



How to assist the students while learning from text? Effects of inserting adjunct questions on text processing

Alba Rubio¹ · Eduardo Vidal-Abarca¹ · Marian Serrano-Mendizábal¹

Received: 7 August 2021 / Accepted: 16 June 2022
© The Author(s) 2022

Abstract

This study analyzes the effect of text-inserted questions and post-text-reading questions, i.e., questions timing, on students' processing and learning when studying challenging texts. Seventy-six freshmen read two science texts and answered ten adjunct questions with the text available, being tested on learning 5 days afterwards. Questions were presented either after reading the whole text or inserted in the text after reading the relevant information. Online processing data were recorded while reading and searching the texts, and measures of processing strategies (i.e., paraphrases, elaborations) while answering the questions were collected. Compared to students in the post-reading condition, those in the inserted condition spent more time reading the text initially, were more efficient at searching for information in the text, and produced more accurate elaborations, all of which may explain why answering inserted questions in an available text were more effective in terms of learning than answering post-reading questions. Limitations and educational implications of these results are also discussed.

Keywords Adjunct questions · Question timing · Text processing · Learning · Comprehension

Introduction

Processing text information for learning purposes can be highly influenced by learning activities, such as question-answering. Questions can be understood as specific relevance instructions that help students generate reading goals, as well as access, locate and retrieve relevant resources of information in a systematic and strategic way (Cerdán et al., 2009; McCrudden & Schraw, 2007). These mental processes can be very helpful for learning purposes. Thus, questions can guide students on what text information to focus on and challenge them to infer relations between queried and answered information (Dirkx et al., 2015; Olson et al., 1985). Given the impact that questions have on text processing, when to present them in a question-answering scenario becomes an important issue.

✉ Alba Rubio
alba.rubio@uv.es

¹ ERI-Lectura, Department of Developmental and Educational Psychology, University of Valencia, Avda. Blasco Ibáñez 21, 46010 Valencia, Spain

The student's processing may be different when questions are presented while reading (i.e., inserted questions) or after reading the whole text (i.e., post-reading questions), and these differences in processing may affect the students' final learning outcome. In this regard, some studies have analyzed the effect of question timing on students' comprehension and learning (Andre & Womack, 1978; Cerdán et al., 2009; Kapp et al., 2015; Peverly & Wood, 2001; Philips et al., 2020; van den Broek et al., 2001; Weinstein et al., 2016), but they reveal no clear findings. While most of the researchers found that inserted questions contributed to greater understanding and learning compared to post-reading questions, some of them found no effect of question timing on students' learning. In addition, no study has recorded and examined in detail how students process the text and the questions under each condition. Thus, the goal of this study is to investigate how the timing of comprehension questions (i.e., inserted versus post-reading) influences college students' text processing and learning of complex conceptual knowledge.

Our study compares a condition in which the adjunct questions were presented after reading the whole text with another condition in which the questions were inserted in the text after reading question-relevant segments. In both cases, students had the text available while answering the questions. We used two science texts with challenging ideas about natural phenomena (e.g., the differences between heat and temperature) and recorded online processing. Our procedure is inspired by the paradigm of adjunct question research. Several reviews concluded that adjunct questions have positive effects on students' learning (Hamaker, 1986; Hamilton, 1985; Roediger & Karpicke, 2006), the so-called *adjunct question effect*. We assumed that inserted questions might improve the students' learning over post-reading questions, as they may support relevant text processing and inference making while building a mental representation of the text.

The role of adjunct questions for prose learning

Answering adjunct questions with an available text is a frequent learning activity in classrooms (Ness, 2011). Three main design features of adjunct questions research were generally assumed to be related to learning (Hamaker, 1986). The first one is the cognitive level of adjunct questions. Several taxonomies have been proposed (Goldman & Duran, 1988; Ozuru et al., 2007; Tawfik et al., 2020), but all of them share the distinction between low- and high-level questions. Most studies conclude that high-level adjunct questions improve deep comprehension and learning (Cerdán et al., 2009; Jensen et al., 2014), whereas others claim that a mix of question types offers the best results (Agarwal, 2019): while low-level questions focus the reader's attention on the recall and understanding of specific text ideas, high-level questions require the reader to analyze, apply, or evaluate textual information. The second feature is the issue of question placement. Recent research has focused on comparing inserted questions versus presenting questions after reading the whole text. This research yields mixed results. Whereas many studies found that inserted questions were more effective than massed post-questions (Kapp et al., 2015; van den Broek et al., 2001; Peverly & Wood, 2001; Philips et al., 2020), other studies found that both produced comparable benefits on final learning (Uner & Roediger, 2018; Weinstein et al., 2016). The last feature affects the relation between the adjunct and the transfer test. The adjunct question effect was mainly found on the retention of related information asked in adjunct questions, but it is rarely observed on the retention of unrelated information (Chan et al., 2006; Dirks et al., 2015; Hamilton, 1985; van den Broek et al., 2001).

Apart from the discrepancies pointed out earlier, adjunct questions research has some limitations. First, most of the classic research did not use long texts requiring the construction of a coherent mental model of the whole passage (e.g., Hamaker, 1986; Hamilton, 1985). When adjunct questions research used long texts, the different text sections were relatively disconnected one from the other, so understanding a section was not necessary to understand any other (e.g., Cerdán et al., 2009; Uner & Roediger, 2018). Second, few studies allowed students to reread the texts while answering the questions, which differ from usual conditions in school settings (Ness, 2011). Third, most of the questions used were quite limited in scope; for instance, low-level questions referred to proper names, places or similar, which demanded extremely superficial processing, but not understanding ideas that required semantic processing (e.g., Weinstein et al., 2016). Further, most of the higher order questions required students to understand ideas explicit in the texts, but not to apply text information to a new situation or to make deep inferences (e.g., van den Broek et al., 2001). Finally, no studies have collected online processing measures that could explain the effect of the question timing. Our study was designed to overcome these important limitations.

Constructing text meaning for learning: how adjunct questions modify processing strategies

It is assumed that understanding rests in the construction of a coherent text representation that involves representing the explicit text information (i.e., text-based representation) plus incorporating text ideas into the reader's prior knowledge, called a situation-model level of understanding (Kintsch, 1998). Considering that learning results from the modification of the readers' knowledge structures for the domain, we can accept that it occurs when readers reach an appropriate situation model level of understanding (Coté et al., 1998).

According to these ideas, when constructing meaning from a text, readers may use processing strategies such as paraphrasing and elaborations (Coté et al., 1998). Paraphrases remain close to the explicit text meaning and are relatively superficial levels of comprehension, whereas elaborations involve generating text-connecting inferences and going beyond the text by integrating prior knowledge (McNamara, 2004). Therefore, paraphrasing may lead to a good-quality text-based, whereas inferences are needed to achieve deep comprehension and learning (Coté et al., 1998). However, although both processing strategies contribute to text comprehension, incorrect paraphrases have been associated with lower comprehension levels. Incorrect elaborations are compatible with the text-based level of understanding, although they do not contribute to deep comprehension (McNamara, 2004).

