1992). However, relational processing itself is in vane as long as the cohesion gaps cannot be adequately closed (O'Brien & Myers, 1985) – previous knowledge is a necessary prerequisite to close the cohesion gaps. There is thus no benefit of cohesion gaps if learners lack the previous knowledge to close the cohesion gaps. In such a case, the advantages of instructional support provided by a fully cohesive text outweigh the advantages of a non-cohesive text

(McNamara et al., 1996; McNamara & Kintsch, 1996). It seems obvious that as long as the text itself does not provide the necessary information for closing the cohesion gaps, the learning success will crucially depend on previous knowledge, resulting in an aptitude-treatment-interaction described above.

The few studies on text sequences that analyzed the moderating impact of previous knowledge found interleaving effects for high knowledge learners, whereas low knowledge learners took no advantage of reading a *compare/contrast* (i.e., interleaved) text (Wiley & McGuinness, 2004). This pattern of results is consistent with our reasoning. Given that an expository text used in the particular studies lacked background information for establishing coherence (and many relations in text were implicit), previous knowledge was necessary to close the cohesion gaps. High knowledge learners could easily recognize the cohesion gaps due to an interleaved sequence and benefitted in terms of situation model construction by closing the gaps. While reading a poorly structured text, in contrast, high knowledge learners were less likely to use their previous knowledge because many gaps remained invisible. However, even if interleaving highlights the cohesion gaps – as we assume – learners who lack the necessary previous knowledge were still not able to close the gaps. Accordingly, previous knowledge correlated with learning success in interleaved conditions, whereas no correlation was observed in blocked conditions (Schnotz, 1982). Different from fully cohesive texts, blocking (e.g., *enumeration* structure in the study of Wiley & McGuinness, 2004) did not support low knowledge learners to a higher extent than interleaving because – as previously explained – a blocked text provides the same amount of information as an interleaved one. That is, while a blocked text is continuously presented and an interleaved text entails continuity disruptions, the extent of cohesion is independent of text sequence. In other words, low knowledge learners did not benefit from blocking because – different from a fully cohesive text – blocking provides no instructions on establishing coherence.

We assume that the studies that found learning advantages of cohesion gaps (and interleaving in particular) only for high knowledge learners used no aids to compensate for the lack of previous knowledge. In Experiments 1a and 1b in contrast, such aids for compensating for the lack of previous knowledge were provided: Coherence establishing conclusions were possible based on adjacent information units. Learners who lacked the previous knowledge to close the cohesion gaps could thus compensate for this lack by generating inferences based on adjacent information units.[15]

Under such circumstances, we would expect interleaving to support not only high knowledge learners, but also low knowledge learners, which would result in a main effect of study sequence and no aptitude-treatment interaction.[16] We would have expected the same pattern of results for reading a non-cohesive text as compared to reading a fully cohesive text (that is a main effect of cohesion gaps and no aptitude-treatment interaction) if readers were able to close the gaps independently of their previous knowledge, but based on information in text.

### *Reading Skill – a Prerequisite of Detecting Cohesion Gaps*

Given that previous knowledge enables learners to close the cohesion gaps while reading an expository text, the question arises of which proficiency enables learners to

[15] In Experiment 2a, further proficiencies (reading skill and word analogy) were predictive for generation accuracy over and above the impact of domain specific previous knowledge. The cohesion gaps could thus be closed not only based on elaborative inferences, but also by making bridging and world-knowledge inferences. This founding speaks in favor of a successful implementation of textual aids that allow learners to close cohesion gaps independently of their previous knowledge. However, it is favorable to improve textual aids in order to further reduce the impact of previous knowledge.

[16] Unfortunately, we have not assessed previous knowledge in Experiment 1a and 1b to underspin our assumption because a preceding assessment of previous knowledge could have driven learners' attention to comparisons and underlying regularities, which we wanted to be accentuated solely by sequence manipulation. However, the main effect of interleaving for both college students (Experiment 1b) and 8th and 9th grade students (Experiment 1a) might speak in favor of a successful implementation of textual aids. That is, learners may have closed the cohesion gaps based on adjacent information in text and by doing so taken advantage of reading an interleaved text, irrespective of their level of previous knowledge. Future studies should directly manipulate previous knowledge (training yes vs. no) and the presence of basic information for bridging inferences (provided vs. not provided) to investigate whether the aptitude-treatment interaction can be eliminated in favor of a main effect of interleaving (that is, equally for learners with a high and a low level of domain specific previous knowledge).

recognize cohesion gaps in the first place. Hannon and Daneman (2001) distinguish processing components of coherence construction while reading an expository text. They found a positive link between reading skill measures and learners' ability to integrate previous knowledge, that is, to make inferences based on previous knowledge (i.e., elaborative inferences). According to this finding, reading skill initiates the use of previous knowledge to establish coherence. That means that a high level of domain specific previous knowledge alone is not sufficient (but necessary) to close a cohesion gap: Poor readers might easily overlook the opportunities to retrieve their knowledge, which actually might be sufficient to establish a coherent relation. Skilled readers in contrast have a smooth access to their previous knowledge. In this sense, the finding of McNamara and colleagues that readers with a high level of previous knowledge benefit from cohesion gaps, irrespective of their reading skill, might be attributed to a generally high level of reading skill across college students.

There are findings that indicate an alternative view (O'Reilly & McNamara, 2007; Ozuru et al., 2009; Voss & Silfies, 1996). Researchers found a link between reading skill and readers' ability to process cohesion. A higher cohesion extent – that is, a greater number of explicit relations between information units – enhances the processing demands. Reading skill in turn is the ability to deal with processing demands.[17] According to this view, reading a fully cohesive text is especially beneficial for low knowledge readers who are able to deal with additional demands imposed by cohesion.

The latter interpretation should be critically reflected because the authors propose an inherent link from cohesion to complexity. According to their view, cohesion implies a fine-grained zoom-in into the underlying relations across the idea units in text, which increases the

[17] A similar line of reasoning concerning the processing demands imposed by cohesion was proposed by de Jonge et al. (2015). According to them, a scrambled (non-cohesive) text imposes less cognitive load than a highly structured (cohesive) text due a lower level of element interactivity (i.e., number of interconnected elements).

processing demands. The assumption of low processing demands imposed by a non-cohesive text implies that cohesion gaps are not recognized by readers – otherwise the processing demands would increase.

