Received: 7 August 2019 | Revised: 14 August 2020 | Accepted: 13 October 2020
DOI: 10.1002/acp.3761

RESEARCH ARTICLE

WILEY

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

Roman Abel | **Luka Maria Niedling** | **Martin Hänze**

Institute of Psychology, University of Kassel, Kassel, Germany

Correspondence
Roman Abel, Department of General Psychology, University of Kassel, Holländische Str. 36-38, 34127 Kassel, Germany.
Email: r_abel@uni-kassel.de

Funding information
Hessian Ministry for Science and the Arts

Abstract

Recent studies on text sequencing found learning advantages of interleaving over blocking in terms of high-level inferences. We conducted a 2 × 2 × 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. 1-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
inductive learning, inductive reasoning, interleaving effect, question generation, text comprehension

1 | INTRODUCTION

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 (1982, 1984) experiments on reading texts about two forms of psychotherapy presented in different sequences, 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 Jr., & 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 et al. (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, for example, 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.[1]

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 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.

2 | PRESENT STUDY

We investigated the following three research questions: (1) whether the main finding of Abel et al. (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 1 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 1-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 et al. (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 1-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 1-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*).

3 | METHOD

We conducted a 2 × 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 1-week delay. The research was conducted in compliance with the Declaration of Helsinki and ethical standards of the DGPS (German Society of Psychology).

3.1 | Sample

Our laboratory-experiment included 117 volunteer participants. Three participants, who did not attend the final test after 1-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.

3.2 | 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 1060 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 (cf. 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.

3.3 | 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.

3.4 | 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.

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 Fin whale: 80–100 Blue whale: 90, single > 200 Sperm whale: ca. 60 Narwhale: 30–40 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).

3.5 | 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 1-week delay.

4 | RESULTS

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

4.1 | Learning processes

We computed separate ANOVAs (analyses of variance) 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).

4.1.1 | 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* (confidence interval) [0.55, 1.75], *MD* (mean difference) = 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], $MD = -0.18$, $SE = 0.31$.

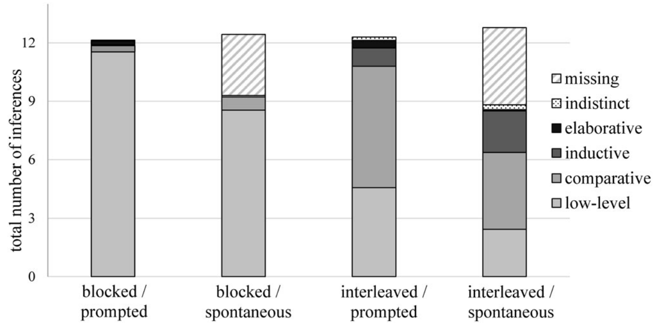

FIGURE 1 The distribution of averaged frequencies in generating inferences of different cognitive levels while reading (collapsed for both reading cycles) as a function of sequence (blocked vs. interleaved) and self-questioning (spontaneous vs. prompted). Participants' text box entries were coded as inferences of different cognitive levels: Either low-level, comparative, inductive, or elaborative. Indistinct and missing responses were also recorded. each bar consists of 12 inferences from the six paragraphs by two reading cycles. A higher number than 12 occurred when a text box entry was assigned to more than one cognitive level

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.

4.1.2 | 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.

4.1.3 | 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, $F < 1$.

4.1.4 | 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, $F < 1$.

4.1.5 | 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 12 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, $F < 1$.

4.2 | 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. 1 week later, T2).

4.2.1 | 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

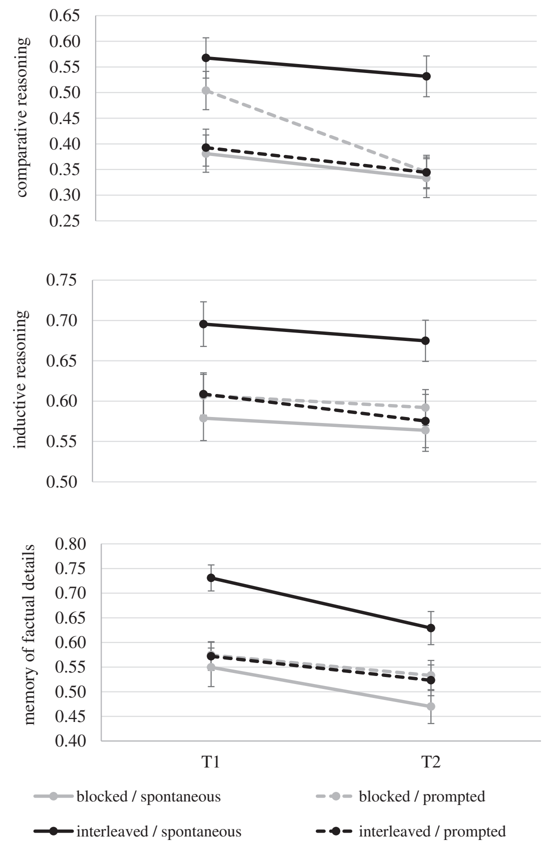

FIGURE 2 Proportion of correctly solved questions on comparative reasoning (above), inductive reasoning (middle), and memorization of factual details (below) in the final test as a function of sequence (interleaved vs. blocked), self-questioning (spontaneous vs. prompted), and testing delay (T1 vs. T2).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

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, $F(1,110) = 1.80$, $p = .183$, $\eta_p^2 = .02$.

4.2.2 | 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 leads 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, *F*s < 1.