When reading expository texts, engaging in deep comprehension processes is not common, not even for college students (Endres et al., 2017; Linderholm & van den Broek, 2002). In this context, adjunct questions are frequently used to help the students construct a coherent mental representation of the text meaning by promoting inferences during and after reading. To understand how questions influence comprehension, it is important to place question-answering processes within the goal-focusing model (McCrudden & Schraw, 2007). This model explains how questions affect text processing, as they may be considered relevance processing instructions. When reading a question, readers form a task model in their mind, which is a representation of the question goal and the means to achieve it (Rouet et al., 2017). This representation guides the student in finding question-relevant information and strategically processing the text (Cerdán & Vidal-Abarca, 2008;

Dirkx et al., 2015). Reading (or rereading) the text with a goal in mind increases the accessibility of relevant background knowledge (McCrudden & Schraw, 2007), making readers more likely to elaborate on text information and increasing the likelihood of deep comprehension.

Assuming that the benefits of adjunct questions rely on their ability to direct the students' attention to relevant information, inserting questions after reading one or two paragraphs (i.e., while constructing a representation of that information) would have more profound effects on the processing strategies than questions placed after reading the whole text, when the representation has already been constructed and is more resistant to modifications (van den Broek et al., 2001; van Oostendorp & Goldman, 1999). However, although it is clear that adjunct questions have an impact on text processing, no studies have explored how the position of these questions (i.e., inserted vs. post-reading) may modify the processing strategies underlying learning.

Reading behavior when answering adjunct questions

Comprehension processes can be passive (i.e., automatic) and reader-initiated (i.e., controlled). Passive processes always take place through the unrestricted spread of activation mechanisms, but reader-initiated processes operate in a restricted manner as a function of the reader's *standards of coherence* and the information returned from passive processes (van den Broek & Helder, 2017). A reader's standards of coherence are the (often implicit) criteria that a reader has for what constitutes adequate comprehension and coherence in a particular reading situation (van den Broek et al., 2011; van den Broek et al., 1995). Standards can be modified by specific reading instructions. For instance, providing instructions for learning from a text elicits more processes for building coherence (e.g., connecting text ideas and elaborative inferences) than reading for entertainment (van den Broek et al., 2001). Answering adjunct questions with the text available may also modify the standards of coherence as they may lead to specific reader-initiated processes (e.g., rereading decisions while searching the text).

Several studies have found that rereading decisions while searching predict text comprehension (Cerdán et al., 2009, 2011; Gil et al., 2015; Mañá et al., 2009; Máñez et al., 2022). For example, Mañá et al. (2009) found that the number of visits to relevant information and its use explained significant variance in question-answering performance beyond general comprehension skills. Similarly, Máñez et al. (2022) found a strong correlation between question-answering performance and the percentage of relevant information selected to answer the questions.

Nevertheless, reading and search processes while question-answering may be different depending on question timing (McCrudden & Schraw, 2007). When answering inserted questions, the short delay between reading and questioning facilitates the access and retrieval of text information (e.g., Carrier & Fautsch-Patridge, 1981; Hamaker, 1986; Rickards & Di Vesta, 1974; Schumacher et al., 1983). This may make students spend a shorter time searching the text, given that the relevant information is still active in their memory. Furthermore, students can become very efficient in locating and selecting the relevant ideas due to the proximity between the question and the initial reading of the text. In this sense, some studies have found that inserted questions lead to increases in reading time for question-relevant information (Lapan & Reynolds, 1994; Reynolds et al., 1979). However, when answering post-reading questions, the interval between reading relevant information and answering the question may lead to that information not being active in the

reader's memory anymore. Consequently, readers might spend more time rereading the text to find the relevant information and their search processes could be relatively inefficient, making them waste time reading information that is not relevant to the questions' goal.

Something similar may occur while task model formation, i.e., when reading an adjunct question for the first time, which involves representing the question goal and a set of means for achieving the goal (Rouet & Britt, 2014). Students not only have to understand the question but also build a plan to give a response, which implies recalling the relevant text information or developing a plan to search for question-relevant content (Mañez et al., 2022). Therefore, question timing may affect the time students spend on task model formation. While answering inserted questions, students might still have in memory the information relevant for task model formation; however, when the questions are presented at the end of the passage, the question-relevant information is no longer active in the students' memory, making it more difficult to represent the question goal and the means to achieve it.

Finally, question timing may also affect the initial reading of the text. Several experiments have found that junior high school students and college students read expository texts relatively quickly when they expect to have the text available to answer post-reading questions (Ferrer et al., 2017; Higgs et al., 2017). Readers may believe that a superficial reading of the text will be enough to form a schema of text information to be used while searching for relevant information. However, when answering inserted questions, students may read the text slowly and more carefully, which may especially favor the generation of inferences between relevant text ideas while reading the text (Olson et al., 1985). That is, under this condition, students may be likely to adopt studying as their primary task, rather than answering the adjunct questions (Hamaker, 1986). In addition, this sort of reading may make searching more efficient in comparison to post-reading question condition since text information is still active in the reader's memory.

The current study

This study aims to examine the effect of question timing (i.e., inserted in the text or presented after reading the whole text) on students' processing strategies and online reading behavior when studying a long passage composed of many interconnected text ideas. Thus, we used a between-subjects design. We assumed that inserting the questions would assist readers while text processing compared to post-reading questions. First-year college students studied two science texts while answering inserted versus post-reading adjunct questions. Online processing data while students read the texts and answered the questions were recorded. Five days later, students' learning was assessed using a test with short-answer questions closely related, but not identical, to the information covered by the adjunct questions.

We formulated five hypotheses regarding the effect of the two experimental conditions on students' text processing and final learning. First, we expected readers in the inserted question condition (hereafter referred to as *inserted condition*) to spend more time reading the text initially than those in the post-reading question condition (hereafter referred to as *post-reading condition*) (H1).

Second, students in the post-reading condition would spend more time reading the questions for the first time than students in the inserted condition, as the task-model formation processes will be more time-consuming in the former (H2).

Third, we expected that students in the inserted condition would allocate their resources more efficiently during the search process than students in the post-reading condition. More specifically, we expected an interaction effect between condition and relevance of rereading (H3a). The interaction is expected to result from two different patterns: first, students in the inserted condition will spend more time rereading question-relevant information than post-reading students, whereas the opposite would be true for reading non-relevant information; second, students in the inserted condition will spend more time rereading relevant information than non-relevant information, whereas the opposite would be true for students in the post-reading condition. We also predicted that students in the post-reading condition would reread more text segments to look for question-relevant information in comparison to students in the inserted condition, which is indicative of inefficiency in the search process (H3b).