The view linking reading skill with processing cohesion is further not compatible with the cognitive load theory. According to the core assumptions of CLT, processing interrelated information is a direct function of previous knowledge, namely of *chunks*, which decrease the element interactivity and in turn allow learners to reduce the demands on working memory capacity (Kalyuga & Singh, 2015). The working memory capacity span also may be helpful in dealing with demands imposed by a high element interactivity because it allows simultaneous processing of a high number of interconnected elements (Kalyuga, 2006). According to CLT, there is no free processing capacity for knowledge construction processes (germane load) if the previous knowledge is not sufficiently high to substantially reduce the element interactivity of the task (intrinsic load) (Sweller, 2010). Thus, the text comprehension research and CLT ascribe the function of processing relations to different proficiencies, either to reading skill (text comprehension research) or previous knowledge and working memory (CLT). Experiment 2b revealed the reducing impact of previous knowledge on intrinsic cognitive load, whereas reading skill had no impact on intrinsic cognitive load, supporting the CLT view.

The pattern of results obtained in Experiment 2b is inconsistent with the view linking reading skill with processing cohesion. According to this view, reading a fully cohesive text should be beneficial for low knowledge/high skilled readers. Reading a non-cohesive text in contrast should be beneficial for high knowledge readers, irrespective of their reading skill. Our results, however, indicate a different pattern, that is, a learning advantage of reading a fully cohesive text for high knowledge/poor readers. Reading a non-cohesive text, in contrast, benefitted high knowledge/high skilled readers. Thus, we did not obtain any support for the assumption that reading skill is a proficiency in processing cohesion.

Our results rather support the view of Hannon and Daneman (2001) that reading skill helps in integrating previous knowledge. Specifically, we assume that reading skill entails the ability to detect implicit cohesion gaps while reading a non-cohesive text. Given that reading skill helps readers to detect the cohesion gaps and previous knowledge helps them to close the gaps, we would have expected the yielded pattern of results in Experiment 2b, namely that only the high knowledge/high skilled readers benefit from reading a non-cohesive text: Cohesion gaps support coherence construction if readers are able of recognizing the gaps (due to a high reading skill) and closing them (due to a high previous knowledge). A fully cohesive text in contrast supported knowledge acquisition when high knowledge readers failed to recognize the cohesion gaps (due to a low reading skill). Contrary to our argumentation – *and* previous research –, in Experiment 2b, low knowledge readers did not benefit from reading a fully cohesive text. Furthermore, the yielded pattern of results was limited to the text-based representation and did not spread over to the situation model construction.

Given that poor readers fail to recognize the cohesion gaps, interleaving and cohesion generation may relieve the strains by making cohesion gaps visible. In this sense, especially poor readers should benefit from learning aids that highlight the cohesion gaps. Skilled readers in contrast may not require such an aid because they may spontaneously recognize the cohesion gaps. In line with this expectation, Experiments 2a and 2b yielded lower forgetting rates for readers in the cohesion generation condition, which made cohesion gaps visible and required learners only to close them. Over and above, in Experiment 2a, poor readers who accurately closed the gaps, achieved higher learning scores than their counterparts who read a fully cohesive text. Contrary to our argumentation, in Experiment 2b, we yielded no superiority of generating cohesion, especially not for poor readers with a high level of previous knowledge.

More research is required to independently assess various processing components (cf. Hannon & Daneman, 2001) and to explore how those processing components are linked to learners' proficiencies such as previous knowledge, reading skill, and working memory capacity. From the educational perspective, it would help researchers to understand the particular deficits of learners and to make tailored recommendations for learning aids compensating for particular deficits. From a theoretical perspective, it would help researchers to make reasonable and accurate predictions with respect to aptitude-treatment interactions.

### Interleaving and Cohesion Generation – Desirable Difficulties?

In this section, we will consider both learning tools, interleaving and cohesion generation, against the background of the *desirable difficulties framework*. Particularly, we will compare interleaving and cohesion generation with respect to their implementation and their impact on relational processing depending on learners' proficiency levels.

An apparent difference between interleaving and cohesion generation is that interleaving is a manipulation of text-characteristics, whereas cohesion generation is a task. Apart from this duality, both tools have several similarities in their implementation and function. Interleaving alike, cohesion generation task is inherently incorporated into the text: Causal connectives are removed, leaving behind open gaps, which learners have to close in order to complete the text. Both tools can serve the same function, namely highlighting the cohesion gaps to compensate for learners' deficit of recognizing the cohesion gaps. Accordingly, we yielded interleaving effects not only for more skilled readers in Experiment 1b, but also for younger (less skilled) readers in Experiment 1a; and generation effect could be obtained only for poor readers in Experiment 2a.

Can we consider our learning tools *desirable* for relational processing while reading? Given that especially poor readers do not notice cohesion gaps while reading, a tool that highlights cohesion gaps compensates for this particular deficit and thus should be considered desirable. The more a tool takes over of a particular demand, the less relational processing will depend on the particular proficiencies. Accordingly, relational processing should be less dependent on reading skill when reading an interleaved text than reading a blocked text. Analogously, relational processing should be less dependent on reading skill when generating causal cohesion than reading a non-cohesive text (for which we found so far no evidence).

When directly comparing interleaving and cohesion generation, interleaving appears less restrictive than cohesion generation with regard to highlighting cohesion gaps because the cohesion gaps in an interleaved text are – more apparent than in a blocked text, but – still implicit. Thus, interleaving may compensate less for the demand to recognize the cohesion gaps than cohesion generation. However, cohesion generation highlights merely the *local* – near at hand – cohesion gaps (between subsequent clauses) but not the *global* ones. Accordingly, in both experiments on causal cohesion (2a and 2b), generation yielded no main effect in terms of situation model. Interleaving in contrast might increase learners' awareness of global cohesion gaps, as has been demonstrated in Experiment 1b. Therein, readers of an interleaved text spontaneously generated inferences on underlying regularities, whereas self-questioning prompts aiming to increase the awareness of cohesion gaps narrowed the attentional focus and highlighted consequently the local rather than the global cohesion gaps. Taken together, it is not clear whether relational processing is more or less dependent on reading skill when generating cohesion in comparison to reading an interleaved text.

The particular implementation of interleaving and cohesion generation was also supposed to support learners who lack the previous knowledge for closing the gaps by providing basic text information for bridging inferences. This aim required additional manipulations on the text surface across *all* conditions. Thus, our expository texts contained the necessary information for closing the cohesion gaps. Given that especially low knowledge readers – even if recognized – are less likely to accurately close the cohesion gaps than high knowledge readers are, an additional support compensating for the lack of previous knowledge should also be considered desirable.