4.2.3 | 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, $F < 1$.

4.3 | 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.[2] 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 $m = 5000$.

TABLE 2 Pearson correlations between dependent measures

Learning processes	2	3	4	5	6	7	8	9	10	11
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: *$p < .05$. **$p < .01$.

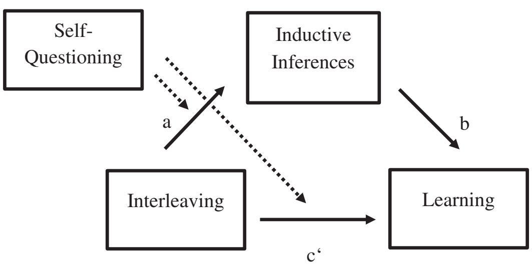

FIGURE 3 The moderated-mediation model. Effect of sequence (interleaving vs. blocking) on learning (comparative reasoning, inductive reasoning, and memorization of factual details; collapsed for T1 and T2) mediated by making inductive inferences and moderated by self-questioning (spontaneous vs. prompted)

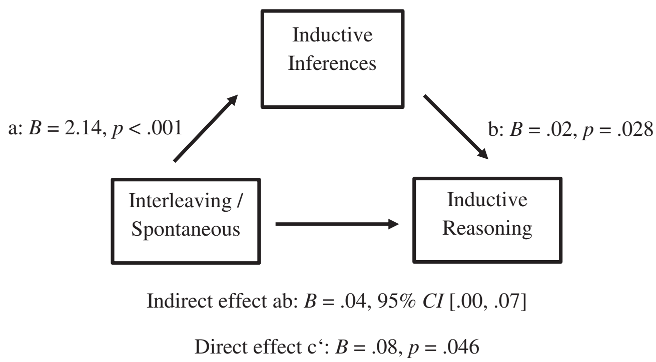

FIGURE 4 Mediation model for spontaneous self-questioning conditions. Effect of sequence (interleaving vs. blocking) on inductive reasoning (collapsed for T1 and T2) mediated by making inductive inferences. Note that this pattern of results (a significant indirect effect and a significant direct effect) 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

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).

4.3.1 | 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%

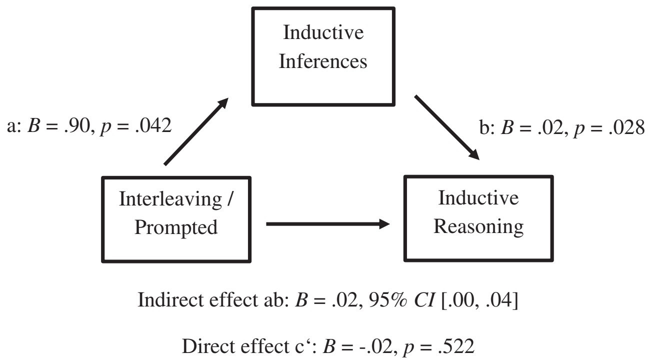

FIGURE 5 Mediation model for prompted self-questioning conditions. Effect of sequence (interleaving vs. blocking) on inductive reasoning (collapsed for T1 and T2) mediated by making inductive inferences. Note that this pattern of results (a significant indirect effect but no direct effect) also applies 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

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.

4.3.2 | 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.

4.3.3 | 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.

5 | DISCUSSION

The present study served three purposes. First, we wanted to replicate the results from the previous research conducted by Abel et al. (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 1 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 et al. (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 et al., 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 indicates 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*.[3] 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.[4] Merely paying attention to factual details without any construction and integration activity thus does not support memory.

5.1 | 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 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 et al., 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).[5]

5.2 | 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.

5.3 | 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, Bjork, & Bjork, 2016). In light of this reasoning, many book designers might design expository texts and nonfiction 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.

ACKNOWLEDGMENTS

This study was supported by a LOEWE grant from the Hessian Ministry for Science and the Arts entitled "desirable difficulties in learning" awarded to Martin Hänze. We thank Matthias Mai and Matthias Brunmair for stimulating discussions. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

AUTHOR CONTRIBUTIONS

RA developed the study concept. RA and LN contributed to the design of the study. LN organized the database. RA performed the statistical analysis and wrote the manuscript. MH contributed to manuscript revisions. All authors read and approved the submitted version.

The research was conducted in compliance with the Declaration of Helsinki and ethical standards of the DGPS (German Society of Psychology).

DATA AVAILABILITY STATEMENT

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

ORCID

Roman Abel https://orcid.org/0000-0002-4589-4306

ENDNOTES

[1] The pattern of results yielded by Abel et al. (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.

[2] 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, $B = -.00$, $p = .513$.

[3] *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.

[4] 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 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 to 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.

[5] 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], $MD = -.06$, $SE = .05$.

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. *Frontiers in Psychology*, *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: Cognitive Science Society.

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. 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, NY: 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*. Paper presented at the 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. https://doi.org/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. https://doi.org/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. 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, 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. https://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*, *35*, 218–227. https://doi.org/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. https://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. https://doi.org/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, The Netherlands: 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. *Research in Science Education*, *48*, 1273–1296. Retrieved from. https://doi.org/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), 7–18.

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, UK: 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. https://doi.org/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.

How to cite this article: Abel R, Niedling LM, Hänze M. Spontaneous inferential processing while reading interleaved expository texts enables learners to discover the underlying regularities. *Appl Cognit Psychol*. 2021;35:258–273. https://doi.org/10.1002/acp.3761

APPENDIX A.

TABLE A1 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.