Fourth, regarding the processing strategies used when answering the questions, we predicted that students in the inserted condition would make more correct elaborations and fewer incorrect elaborations than students in the post-reading condition (H4a), due to the better task model formation and search processes in that condition. However, no significant differences between the two experimental conditions were expected regarding correct and incorrect paraphrases (H4b).

Fifth, as a consequence of the processing mentioned, the inserted condition was expected to be more effective than the post-reading condition for learning (H5).

Method

Participants and design

The total sample consisted of 84 freshmen from the Faculty of Teacher Training at the University of Valencia, Spain. We excluded eight students because of missing sessions. The final sample included 76 participants (M age = 18.89, SD = 2.41; 76.30% female): 39 for inserted condition and 37 for post-reading condition. At least 34 participants per condition were needed to detect a medium-to-large effect size (Cohen's d = 0.81), at an alpha error of .05, and statistical power of 0.95. This effect size was based on studies with similar outcomes (e.g., van den Broek et al., 2001). All participants were native Spanish speakers.

Participants were randomly assigned to one of the two experimental conditions, ensuring equality in prior knowledge between conditions, measured with a test (see below). No difference in prior knowledge between inserted condition (M = 8.46; SD = 4.64) and post-reading condition (M = 8.59; SD = 4.58) was apparent, $t(74) = 0.13$, $p = .900$, $d = 0.03$.

Materials

Materials included a test on prior knowledge, texts and adjunct questions, and a transfer test. These materials were validated in previous studies (e.g., M \acute{a} ñez, 2020).

Test on prior knowledge

It included 30 items about science with three options (i.e., True/False/I don't know). Items were about general knowledge in science (e.g., "Density is the proportion of mass to volume"), and specific knowledge related to both text topics, i.e., *Atmospheric Pressure*

and *Heat Transmission*, not included in the texts (e.g., “The thermometer is the instrument with which we measure heat”). Cronbach’s alpha revealed reasonable reliability, $\alpha = .74$.

Texts

Two science texts about *Atmospheric Pressure* and *Heat Transmission* were used. The text order was counterbalanced. Although the students had some prior background knowledge (since the topics were addressed in the secondary school curriculum), we selected them for being sufficiently challenging for students to still struggle to comprehend. *Atmospheric Pressure* text included 965 words in length distributed in four sections: the weight of the air, the Torricelli’s experiment and the discovery of the barometer, the influence of altitude and temperature on atmospheric pressure, and the origin of the wind and its displacement. *Heat Transmission* text consisted of 895 words in length distributed in two sections: differences between *heat*, *temperature* and *internal energy*, and the thermal conductivity plus the different thermal sensations depending on the material type. A group of experts divided both texts into segments by idea-units, with 37 segments for the *Atmospheric Pressure* text and 26 segments for the *Heat Transmission* text. A segment may include only one sentence, e.g., “Atmospheric pressure is the force exerted, at a given point, by the weight of the column of air extending above that point, up to the upper limit of the atmosphere”, or several sentences closely related by the unit-idea, e.g., “The Earth is surrounded by a layer of gases that separates it from the space that constitutes, for the most part, the Universe. This layer is called the atmosphere and is made up of a mixture of gases that we call air” (see Appendix 1).

Adjunct questions

We developed a set of open-ended questions referring to the above-mentioned ideas. After testing the questions in a pilot study, we selected five high-level (e.g., “If Torricelli’s experiment were replicated at the top of Everest, would more or less mercury come out of the tube into the bucket? Why?”) and five low-level questions (e.g., “What happens when there are high-pressure air masses next to low-pressure air masses?”). Please, note that low-level questions required understanding text ideas, rather than identifying factual information (e.g., names, locations, etc.), whereas high-level questions required making inferences by applying text information to new situations not described in the text. The distribution was three low-level and two high-level questions for “Atmospheric Pressure”, and the opposite for “Heat Transmission”. In the inserted condition, both types of questions were inserted immediately after reading all the information needed to answer the question. However, the last segment read does not provide all the information to answer the question, especially in high-level questions, in which the relevant information is located in several segments that may not be contiguous. Both the text and the questions have been used in previous experiments with good results.

Transfer test

It included 10 open-ended questions that addressed the same key information as the adjunct questions, so they were near transfer questions (e.g., “In which direction will the wind move when there are nearby areas with different atmospheric pressure?” and “Someone says to you: If you replicated Torricelli’s experiment on top of a mountain would less mercury come out of the tube. Would you agree, and why?”). There were five high- and five low-level questions with the same distribution that the adjunct questions.

Measures

The measures are divided into three categories: processing strategies, online reading behavior, and transfer test. Both processing strategies and transfer test were manually coded by two examiners. After several training sessions, they independently coded approximately 15% of the sample. The first examiner continued coding the remaining responses after resolving disagreements. It should be noted that data from both texts were aggregated to obtain the measures.

Processing strategies

We distinguished between paraphrases and elaborations in the students’ responses to adjunct questions. Paraphrases were counted when students’ responses included idea-units from the text. A paraphrase would be correct if the idea of the text was correctly reported, but it would be incorrect if the idea-unit implied misunderstanding. Elaborations referred to idea-units not present in the text. Depending on the meaning, elaborations were also coded as correct. Please note that a student’s response may include a combination of correct and incorrect paraphrases and elaborations. The total number of paraphrases and elaborations were computed by adding correct and incorrect strategies. For example, “*Atmospheric pressure is a force (1st correct paraphrase) caused by the weight of the column of air on a point (2nd correct paraphrase). The amount of pressure will depend on the length of the air column (1st correct elaboration)*”. The students’ responses may also be coded as non-analyzable (NA) when they were too short, incomplete, or with incongruent meaning. Cohen’s kappa indicated a high inter-rater agreement ($\kappa = .81, p < .001$).