Although the number of information units was equal in interleaved and blocked conditions in Experiments 1a and 1b, we assume that especially readers of an interleaved text were more likely to bridge basic information (=making comparative inferences) not primarily because the text was enriched with the necessary information, but specifically because those information units were adjacently placed. Thus, to close the near at hand cohesion gaps, readers of an interleaved text were merely required to link adjacent information units (that is, making local comparisons), whereas the readers of a blocked text were required to link distant information units (that is, making global bridging inferences).

In this sense, we assume that readers of an interleaved text simultaneously process critical pieces of information and spontaneously make comparisons. We thus consider the awareness of cohesion gaps not the necessary condition for closing the near at hand gaps, but a consequence. By spontaneously bridging adjacent information, readers in turn detect more global cohesion gaps, especially concerning the underlying regularities.

When directly comparing interleaving and cohesion generation, it is unclear which tool is more restrictive with regard to closing cohesion gaps. On the one hand, cohesion generation is very restrictive because the connective choice involved solely four alternatives (*because*, *therefore*, *although*, and *however*). On the other hand, the generation accuracy in Experiments 2a and 2b still depended on previous knowledge (and further proficiencies) probably due to the requirements imposed by the generation task: To make a valid connective choice, readers were required to reflect on the *direction* of a causal link (which of two clauses is the cause and which one is the consequence) and on its *polarity* (whether a particular causal link matches or mismatches the own expectation). Such reflections probably cannot be made solely based on the contents in text and require readers to make *world knowledge inferences*. Readers might thus require further support to improve their generation accuracy.

Due to the mentioned reasons, we consider interleaving more supportive for closing the cohesion gaps than cohesion generation. Accordingly, the quality of relational processing should depend less on previous knowledge when reading an interleaved text than generating cohesion.[18] However, our interpretation should be considered with caution because the complexity of learning materials may also affect the support for closing the gaps. The topic of the text used in Experiments 2a and 2b – *greenhouse effect and climate change* – is more complex than *life of marine mammals* used in Experiments 1a and 1b. Therefore, several information units necessary to close a gap were inherently spaced across the text, imposing higher demands on readers who were engaged in cohesion generation.

To sum up, interleaving appears more desirable than cohesion generation with respect to compensating for both deficits (recognizing and closing the cohesion gaps). We observed these benefits for young (Experiment 1a) and more experienced readers (Experiment 1b), whereas cohesion generation was desirable only for poor readers who succeeded to accurately close the gaps.

The desirable difficulties framework not only emphasizes the desirable impact of learning methods, but also attributes their desirable impact to a higher difficulty compared

[18] We should be cautious in analyzing the relation between previous knowledge and relational processes depending on condition within an experiment and between experiments. Because the recognition of cohesion gaps is preceding to closing the gaps – which is not necessarily true for near at hand cohesion gaps – and conditions (but also expository texts) may differ in the extent that cohesion gaps are highlighted, a between-condition comparison concerning the relation between previous knowledge and relational processes requires to control for readers' recognition performance in the first place. Otherwise, the correlation between previous knowledge and relational processes inherently would be lower in conditions that do not support readers in detecting the gaps: Readers cannot use their previous knowledge to close a gap if this gap remained undetected.

with passive methods, which do not engage learners in deep processing (Bjork & Bjork, 2014). Should our learning tools thus be considered *difficult*? We pursue a different perspective. We consider interleaving and cohesion generation not difficulties but aids compensating for readers' particular deficits in recognizing and closing the cohesion gaps when reading a non-cohesive expository text. Thus, the cohesion gaps engage readers in relational processing (McNamara et al., 1996; O'Brien & Myers, 1985) and our learning tools *lower* their difficulty. In a nutshell, interleaving and cohesion generation provide support to compensate for the difficulty imposed by cohesion gaps and in turn unfold the desirable potential of cohesion gaps for relational processing.

### *Why Learners Perceive Interleaving and Cohesion Generation Not Desirable?*

Taking our latter argument into account, interleaving and cohesion generation compensate for the difficulties imposed by cohesion gaps, rather than being difficulties themselves. Does it thus mean that the label *desirable difficulties* is not justified in the case of interleaving and cohesion generation? We consider both tools desirable difficulties in defiance of our latter argument. The seeming contradiction resolves when taking learners' perspective into account, particularly when distinguishing between objective difficulty, subjective experience, and metacognitive effectivity judgements on the one hand and between creating difficulty in the first place and facing the existing difficulty on the other hand.

The vast majority of students is convinced that learning in an interleaved sequence is less beneficial than in a blocked one (Kornell & Bjork, 2008; McCabe, 2011). People erroneously believe that an interleaved sequence makes a mess of everything (Yan et al., 2016). Thus, learners are not aware of the benefits of juxtaposing categories on relational processing. On the contrary, learners consider the *within* category comparisons essential for category learning (Yan et al., 2017). Not only is the majority erroneously convinced that blocking is the superior sequence, this misbelief is also relatively resistant against resolution

(Yan et al., 2016). Yan and colleagues revealed cognitive biases that undermine conceptual change, but make learners sticking to their previous misbeliefs: For example, when making learners aware of interleaving advantage for 90% of people, 90% of people actually assign themselves to the 10% that benefit from blocking. Their further quasi-experimental investigation showed that students attribute superior learning of interleaved categories to the easiness of interleaved compared to blocked categories. Thereby, students attribute superior learning of blocked categories to the general effectiveness of blocking. Finally, students prefer to block their study materials (Carvalho et al., 2016; Cohen et al., 2013; Kornell & Vaughn, 2018; Tauber et al., 2013; Yan et al., 2017).

Students alike, book designers might design textbooks in a one-category-at-a-time (blocked) manner due to this common misbelief and in anticipation of learners' expectations. Briefly, because learners generally underestimate the benefits of interleaving textual materials (Zulkiply et al., 2012) and particularly the compensatory functions of detecting and closing the cohesion gaps, we consider the label *desirable difficulty* for interleaving textual materials justified.

Researchers explain the incongruence between the objective learning progress and the subjective experience with a low validity of cues learners use to diagnose their learning progress. Processing dis/fluency is such a clue – it gives the feeling of being un/knowing. Thus, learners attribute the fluency of learning to the progress of learning (Koriat & Bjork, 2006b). Based on the fluency of their learning experience, learners may also conclude on the effectiveness of their learning methods (Bjork et al., 2013; Yan et al., 2016).