Online reading behavior. We used these indices:

- (a) *total number of text segments* while searching is the sum of unmasked segments in search time to answer the total number of questions. Each segment may have been counted more than once in case it has been consulted by the student several times.
- (b) *time reading the questions for the first time* is the sum of the time spent during the first unmasking of each question.
- (c) *time reading the text for the first time* is the time spent prior to accessing the questions. For the post-reading question condition, this measure corresponds to the time reading text segments before moving to the question screen. For the inserted question condition, this time corresponds to the sum of the time reading the different text segments before accessing each question.
- (d) Other measures are *time rereading relevant information* and *time rereading non-relevant information* while searching to answer the total number of questions. The protocol for extracting times by rereading relevant information is adapted from Vidal-

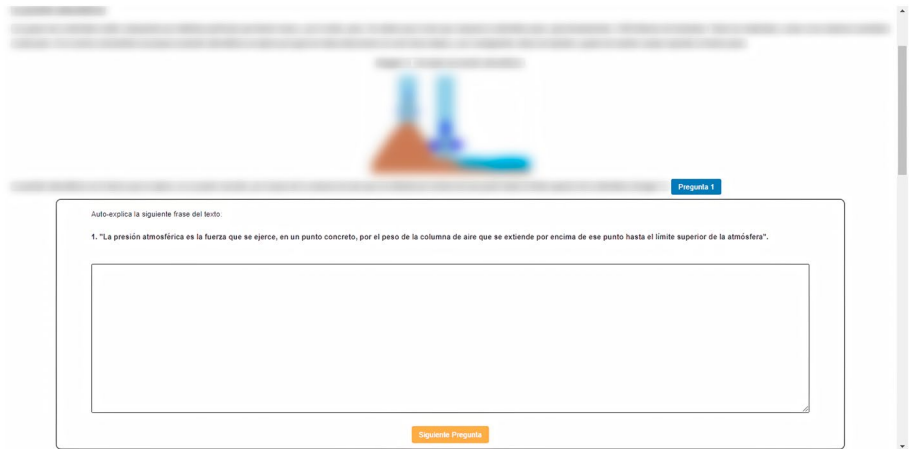


Fig. 1 Screenshot of inserted question condition: text screen and question unmasked

Abarca et al. (2010). In the output file provided by Read&Learn, every text segment was previously classified as relevant or non-relevant information for each question depending on its goal. A segment includes one or several sentences that are unmasked at a time. Therefore, the system registers its rereading as relevant or not relevant depending on the question. Consequently, a segment may be reread before answering questions 1 and 2, but be only relevant for question 1. Therefore, the time spent on this segment would be classified as time rereading relevant information for question 1 and as time rereading non-relevant information for question 2. The measure used here is the sum of relevant and non-relevant search time for each question, taking into account which information is key or not depending on the question target.

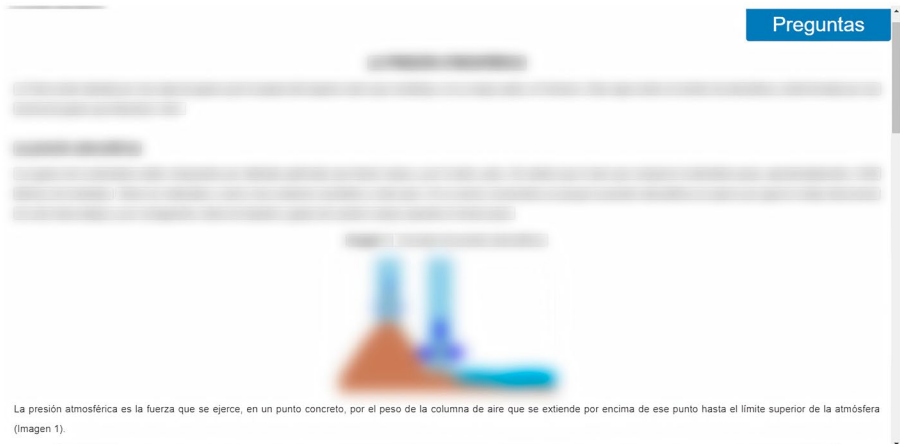
Learning

It refers to the students' outcome on the transfer test. Each correct response scored 1 point; partial responses scored 0.5; and incorrect 0, so the maximum was 10. The coding was carried out using an answers key developed by the experts in the field. Cohen's kappa indicated a high inter-rater agreement ($\kappa = .85, p < .001$).

Apparatus

Adjunct questions were answered on a computer web-based system called Read&Learn (Vidal-Abarca et al., 2018). Students in the inserted condition completed the task on a single screen because questions were inserted in the text (see Fig. 1), whereas text and questions were displayed on two separate screens for the post-reading condition (see Fig. 2a, b). Please note that text and questions displayed in Figs. 1 and 2 were masked. To read a masked segment, students had to click on it. This segment remained unmasked until the participant clicked on another segment, then the previous segment was masked again. Through this masking procedure, it was possible to record the students' reading actions. In both conditions, students could reread or search for information in the text at any time.

(a)



La presión atmosférica es la fuerza que se ejerce, en un punto concreto, por el peso de la columna de aire que se extiende por encima de ese punto hasta el límite superior de la atmósfera (Imagen 1).

(b)



Responde la siguiente pregunta.

1. ¿Qué da lugar a la presión atmosférica?

Sigüiente Pregunta

La presión atmosférica es la fuerza que se ejerce, en un punto concreto, por el peso de la columna de aire que se extiende por encima de ese punto hasta el límite superior de la atmósfera (Imagen 1).

Fig. 2 Screenshots of post-reading question condition: (2a) Text Screen With a Text Segment Unmasked; (2b) Question Unmasked

Procedure

The experiment was conducted in three different sessions with a time limit. In the first session, students completed the test on prior knowledge in a paper-and-pencil format. In the second session (i.e., study phase), participants read the two texts and answered the adjunct questions. Before that, students were instructed on how to use Read&Learn for the experimental task. Whereas students in the post-reading condition were asked first to read the text and then to answer the questions, students in the inserted condition were instructed to read and answer questions in a continuous and sequential way. After 5 days, students completed the transfer test in a paper-and-pencil format (i.e., assessment phase).

Data analyses

Statistical analyses were conducted using R (R Core Team, 2019). Normality assumptions were tested using the Shapiro–Wilk test. Homogeneity of variance assumption was further considered using Levene tests when normality of the response variable was met, and Fligner Killeen tests were employed otherwise. Assumption of multivariate normality was examined with Shapiro–Wilk tests of multivariate normality and homogeneity of covariance matrices assumption was tested using Box’s M test. If assumptions were not met, robust estimation was conducted (Wilcox, 2013).

Descriptive analyses and Spearman correlations were performed as preliminary analyses. Mixed ANOVAs and Student’s unpaired t test, or their robust equivalents were performed to test the differences between inserted and post-reading conditions in processing strategies, online reading behavior, and learning. Effect sizes were also computed, concretely partial eta squared for ANOVA and Cohen’s d (d ; Cohen, 1988). The robust Cohen’s d (d_R ; Algina et al., 2005) was provided in case assumptions for unpaired t-test were not met.

Additional to system library, R packages employed in the analyses were: car (v3.0–10; Fox & Weisberg, 2019), effsize (0.8.1; Torchiano, 2020), WRS2 (1.1–1; Mair & Wilcox, 2020), reshape2 (1.4.4; Wickham, 2007), rstatix (0.6.0; Kassambara, 2020) and heplots (1.3–8; Fox et al., 2018).