Leaners in our experiments may have processed disfluency while reading an interleaved text and especially while generating cohesion probably because of inherent continuity disruptions. However, it is also reasonable to assume that specifically the supposed main function of both tools – namely highlighting the cohesion gaps and compensating for the deficit of recognizing the cohesion gaps – was responsible for processing disfluency. Given that interleaving and cohesion generation highlight the cohesion gaps that otherwise remain subtle, learners cannot ignore open cohesion gaps and consequently experience disfluency. The experience of disfluency might be especially striking if learners lack the necessary knowledge and ought bridging text information to close a gap.

The results on cognitive load measures in Experiments 2a and 2b support this view. In Experiment 2a, learners in the generation condition reported higher scores on extraneous cognitive load. Critically, a simple comparison between the generation condition and the non-cohesive text condition in Experiment 2b revealed a higher intrinsic cognitive load in the generation condition: Readers erroneously attribute the difficulty imposed by the generation task to the difficulty of topic. Furthermore, the generation accuracy negatively correlated with extraneous cognitive load and reaction times during the dual task in Experiment 2a, indicating disfluent processing. Thus, because noticeable cohesion gaps engage readers in processes necessary to close the gaps – and enhance readers' awareness of their own knowledge gaps – readers experience less fluency. As a result, learners generally may underestimate their own learning progress and the effectiveness of interleaving and cohesion generation. In contrast, when cohesion gaps remain subtle, the contents in text appear less related than they actually are (i.e., indicating a low element interactivity). Then, learners might experience the illusion of understanding and erroneously judge the learning setting without those tools superior.

### *Combining Interleaving and Generation – Desirable or Redundant?*

We would like to emphasize the importance of paying attention to the interplay of combined learning tools, based on text-characteristics and instruction. Particular instructional prompts can match or mismatch a particular text sequence. Thus, the future research should pay more attention to the question of which instructional prompts and study sequences fit together.

A promising direction comes from the material-appropriate-processing account (McDaniel & Butler, 2011; McDaniel & Einstein, 1989). According to this account, a learning strategy should be chosen based on the consideration of how well it complements the type of text (expository vs. narrative) with respect to triggering processes that are relevant for the learning objective, but are not automatically triggered by the text itself. The MAP account presumes that expository texts automatically drive readers' attention to factual details, but not to relations across idea units (i.e., inferences). The present work investigates thus how to promote relational processing while reading an expository text.

Taking up the MAP presumptions, we consider the text sequence x learning instruction interplay. We argue that an interleaved study sequence automatically directs readers' attention toward comparisons, which in turn might increase readers' awareness of underlying relations not explicitly addressed in the text (i.e., awareness of cohesion gaps). Accordingly, a learning task driving readers' attention to relations by highlighting the cohesion gaps – such as the sentence reordering task (McDaniel et al., 2002), but also our cohesion generation task – should be redundant. In line with this reasoning, the question generation instruction while reading an interleaved text did not improve relational processing in Experiment 1b because it served no complement function. In contrast, readers of a blocked text – but probably of any expository text that inherently lacks comparisons – require a considerable instructional support to be engaged in relational processing. We will address the ways of investigating the interplay of text-characteristics and learning instruction in future directions, particularly with respect to how instructional support can look like to compensate for the disadvantages of blocking.

## Future Directions

Some of our assumptions concerning the explanatory mechanisms behind the interleaving and cohesion generation effects were not explicitly addressed by the experiment designs. For example, we have not investigated the assumption that incorporating basic text information for bridging inferences may compensate for the lack of previous knowledge in closing cohesion gaps since no manipulation of this aid was made. Thus, some of our assumptions might actually appear speculative to the reader. On the one hand, it is a limitation of the present work. On the other hand, the present work points in numerous directions of future research and makes the first steps in those directions.

We have mentioned some ideas for future designs in the published articles that are included in the present work such as the article presenting Experiment 2a. Therein, we suggest the potential of a pre-training with the aim to increase the generation accuracy for the cohesion generation task applied in Experiments 2a and 2b. The cohesion generation task highlights the cohesion gaps and thus compensates for the difficulty of recognizing them. However, the accuracy of closing the gaps highly depended on learners proficiencies (among others, previous knowledge). We consider the pre-training an important aid in supporting learners in closing the cohesion gaps (and reducing the dependency on proficiencies).

We have mentioned some important directions for future research also throughout this discussion. For example with respect to tracking of readers' attempts in detecting the cohesion gaps (e.g., via eye fixation patterns such as lookbacks), the link between reading skill and the ability to detect cohesion gaps, and the complex interacting pattern between the demands imposed by the text, learners' proficiencies, and learning aids. Especially the latter entails a plenty of possibilities. For example, within an experiment design, the demands of detecting and closing the gaps can be respectively increased vs. reduced. The proficiencies in turn can be measured, manipulated (e.g., via pre-training), or compensated by learning aids. In the following, we will describe one idea for future studies more detailed.

### *How to Compensate for the Disadvantages of Blocked Sequences?*

We are currently working on a follow-up investigation, in which the in/visibility of cohesion gaps should be manipulated independently of text sequence. Such a design would allow us to directly investigate the proposed explanatory mechanism behind the interleaving effects, namely the in/visibility of cohesion gaps. The in/visibility of cohesion gaps should be manipulated via test expectations, *specific* vs. *general*. A specific test expectation means that we will inform learners about the types of final test questions (on comparative and inductive reasoning). By this manipulation, we will instruct learners to pay attention to comparisons and underlying regularities while reading (cf. McCrudden & Schraw, 2007; Thiede et al., 2011). We presume that a specific test expectation (compared to the general one) only marginally raises the awareness of cohesion gaps in the interleaved conditions because an interleaved text may engage readers in spontaneous relational processing anyway. Specific test expectations should in contrast substantially enhance the awareness of cohesion gaps while reading a blocked text: Only then, readers should notice the lack of relational statements in the text. Thus, we expect the specific test-expectations to increase the level of cognitive engagement in the blocked conditions to a higher degree than in the interleaved conditions (as we have proposed, but failed to confirm in the Experiment 1b with prompted vs. spontaneous questioning factor).

Yet, would we actually expect the specific test expectations to augment the quality of inferences and learning (in terms of comparative and inductive reasoning) in the blocked conditions? It should depend on whether learners would be able to close the cohesion gaps. Given that learners in the experiment would lack the necessary knowledge to close the gaps, but the text itself provides the basic information for bridging inferences,[19] the moderated

[19] An experiment design can also apply the manipulation of supporting learners in closing the cohesion gaps, that is, providing basic text information for bridging inferences vs. not providing. If no support for closing the cohesion gaps is provided, readers who lack the previous knowledge are expected to take no advantage of

impact of sequence and test-expectations on learning should depend on the extent of *navigational control* while reading the text.

*Navigational control* herein means whether learners have the opportunity to jump across the paragraphs. Navigational control is provided simply when reading a printed text (as has been the case in Experiment 1a), when the whole text is presented at once, or within a digital environment giving learners the option to frequently switch across the slides (Naumann et al., 2007). The latter option allows researchers to record navigational behavior, that is, the pattern and frequency of switches. Navigational control is limited when the paragraphs are presented on separate slides and learners have no other options than to click on *continue*.

When navigational control is limited, readers are probably less able to compensate for a further disadvantage of blocking (beyond making cohesion gaps less noticeable), namely making bridging inferences more challenging due to greater distances across comparable information units. Hence, we expect that even if readers of a blocked text become aware of cohesion gaps by specific test expectations, they are still limited in their opportunities to bridge comparable information units to close the cohesion gaps when navigational control is limited. We thus expect the readers of a blocked text to benefit from specific test expectations only if navigational control is provided. Herein, specific test expectations would compensate readers of a blocked text for the invisibility of cohesion gaps, but navigational control would compensate them for the greater difficulty to close the gaps.

cohesion gaps, irrespective of whether the cohesion gaps are highlighted. However, whether learners with a high level of previous knowledge take advantage of cohesion gaps is expected to depend on whether the cohesion gaps are highlighted.

## Educational Implications

The present work makes an educationally relevant contribution to research on learning from more or less non-cohesive expository texts. A tailored educational recommendation should be made by considering 1) particular demands imposed by the expository text materials 2) whether learner's proficiencies are sufficient for overcoming those demands, and 3) compensatory potential of learning aids for the lack of particular proficiencies.

An expository text usually entails more or less cohesion gaps. Cohesion gaps promote relational processing if learners succeed in both recognizing and closing the cohesion gaps. To overcome the demand of recognizing the gaps, learners might rely on their reading skill, but the ability to close the gaps depends on learners' previous knowledge. Learning aids should compensate for the *invisibility* of cohesion gaps and the lack of previous knowledge respectively. Higher skilled readers with a high level of previous knowledge thus do not require any aids in learning from non-cohesive expository texts. However, cohesion gaps might be often overlooked by poor readers. Thus, especially poor readers might require learning aids that compensate for the *invisibility* of cohesion gaps by highlighting them. *Interleaved sequence* (i.e., juxtaposing of comparable information in text) and *causal cohesion generation* (i.e., fill-in-the-blank task using connectives) are two tools making cohesion gaps visible. Moreover, even if a gap was recognized, low knowledge readers might often fail to close the gaps. Then, low level of previous knowledge needs to be compensated by incorporating basic text information that can be bridged by readers to close the gaps independently of their previous knowledge. In a nutshell: The *difficulty* imposed by cohesion gaps is *desirable* in two cases, either when learners have the necessary proficiencies or – if not – when they receive learning aids that compensate for the invisibility of cohesion gaps and the lack of ability to close them.

## References

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction*, *17*(3), 286–303. https://doi.org/10.1016/j.learninstruc.2007.02.004

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.

Allen, L. K., McNamara, D. S., & McCrudden, M. T. (2015). Change your mind: Investigating the effects of self-explanation in the resolution of misconceptions. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 78–83). Pasadena, CA.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295.

Asterhan, C. S. C., & Resnick, M. (2020). Refutation texts and argumentation for conceptual change: A winning or a redundant combination? *Learning and Instruction*, *65*. https://doi.org/10.1016/j.learninstruc.2019.101265

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

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392–402. https://doi.org/10.3758/s13421-012-0272-7

Bjork, E. L., & Bjork, R. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed., pp. 59–68). Worth.

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

Britt, M. A., Richter, T., & Rouet, J.-F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist*, *49*(2), 104–122. https://doi.org/10.1080/00461520.2014.916217

Bugg, J. M., & McDaniel, M. A. (2012). Selective benefits of question self-generation and answering for remembering expository text. *Journal of Educational Psychology*, *104*(4), 922–931.

Canestrelli, A. R., Mak, W. M., & Sanders, T. J. M. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, *28*(9), 1394–1413. https://doi.org/10.1080/01690965.2012.685885

Carvalho, P. F., Braithwaite, D. W., de Leeuw, J. R., Motz, B. A., & Goldstone, R. L. (2016). An in vivo study of self-regulated study sequencing in introductory psychology courses. *PloS One*, *11*(3), 1-16. https://doi.org/10.1371/journal.pone.0152115

Catrysse, L., Gijbels, D., Donche, V., Maeyer, S. de, van den Bossche, P., & Gommers, L. (2016). Mapping processing strategies in learning from expository text: An exploratory eye tracking study followed by a cued recall. *Frontline Learning Research*, *4*(1), 1–16. https://doi.org/10.14786/flr.v4i1.192

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, *107*(3), 689–704. https://doi.org/10.1037/edu0000018

Chen, O., Kalyuga, S., & Sweller, J. (2016). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction*, *45*, 20–30. https://doi.org/10.1016/j.learninstruc.2016.06.007

Cohen, M. S., Yan, V. X., Halamish, V., & Bjork, R. A. (2013). Do students think that difficult or valuable materials should be restudied sooner rather than later? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(6), 1682–1696. https://doi.org/10.1037/a0032425

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology*, *80*(4), 448–456.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*(1), 1–53. https://doi.org/10.1080/01638539809545019

Cozijn, R. (2000). *Integration and inference in understanding causal sentences*. Faculteit der Letteren, KUB.

Cozijn, R., Noordman, L. G. M., & Vonk, W. (2011). Propositional integration and world-knowledge inference: Processes in understanding because sentences. *Discourse Processes*, *48*, 475–500.

de Jonge, M., Tabbers, H. K., & Rikers, R. M. J. P. (2015). The effect of testing on the retention of coherent and incoherent text material. *Educational Psychology Review*, *27*(2), 305–315. https://doi.org/10.1007/s10648-015-9300-z

Degand, L., Lefevre, N., & Bestgen, Y. (1999). The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design*, *1*, 39–51.

Degand, L., & Sanders, T. J. M. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing*, *15*, 739–757.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, *32*(6), 945–955.