Results

Descriptive statistics and correlations of measures

Table 1 presents the descriptive statistics of the investigated variables for each question condition. Furthermore, Table 2 shows the correlations between processing strategies, online reading behavior, and learning. The results revealed significant correlations between processing strategies and learning in the post-reading condition, being positive in the case of correct paraphrases and elaborations and negative with incorrect paraphrases and

Table 1 Descriptive statistics of the measures

	Post-reading condition		Inserted condition	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Correct paraphrases	18.11	8.14	18.79	4.76
Incorrect paraphrases	1.59	1.44	1.10	1.57
Correct elaborations	7.92	4.16	10.44	3.68
Incorrect elaborations	3.05	2.52	1.28	1.34
Total number of text segments	99.16	70.16	59.18	40.02
First time reading the questions for the first time	86.59	29.40	693.85	332.32
Time reading the text for the first time	493.50	310.96	74.38	24.67
Time rereading relevant information	208.66	139.05	244.62	146.77
Time rereading non-relevant information	310.09	258.21	183.45	154.69
Learning	5.08	2.01	6.22	1.81

Table 2 Spearman correlations of the measures

	1	2	3	4	5	6	7	8	9	10	
Post-reading condition	1	1.00	-.30	.46**	-.25	.00	-.31	-.14	.32	-.10	.57***
	2	1.00	1.00	.58***	.32	-.06	.02	.07	-.20	-.04	-.46**
	3		1.00	1.00	-.51**	-.13	-.06	.11	.06	-.05	.71***
	4			1.00	.12	.22	-.27	.00	.14	.14	-.60***
	5				1.00	-.01	-.53***	.75***	.91***	-.24	-.24
	6					1.00	.11	-.15	.06	-.11	-.11
	7						1.00	-.66***	-.50**	.08	.08
	8							1.00	.60***	.00	.00
	9								1.00	-.20	-.20
	10									1.00	1.00
Inserted condition	1	1.00	-.18	.52**	.00	-.03	-.41**	.14	-.08	-.09	.38*
	2	1.00	1.00	1.00	.20	-.04	.38*	-.29	-.10	.11	-.42**
	3		1.00	1.00	1.00	-.41*	-.21	.21	-.11	-.13	.57***
	4			1.00	1.00	.09	.19	-.26	.04	.19	-.50**
	5				1.00	1.00	1.00	-.58***	.88***	.90***	-.28
	6					1.00	1.00	.04	.04	.06	-.27
	7						1.00	1.00	-.47**	-.70***	.54***
	8							1.00	.80***	-.26	-.26
	9								1.00	-.40*	-.40*
	10									1.00	1.00

1 = Correct paraphrases; 2 = Incorrect paraphrases; 3 = Correct Elaborations; 4 = Incorrect elaborations; 5 = Total number of text segments; 6 = First time reading the questions for the first time; 7 = Time reading the text for the first time; 8 = Time rereading relevant information; 9 = Time rereading non-relevant information; 10 = Learning

* $p < .05$; ** $p < .01$; *** $p < .001$

elaborations. In addition to these correlations, in the inserted questions condition the time reading the text for the first time correlated significantly and positively with learning, while time rereading non-relevant information correlated negatively.

Effect of question timing on students' reading behavior

A *t* test was conducted to examine differences in *time reading the text for the first time* as a function of experimental conditions (H1). Shapiro–Wilk normality test results was $W=0.98$ ($p=.140$). Levene's test indicated equal variances ($F=0.04$, $p=.852$). Results were consistent with the hypothesis H1. Readers in the inserted condition ($M=693.85$, $SD=332.32$) spent more time reading the text initially compared to readers in the post-reading condition ($M=493.50$, $SD=310.96$), $t(74)=-2.71$, $p=.008$, $d=-0.62$.

To test whether students in the post-reading condition would spend more *time reading the questions for the first time* than students in the inserted condition (H2), a robust *t* test was conducted. Shapiro–Wilk normality test results was $W=0.92$ ($p<.001$). Fligner-Killeen test yielded a statistically non-significant result regarding homogeneity of variance between experimental conditions for time for the first reading of the question ($FK(1)=1.61$, $p=.204$). Results showed statistically significant differences between post-reading condition ($M=86.59$, $SD=29.40$) and inserted condition ($M=74.38$, $SD=24.67$), $t(53.71)=2.01$, $p=.049$, $d_R=0.43$, which was consistent with our prediction.

Regarding *relevance of rereading*, Shapiro–Wilk normality test results were $W=0.95$ ($p=.003$) for relevant time and $W=0.85$ ($p<.001$) for non-relevant time. Fligner-Killeen test yielded a statistically non-significant result regarding homogeneity of variance between experimental conditions for time rereading relevant information ($FK(1)=1.15$, $p=.287$) and non-relevant ($FK(1)=3.11$, $p=.07$). To test whether inserted condition would lead students to spend more time rereading relevant text information than post-reading

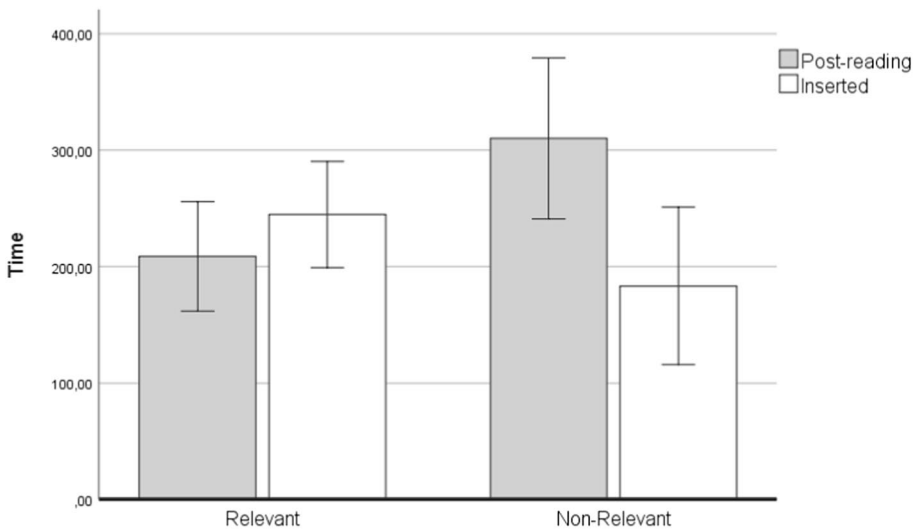


Fig. 3 Interaction effect between question condition and relevance of rereading. Relevant=Total time rereading relevant text information while searching; Non-relevant=Total time rereading non-relevant text information while searching

condition (H3a), we conducted a robust two-way mixed ANOVA, with relevance of rereading (i.e., time reading relevant and non-relevant information while searching the text) as a within-subjects variable, and question-answering condition (inserted, post-reading) as a between-subjects variable. Results showed a significant interaction effect between question condition and relevance of rereading, $F(1, 32.57) = 15.49, p < .001$ (see Fig. 3). Students in the post-reading condition spent more average time rereading non-relevant information ($M = 310.09, SD = 258.21$) than relevant information ($M = 208.66, SD = 139.05$), whereas students in the inserted condition spent more average time rereading relevant information ($M = 244.62, SD = 146.77$) than non-relevant information ($M = 183.45, SD = 154.69$). In addition, students in the inserted condition spent more time rereading relevant information than post-reading students, whereas time rereading non-relevant information was higher in the post-reading condition than in the inserted condition. These results were consistent with our predictions. There were no statistically significant differences between relevant and non-relevant time of rereading, $F(1, 32.57) = 0.02, p = .879$. Finally, there were no statistically significant differences between experimental conditions in time rereading, $F(1, 43.57) = 0.66, p = .431$.