Dirkx, K. J. H., Camp, G., Kester, L., & Kirschner, P. A. (2019). Do secondary school students make use of effective study strategies when they study on their own? *Applied Cognitive Psychology*, 1–6. DOI: 10.1002/acp.3584

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 and the Public Interest*, *14*, 4–58.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fleischer, J., Wirth, J., & Leutner, D. (2014). Effekte der kontextuellen Einkleidung von Testaufgaben auf die Schülerleistungen im analytischen Problemlösen und in der Mathematik. *Zeitschrift Für Pädagogische Psychologie*, *28*(4), 207–227. https://doi.org/10.1024/1010-0652/a000135

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language*, *27*, 235–244.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Lawrence Erlbaum.

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137–153. https://doi.org/10.1016/j.cognition.2014.12.001

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*(1), 103–128. https://doi.org/10.1037/0022-0663.93.1.103

Helsdingen, A., van Gog, T., & van Merriënboer, J. J. G. (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgment task. *Journal of Educational Psychology*, *103*(2), 383–398. https://doi.org/10.1037/a0022370

Holley, C. D., Dansereau, D. F., McDonald, B. A., Garland, J. C., & Collins, K. W. (1979). Evaluation of a hierarchical mapping technique as an aid to prose processing. *Contemporary Educational Psychology*, *4*, 227–237.

Hyönä, J., Lorch, R. F., Jr., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*(1), 44–55. https://doi.org/10.1037//0022-0663.94.1.44

Ionas, I. G., Cernusca, D., & Collier, H. L. (2012). Prior knowledge influence on self-explanation effectiveness when solving problems: An exploratory study in science learning. *International Journal of Teaching and Learning in Higher Education*, *24*(3), 349–358.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory.* Erlbaum.

Kalyuga, S. (2006). Assessment of learners' organised knowledge structures in adaptive learning environments. *Applied Cognitive Psychology*, *20*(3), 333–342. https://doi.org/10.1002/acp.1249

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

Kalyuga, S., & Singh, A.-M. (2015). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review.* Advance online publication. https://doi.org/10.1007/s10648-015-9352-0

Kamalski, J., Sanders, T. J. M., & Lentz, L. (2008). Coherence marking, prior knowledge, and comprehension of informative persuasive texts: Sorting things out. *Discourse Processes*, *45*, 323–345.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in students learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479.

Kendeou, P., Muis, K. R., & Fulton, S. (2011). Reader and text factors in reading comprehension processes. *Journal of Research in Reading*, *34*(4), 365–383. https://doi.org/10.1111/j.1467-9817.2010.01436.x

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction*, *7*(3), 161–195.

Kintsch, W. (1988). The role of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*(2), 163–182.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037//0033-295X.85.5.363

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimmy, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*, 133–159.

Koriat, A., & Bjork, R. A. (2006a). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition*, *34*(5), 959–972.

Koriat, A., & Bjork, R. A. (2006b). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1133–1145. https://doi.org/10.1037/0278-7393.32.5.1133

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*, 585–592.

Kornell, N., & Vaughn, K. E. (2018). In inductive category learning, people simultaneously block and space their studying using a strategy of being thorough and fair. *Archives of Scientific Psychology*, *6*(1), 138–147. https://doi.org/10.1037/arc0000042

Kowalski, P., & Kujawski Taylor, A. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology*, *36*, 153–159. DOI: 10.1080/00986280902959986

Kraal, A., Koornneef, A. W., Saab, N., & van den Broek, P. (2017). Processing of expository and narrative texts by low- and high-comprehending children. *Reading and Writing*. DOI 10.1007/s11145-017-9789-2

Kurby, C., Magliano, J. P., Dandotkar, S., Woehrle, J., Gilliam, S., & McNamara, D. S. (2012). Changing how students process and comprehend texts with computer-based self-explanation training. *Faculty Research and Creative Activity*, *25*, 1–48.

Lagerwerf, L. (1998). *Causal connectives have presuppositions: Effects on coherence and discourse structure*. Holland Academic Graphics.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of*

*Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787–1794. https://doi.org/10.1037/xlm0000012

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: an aptitude-treatment-interaction study. *Metacognition and Learning*, *11*(1), 89–105. https://doi.org/10.1007/s11409-015-9149-z

Leppink, J., Paas, F. G. W. C., Van der Vleuten, C. P. M., van Gog, T., & van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction*, *18*(4), 525–556.

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction*, *34*, 74–85. https://doi.org/10.1016/j.learninstruc.2014.08.004

Lorch, R. F., Jr. (2015). What about expository text? In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 348–361). Cambridge University Press.

Louwerse, M. (2001). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, *12*(3), 291–315.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, *21*(3), 251–283.

Maier, J., Richter, T., & Britt, M. A. (2018). Cognitive processes underlying the text-belief consistency effect: An eye-movement study. *Applied Cognitive Psychology*, *32*, 171–185. DOI: 10.1002/acp.3391

Maury, P., & Teisserenc, A. (2005). The role of connectives in science text comprehension and memory. *Language and Cognitive Processes*, *20*(3), 489–512.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press.

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

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, *19*(2), 113–139. https://doi.org/10.1007/s10648-006-9010-7

McCrudden, M. T., Schraw, G., Lehman, S., & Poliquin, A. (2007). The effect of causal diagrams on text learning. *Contemporary Educational Psychology*, *32*(3), 367–388. https://doi.org/10.1016/j.cedpsych.2005.11.002

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–199). Taylor & Francis.

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

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, *20*, 516–522.

McKeown, M. G., Beck, I. L., Sinatra, G. M., & Loxterman, J. A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly*, *27*(1), 78–93.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–466. https://doi.org/10.1037//0033-295X.99.3.440

McNamara, D. S. (1992). The generation effect: A detailed analysis of the role of semantic processing. *Technical Report*, *2*, 1–48.

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

McNamara, D. S., Kintsch, E., Butler Songer, N., & 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.

Meyer, B. J. F. (1975). *The organization of prose and its effect on memory*. North-Holland.

Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, *21*(1), 121–143.

Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 237–242.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*, 453–465.

Naumann, J., Richter, T., Flender, J., Christmann, U., & Groeben, N. (2007). Signaling in expository hypertexts compensates for deficits in reading skill. *Journal of Educational Psychology*, *99*(4), 791–807. https://doi.org/10.1037/0022-0663.99.4.791

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*, 413–448.

Nguyen, K., & McDaniel, M. A. (2016). The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology. Applied*, *22*(1), 59–71. https://doi.org/10.1037/xap0000066

Noordman, L. G. M., & Vonk, W. (1997). The different functions of a conjunction in constructing a representation of the discourse. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships: Studies in the production and comprehension of text* (pp. 75–93). Lawrence Erlbaum.