To examine differences in *total number of text segments read while searching* between experimental conditions (H3b), a robust *t*-test and associated robust effect size were computed. Shapiro–Wilk normality test showed a violation of the normality assumption ($W = 0.87, p < .001$), although Fligner–Killeen indicated equal variances, $FK(1) = 3.48, p = .06$. Results displayed post-reading condition to search significantly more text segments ($M = 99.16, SD = 70.16$) compared to inserted condition ($M = 59.18, SD = 40.02$), $t(49.89) = 2.99, p = .004, d_R = 0.63$. This result was consistent with our prediction.

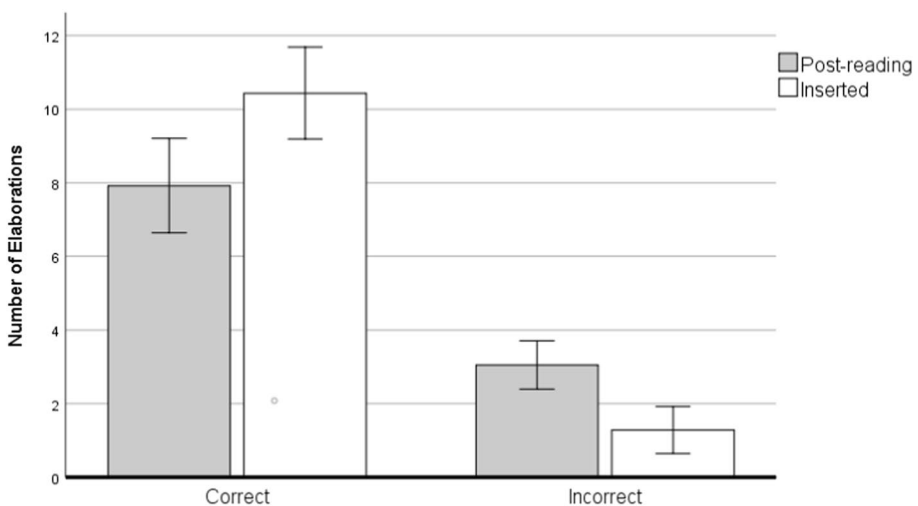


Fig. 4 Interaction effect between question condition and elaborations. Correct=Total number of correct elaborations; Incorrect=Total number of incorrect elaborations

Effect of question timing on processing strategies

Shapiro–Wilk normality tests showed incorrect elaborations to be non-normally distributed ($W=0.81$, $p<.001$). A robust two-way mixed ANOVA revealed a statistically non-significant effect of experimental condition ($F(1, 39.56)=3.43$, $p=.071$) and a statistically significant difference between correct ($M=9.21$, $SD=4.09$) and incorrect ($M=2.14$, $SD=2.17$) elaborations, $F(1, 43.10)=188.31$, $p<.001$. Moreover, a statistically significant interaction was found between correct/incorrect elaborations and experimental condition, $F(1,43.10)=24.35$, $p<.001$ (see Fig. 4). More concretely, students in the inserted condition made more correct elaborations ($M=10.44$, $SD=3.68$) than those in the post-reading condition ($M=7.92$, $SD=4.16$), and fewer incorrect elaborations ($M=1.28$, $SD=1.34$) than post-reading condition ($M=3.05$, $SD=2.51$). This result provided support for our hypothesis (H4a). Regarding correct and incorrect paraphrases, Shapiro–Wilk normality tests showed correct paraphrases ($W=0.88$, $p<.001$) and incorrect paraphrases ($W=0.82$, $p<.001$) to be non-normally distributed. Fligner–Killeen test proved correct paraphrases variances of inserted and post-reading conditions to be equal, $FK(1)=0.81$, $p=.368$, and this was also the case for incorrect paraphrases variances of inserted and post-reading, $FK(1)=0.59$, $p=.442$. Consequently, a robust two-way mixed ANOVA was conducted. Results showed no effect of experimental condition ($F(1, 43.96)=0.09$, $p=.766$) nor interaction between condition and correct/incorrect paraphrases ($F(1, 42.67)=1.96$, $p=.168$). However, a statistically significant difference was found between correct ($M=18.46$, $SD=6.58$) and incorrect ($M=1.34$, $SD=1.52$) paraphrases, $F(1,42.67)=742.71$, $p<.001$. The result was in line with our prediction (H4b).

Impact of adjunct question timing on learning outcome

To examine whether inserted condition would be more effective than post-reading condition in the students' learning (H5), we conducted an unpaired Student's t-test with learning as a dependent variable, and question-answering condition as a between-subjects variable. Shapiro–Wilk normality test results was $W=0.97$ ($p=.060$). Levene's test indicated equal variances ($F=0.90$, $p=.345$). Results showed that learning was higher for inserted condition ($M=6.22$, $SD=1.81$) than for post-reading condition ($M=5.08$, $SD=2.01$), $t(74)=-2.59$, $p=.011$, $d=-0.59$, which is consistent with the prediction.

Discussion

We investigated the effect of question timing on students' processing strategies and online reading behavior when studying a challenging text, as well as its impact on learning. For this purpose, students either answered adjunct questions after reading the whole text or inserted questions placed right after the question-relevant information. We find that answering inserted adjunct questions is more effective for learning than answering post-reading questions, which confirms our fifth hypothesis. The analysis of online reading behavior and processing strategies may contribute to explaining this advantage. First, we analyzed a series of reading behavior indices that are indicative of cognitive processes. We first predicted that readers in the inserted condition would spend more time reading the text initially than those in the post-reading condition, which was confirmed. Our interpretation is that presenting questions inserted in the text improved the students' standards for

coherence when constructing meaning from a text, whereas post-reading questions induced a quick reading of the text aimed at forming an outline of the text that enables later location of relevant information (Ferrer et al., 2017; Higgs et al., 2017).