Noordman, L. G. M., Vonk, W., & Kempf, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, *31*, 573–590.

O'Brien, E. J., & Myers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 12–21.

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.

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*, 228–242.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*(4), 737–743.

Prinz, A., Golke, S., & Wittwer, J. (2019). Refutation texts compensate for detrimental effects of misconceptions on comprehension and metacomprehension accuracy and support transfer. *Journal of Educational Psychology*, *111*(6), 957–981. https://doi.org/10.1037/edu0000329

Reinhard, M.-A., Weissgerber, S. C., & Wenzel, K. (2019). Performance expectancies moderate the effectiveness of more or less generative activities over time. *Frontiers in Psychology*, *10*(1623), 1–19. https://doi.org/10.3389/fpsyg.2019.01623

Roelle, J., Berthold, K., & Renkl, A. (2014). Two instructional aids to optimise processing and learning from instructional explanations. *Instructional Science*, *42*, 207–228. DOI 10.1007/s11251-013-9277-2

Roelle, J., & Nückles, M. (2019). Generative learning vs. retrieval practice in learning from text: The cohesion and elaboration of the text matters. *Journal of Educational Psychology*, *111*(8), 1341–1361. https://doi.org/10.1037/edu0000345

Roelle, J., Schmidt, E. M., Buchau, A., & Berthold, K. (2017). Effects of informing learners about the dangers of making overconfident judgments of learning. *Journal of Educational Psychology*, *109*(1), 99–117. https://doi.org/10.1037/edu0000132

Saffran, A., Barchfeld, P., Alibali, M. W., Reiss, K., & Sodian Beate (2019). Children's interpretations of covariation data: Explanations reveal understanding of relevant comparisons. *Learning and Instruction*, *59*, 13–20.

Sana, F., Yan, V. X., Kim, J. A., Bjork, E. L., & Bjork, R. A. (2018). Does working memory capacity moderate the interleaving benefit? *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1016/j.jarmac.2018.05.005

Sanders, T., Land, J., & Mulder, G. (2007). Linguistic markers of coherence improve text comprehension in functional contexts. *Information Design Journal*, *15*(3), 219–235.

Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in processing. *Discourse Processes*, *29*(1), 37–60.

Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*(1), 1–35. https://doi.org/10.1080/01638539209544800

Schindler, J., Schindler, S., & Reinhard, M.-A. (2019). Effectiveness of self-generation during learning is dependent on individual differences in need for cognition. *Frontline Learning Research*, *7*(2), 23–39. https://doi.org/10.14786/flr.v7i2.407

Schnotz, W. (1982). How do different readers learn with different text organizations. In A. Flammer & W. Kintsch (Eds.), *Discourse processing* (pp. 87–97). North-Holland.

Schnotz, W. (1984). Comparative instructional text organization. In H. Mandel, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 53–81). Erlbaum.

Seifert, T. L. (1994). Enhancing memory for main ideas using elaborative interrogation. *Contemporary Educational Psychology*, *19*, 360–366.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*(6), 592–604.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review*, *20*, 356–363.

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. https://doi.org/10.1037/0022-0663.95.1.66

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology*, *81*, 264–273.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.

Traxler, M. J., Sanford, A. J., Aked, J. P., & Moxey, L. M. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 88–101.

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. O’Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 94–121). Cambridge University Press.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

van Silfhout, G., Evers-Vermeul, J., Mak, W. M., & Sanders, T. J. M. (2014). Connectives and layout as processing signals: How textual features affect student's processing and text representation. *Journal of Educational Psychology*, *106*(4), 1036–1048.

van Silfhout, G., Evers-Vermeul, J., & Sanders, T. J. M. (2014). Establishing coherence in schoolbook texts: How connectives and layout affect students’ text comprehension. *Dutch Journal of Applied Linguistics*, *3*(1), 1–29.

van Silfhout, G., Evers-Vermeul, J., & Sanders, T. J. M. (2015). Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, *52*(1), 47–76.

Vidal-Abarca, E., & Sanjose, V. (1998). Levels of comprehension of scientific prose: The role of text variables. *Learning and Instruction*, *8*(3), 215–233.

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.

Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, *36*(2), 109–129. DOI: 10.1207/S15326950DP3602_2

Wiley, J., & McGuinness, C. (2004). The interactive effects of prior knowledge and text structure on memory for cognitive psychology texts. *British Journal of Educational Psychology*, 497–514.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376.

Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (2nd ed., pp. 413–432). Cambridge University Press.

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918–933. https://doi.org/10.1037/xge0000177

Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, E. L., & Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied*, *23*(4), 403–416. https://doi.org/10.1037/xap0000139

Zulkiply, N. (2013). Effect of interleaving exemplars presented as auditory text on long-term retention in inductive learning. *Procedia - Social and Behavioral Sciences*, *97*, 238–245. https://doi.org/10.1016/j.sbspro.2013.10.228

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*(1), 16–27. https://doi.org/10.3758/s13421-012-0238-9

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215–221.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185.

## Acknowledgements

Die abschließenden Worte meiner Doktorarbeit möchte ich auf Deutsch schreiben.

Ich hätte niemals gedacht, dass ich in 5 Minuten Entfernung von der Goethe Schule, an der ich mein Abi machte, promovieren werde; dass ich nach fast 10 Jahren wieder nach Kassel ziehe und Kassel ganz neu für mich entdecke. Aber *Wünschenswerte Erschwernisse beim Lernen* haben es möglich gemacht.

Die allererste Person, der ich in diesem Zuge danken möchte, ist **Martin Hänze**. Der Freiraum, den du mir gelassen hast, die Zeit, finanzielle Ressourcen und Möglichkeiten, die ich hatte im Verlauf der Loewe-Jahre, und nicht zuletzt deine Spontanität und Offenheit für meine Ideen – all das ist von unschätzbarem Wert für die Entwicklung dieser Arbeit sowie meine eigene Entwicklung auf diesem höchst interessanten Forschungsgebiet gewesen. Ich habe über die Jahre hinweg kein einziges Mal das Gefühl gehabt, man schaue mir über die Schulter. Ich danke dir vor allem für dein Vertrauen.

Ich blicke zurück auf eine spannende Zeit. Es waren z.B. solche Momente, in denen man zum ersten Mal Wissenschaftlern begegnete, die man vorher nur aus Artikeln kannte und zu ihnen wie zu Stars aufblickte, die aber letztendlich zu vertrauten, altbekannten Gesichtern wurden. So war es z.B. in Bochum 2016 auf der International Cognitive Load Theory Conference und 2018 in Wollongong.

Meine Forschungstätigkeit sehe ich mit dem Abschluss der Dissertation erst „am Aufkeimen". Ich habe in den vergangenen Jahren zahlreiche Erkenntnisse sammeln können. Es sind aber vor allem offene Fragen, ausgearbeitete Pläne für zukünftige Untersuchungen und Manuskripte in Vorbereitung, die meine Motivation untermauern, mit der Forschung weiterzumachen. In diesem Zuge möchte ich **Julian Roelle** für die Möglichkeit danken, meine wissenschaftliche Karriere fortzusetzen. Die Aussicht auf die Stelle gab mir den richtigen Schwung, endlich zum Abschluss zu kommen. Die letzten zwei Monate vor der

Abgabe waren zwar stressvoll wegen zeitlicher Knappheit, aber erfüllt von Zuversicht, *dass* es weitergeht. Ich möchte an dieser Stelle ferner zum Ausdruck bringen, dass ich mich außerordentlich freue, mit Menschen zusammen zu arbeiten, vor deren Forschung ich Hochachtung habe.

**Ralf Rummer**, dir danke ich ganz herzlich vor allem für die *Brücken*, die du für mich gebaut hast. Eine solche Brücke hast du z.B. zwischen Pädagogischer und Kognitiver Psychologie geschlagen. Ich danke dir für die Unterstützung in der Zeit, die finanziell überbrückt werden musste bis meine Dissertation abgeschlossen war. Eine andere Brücke führt direkt nach Bochum. Ich bin zuversichtlich, dass die Zukunft viele gemeinsamen Projekte bringt.

Ich schulde einen großen Dank den studentischen Mitarbeiter*innen, die meine Forschung, die weit über die Dissertationsarbeit hinausgeht, unterstützt haben und ohne deren Fleiß, Organisationstalent und Expertise ich aufgeschmissen wäre: **Christoph Kissel**, **Marion Kritsch**, **Ida Brandenburger**, **Jan Steinhauer**, **Marcia Steinhäuser**, **Amira Mehr** und andere. Ich danke im gleichen Atemzug Studierenden, deren Qualifikationsarbeiten und/oder Empra-Projekte ich (mit-)betreuen durfte: **Daria Mundt** (nun Kollegin), **Luka Maria Niedling**, **Sebastian Vogel**, **Anika Simon**, **David Wagner**, **Klara Rogge**, **Pia Göller**, **Sina Lippold**, **Manuel Hemmerich**, **Therese Watolla**, **Nico Harhoff**, **David Schlarbaum**, **Jessika Klassen**, **Marie Beil** und viele andere. Es hat mich immer wieder in der Annahme bekräftigt, solche Anlässe als Möglichkeit für gemeinsames Forschen auf Augenhöhe zu nutzen.

Es ist toll, wenn Kollegen zu Freunden werden: **Sophia Christin Weissgerber**, **Matthias Brunmair**, **Benjamin Harders** – es ist wunderbar, euch nicht nur als kompetente Gesprächspartner, sondern vor allem als Freunde gewonnen zu haben. Ich kann nur hoffen, dass weder unsere Diskussionen, noch unsere Brettspielpartien abreißen. Ich wäre glücklich, mit euch gemeinsam in der Zukunft zu forschen.

Ein besonderer Dank geht an **Matthias Mai** – unsere thematischen Gespräche seit der gemeinsamen Studienzeit in Hildesheim (und Salzburg) sind essentiell für die Art und Weise, wie ich heute denke. Schade, dass wir die *Weltformel* immer noch nicht gefunden haben. Aber wir dürfen nicht aufgeben! Bis es soweit ist, müssen wir viel gemeinsam nachdenken und forschen – es gibt noch so viel zu entdecken!

Bleiben wir kurz bei Hildesheim – der Stadt, in der ich meinen Bachelor- und Masterabschluss gemacht sowie meine erste Stelle an der Universität hatte. Hier muss ich drei weitere Menschen erwähnen. **Werner Greve** – es gibt kaum einen Menschen, der mich mehr für die Wissenschaft inspiriert hat und von dem ich im Verlauf meines Studiums mehr gelernt habe. **Elke Montanari** – es war sehr bereichernd für mich, mit dir gleich im Anschluss an mein Studium zur Entwicklung mehrsprachigen Wortschatzes zu forschen. Auf diese Weise habe ich einen Zugang gefunden, ein Phänomen, das mich unmittelbar betrifft – nämlich die Mehrsprachigkeit und bilinguale Identitätskonstruktion – wissenschaftlich zu adressieren. Ich danke weiterhin meinem *flatmate buddy* aus Hildesheimer Zeit **Ahmed Zaher Elgohary** in diesem Rahmen dafür, dass du mein Englisch ganz beiläufig gepuscht hast. Noch wenige Jahre zuvor konnte ich mir nicht ausmalen, meine Doktorarbeit komplett auf Englisch zu verfassen (also *komplett*, wenn man von der Dankesrede absieht). Ich möchte an der Stelle nicht unerwähnt lassen, dass meine Literaturrecherche zu *causal cohesion* zu einem großen Teil in der Bibliothek von Alexandria und in den Cafés von Kairo, wo Naguib Mahfuz an seinen Büchern gearbeitet hat, erfolgte.

Und wenn ich schon bei den Cafés bin, gilt mein Dank zuletzt den Mitarbeiter*innen zahlreicher gemütlichen Cafés, die ich in den letzten Jahren Tag für Tag besucht habe. Um ein paar in Kassel aufzulisten: Kollektivcafé Kurbad, Café Desasta, Stamm Kaffee, Christian

Bach, Meyerbeers Coffee, Rokkeberg, Bistro Hahn, Suppenplantage, Café Seegert, Sapori di Italia, Buch Oase, Holy Nosh, Boulangerie-Patisserie, Falada mit Grimms Garten, Salotti, Café im Fridericianum und andere. In Hildesheim waren es vor allem das kleine Röstwerk, Black Apron und Le Garçon. Danke für die Atmosphäre, leckeren Kuchen und den Geruch frisch gebrühten Kaffes in der Luft. Ich brauche es, damit meine Gedanken in Gang kommen. Es schmerzt mich zu denken, was die Gastronomie im Zuge der Pandemie alles einstecken musste und vermutlich noch erleiden wird. Ich kann nur hoffen, dass die Cafés in dieser schwierigen Zeit es schaffen, sich über Wasser zu halten. Mein beruflicher Erfolg wird maßgeblich davon abhängen, dass es euch gibt.