Our second hypothesis predicted that students in the post-reading condition would spend more time reading the questions for the first time than students in the inserted condition, which was confirmed by the results. This can be explained by the processes involved in task-model formation. When reading a question, the student builds a representation of the end goal and the information needed for the answer. When the question is inserted in the text, that relevant information can be easily reactivated. In contrast, when the questions are presented at the end of the passage, the relevant information is no longer active in memory. Therefore, the reader has to think about where it was located and possibly start mental search processes, which takes time.

The results also confirmed our third hypothesis, i.e., students in the inserted condition would allocate their resources more efficiently during the search process than students in the post-reading condition. Inserted-questions students spent more time rereading relevant than non-relevant information, but the opposite was true for students in the post-reading condition. Further students in the inserted condition spent more time rereading relevant information than post-reading students, whereas the opposite was true for non-relevant information. We had also predicted that students in the post-reading condition would search a higher number of text segments in comparison to students in the inserted questions condition. The effect of questioning to cue students' attention toward specific information is well-known (Dirkx et al., 2015; McCrudden & Schraw, 2007). However, no research had explored how this effect varies as a function of question timing. The delay between reading relevant information and reading the question in the post-reading condition, plus the quick initial reading of the text makes searching for relevant information difficult for the students. On the contrary, that search is facilitated when the delay is very short (i.e., inserted condition), and the text has been read more carefully.

Regarding the processing strategies when answering questions, our fourth hypothesis predicted that students' responses in the inserted condition would include more correct elaborations and fewer incorrect elaborations compared to students in the post-reading condition. Our results were consistent with this prediction. When answering inserted questions, especially high-level questions, students are challenged to apply text information to new situations. Students may easily reread relevant information, which activates the readers' prior background knowledge to make the appropriate inferences. As a result, correct elaborations are quite likely. However, when answering post-reading questions students struggle to find relevant information, which interferes with the activation of prior background knowledge, and the corresponding correct inferences. As a result, the probability of correct elaborations decreases, while that of incorrect elaborations increases. However, we predicted no significant differences between the two experimental conditions regarding correct and incorrect paraphrases, which was confirmed. For freshmen students, paraphrasing depends on understanding explicit text ideas, which may be produced by recalling text ideas, or by accessing relevant information, at least when text content is accessible, which was the case. Since the text was available to answer the questions, paraphrasing the text was not difficult, independently of the question timing.

In summary, the analysis of online behavior variables while reading the text, the questions and making search decisions, as well as the students' processing strategies while responding to the questions contribute to explaining why answering inserted adjunct questions enhanced the students' learning compared to post-reading questions. Despite this, the present study has several limitations. First, the benefits of inserted questions

should not be generalized to all types of texts and questions. They may not be as favorable for less challenging texts, or texts with sections relatively disconnected one from the other so that understanding a section was not necessary to understand any other. We also cannot confirm which condition is better for the retention of non-question information in the study phase (i.e., *general attention perspective*). Second, Read&Learn has some additional registration possibilities, but it also has some costs (e. g., unmasking may make scanning the text slightly more difficult). Although Vidal-Abarca et al. (2010) found no significant differences in strategic patterns between eye-tracking and the previous version of Read&Learn (i.e., Read&Answer), it would be advisable to compare the effect of question timing in a more natural environment (i.e., with unmasked texts. Third, we only examine the learning of ideas closely related to those addressed in the adjunct questions. We made this decision because our goal was to promote the students' learning of important challenging text ideas. Future studies might examine how the effect of the adjunct question can be extended to information less connected to the questions. Other limitations are the students' age and the absence of a control group. Maybe younger students do not benefit from inserted questions. Actually, van den Broek et al. (2001) found that inserted questions were harmful to young readers when reading narratives because of the overloading of the readers' working memory when processing text information simultaneously with answering questions. In addition, we cannot say that either timing of questions favors greater learning compared to another condition without questions (e.g., just reading the text). These limitations should be addressed in future studies.

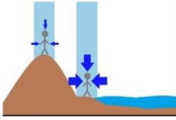
Our findings have important educational implications. Textbook publishers and educational institutions who publish instructional materials should understand *when* to present questions when developing student learning materials, either inserted into the text or after reading the text. Electronic materials such as e-textbooks open new possibilities in this regard. The present study provides arguments to make decisions on these two points. Furthermore, if teachers are aware of the underlying question-answering processes, they can teach students comprehension strategies to answer different types of questions depending on when they are presented. At a more global level, educational institutions can include the main conclusion of this study when preparing training courses for in-service and pre-service teachers.

Appendix 1 Text segments

-
- 1 Atmospheric Pressure
-
- 2 The Earth is surrounded by a layer of gases that separates it from the space that makes up, for the most part, the Universe. This layer is called the atmosphere and is made up of a mixture of gases that we call "air".
 - 3 **Atmospheric pressure**
 - 4 The gases of the atmosphere are composed of different particles that have mass and, therefore, weight. It is estimated that the air that makes up the atmosphere weighs approximately 5500 billion tons.
 - 5 All materials and living beings are subjected to this weight. If we are not aware of it, it is because atmospheric pressure is exerted equally in all directions (not only downwards) and, consequently, all liquids and gases in our body bear the same weight.
 - 6 **Fig. 1** *Concept of atmospheric pressure.*
-

 1 Atmospheric Pressure

7



8 Atmospheric pressure is the force exerted at a particular point by the weight of the column of air extending above that point to the upper boundary of the atmosphere (Fig. 1). (**Question 1**)

9 **The discovery of atmospheric pressure**

10 In 1643, the Italian scientist Evangelista Torricelli discovered the force of the weight of air through the following experiment:

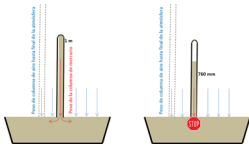
11 Torricelli filled a one-meter-long tube closed at one end with mercury and covered the open end with his finger. Subsequently, he introduced it inverted in a bucket filled with mercury and removed his finger carefully so that no air could enter. At this point, the mercury in the tube started to flow out into the cuvette (Fig. 2a), but when it dropped to a height of 760 mm it stopped flowing out (Fig. 2b).

12 Why did the mercury start flowing out of the tube and then stop? Because, at first, the weight of the mercury in the tube was greater than the weight of the column of air above the mercury in the cuvette, but there came a time when the weight of the mercury in the tube was equal to the weight of the column of air.

13 Torricelli's conclusion was clear: the weight of the column of air (atmospheric pressure) was equal to the weight of a 760 mm column of mercury.

14 **Fig. 2 Torricelli's experiment.**

15



16 (a) The mercury leaves the tube because the weight of the mercury is greater than the weight of the column of air.

17 (b) The mercury stops leaving because the weight of the 760 mm column of mercury is equal to the weight of the column of air.

18 Therefore, it is known that the air pressure at sea level is equal to the pressure exerted by a column of mercury (Hg) 760 mm high, a value that is equivalent to one atmosphere ($760 \text{ mmHg} = 1 \text{ atm}$).

19 Through this experiment, in addition to demonstrating that air exerted pressure, Torricelli found a way to measure this pressure, which allowed him to design an apparatus he called a barometer. (**Question 2**)

20 **Atmospheric pressure varies**

21 The particles of the gases move freely, occupying the entire volume, although the distribution of these particles is not uniform, since the atmosphere is a very large space and the conditions in each place are different.

22 Thus, the density of the air layers varies according to height, with the air in the lower layers being denser because they support the weight of the air in the layers above them.

23 The lower layers are crushed by the upper layers, the air particles in these layers being closer together and more compressed, i.e., there are more particles per unit volume.

24 Since the pressure depends on the weight of the air above us, as we ascend, the pressure decreases. Thus, the normal pressure at sea level is 1 atm, while the pressure at the top of a mountain is much lower.

25 For this reason, if we were to try to replicate Torricelli's experiment at the top of Everest, instead of at sea level as in the original experiment, the amount of mercury that would come out of the tube would vary. (**Question 3**)

26 Atmospheric pressure, in addition to varying with altitude, also varies with temperature.

1 Atmospheric Pressure

27 When air is heated, its particles accelerate and tend to separate. Thus, hot air increases in volume and decreases in density. Remember that less dense objects weigh less per unit volume (1 m^3 of hot air weighs less than 1 m^3 of cold air). The opposite happens when gases are cooled.

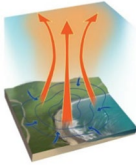
28 Therefore, the size of a balloon filled with air and hermetically sealed will change if we lower the air temperature by $10\text{ }^\circ\text{C}$. (**Question 4**)

29 Atmospheric pressure and wind

30 Differences in atmospheric density and pressure in different parts of the planet are responsible for winds and other meteorological phenomena.

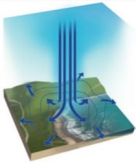
31 **Fig. 3** *Formation of a low-pressure zone or squall.* Low atmospheric pressure zones, called cyclones or squalls, are formed by masses of warm air that, when they rise, leave behind an area of low density that is filled in by neighboring air masses.

32



33 **Fig. 4** *Formation of a high-pressure area or anticyclone.* In high-pressure zones, called anticyclones, it is the cold, denser air masses that tend to descend from the upper layers, causing compression of the lower air masses and their dispersion when they reach the surface.

34



35 If we combine these two phenomena we can understand how the dynamics of the atmosphere work and how the wind is produced.

36 The wind is the movement of large air masses through the troposphere (the lower layer of the atmosphere).

37 The existence of low-pressure zones close to high-pressure zones causes air movements from the higher pressure zones to the lower pressure zones until these pressures equalize. (**Question 5**)

Acknowledgements This work was supported by the Spanish Ministry of Universities [grant number FPU15/02280], and the Spanish Ministry of Economy, Industry and Competitiveness [grant number EDU2017-86650-R].

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was supported by the Spanish Ministry of Universities [Grant number FPU15/02280], and the Spanish Ministry of Economy, Industry and Competitiveness [Grant number EDU2017-86650-R].

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures followed were in accordance with the ethical standards of the University of Valencia and with the Helsinki Declaration.

Consent to participate All subjects gave informed consent prior to participating in the present experiment.

Consent for publication All authors gave consent for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, *111*(2), 189–209. <https://doi.org/10.1037/edu0000282>
- Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, *10*(3), 317–328. <https://doi.org/10.1037/1082-989X.10.3.317>
- Andre, T., & Womack, S. (1978). Verbatim and paraphrased adjunct questions and learning from prose. *Journal of Educational Psychology*, *70*(5), 796–802. <https://doi.org/10.1037/0022-0663.70.5.796>
- Carrier, C. A., & Fautsch-Patridge, T. (1981). Levels of questions: A framework for the exploration of processing activities. *Contemporary Educational Psychology*, *6*(4), 365–382. [https://doi.org/10.1016/0361-476X\(81\)90019-9](https://doi.org/10.1016/0361-476X(81)90019-9)
- Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences*, *21*(2), 201–205. <https://doi.org/10.1016/j.lindif.2010.11.007>
- Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*(1), 209–222. <https://doi.org/10.1037/0022-0663.100.1.209>
- Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*(1), 13–27. <https://doi.org/10.1016/j.learninstruc.2007.12.003>
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. I. I. I. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*(4), 553–571. <https://doi.org/10.1037/0096-3445.135.4.553>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.
- Coté, N. A., Goldman, S. R., & Saul, E. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*, 1–53. <https://doi.org/10.1080/01638539809545019>
- Dirkx, K. J., Thoma, G. B., Kester, L., & Kirschner, P. (2015). Answering questions after initial study guides attention during restudy. *Instructional Science*, *43*(1), 59–71. <https://doi.org/10.1007/s11251-014-9330-9>
- Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, *49*, 13–20. <https://doi.org/10.1016/j.learninstruc.2016.11.010>
- Ferrer, A., Vidal-Abarca, E., Serrano, M. Á., & Gilabert, R. (2017). Impact of text availability and question format on reading comprehension processes. *Contemporary Educational Psychology*, *51*, 404–415. <https://doi.org/10.1016/j.cedpsych.2017.10.002>
- Fox, J., Friendly, M., & Monette, G. (2018). *heplots: Visualizing tests in multivariate linear models*. R package version 1.3–5. <https://CRAN.R-project.org/package=heplots>.
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.
- Gil, L., Martínez, T., & Vidal-Abarca, E. (2015). Online assessment of strategic reading literacy skills. *Computers and Education*, *82*, 50–59. <https://doi.org/10.1016/j.compedu.2014.10.026>

- Goldman, S. R., & Durán, R. P. (1988). Answering questions from oceanography texts: Learner, task, and text characteristics. *Discourse Processes*, 11(4), 373–412. <https://doi.org/10.1080/01638538809544710>
- Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, 56(2), 212–242. <https://doi.org/10.2307/1170376>
- Hamilton, R. J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research*, 55(1), 47–85. <https://doi.org/10.2307/1170407>
- Higgs, K., Magliano, J. P., Vidal-Abarca, E., Martínez, T., & McNamara, D. S. (2017). Bridging skill and task-oriented reading. *Discourse Processes*, 54(1), 19–39. <https://doi.org/10.1080/0163853X.2015.1100572>
- Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test ... or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, 26, 307–329. <https://doi.org/10.1007/s10648-013-9248-9>
- Kapp, F., Proske, A., Narciss, S., & Körndle, H. (2015). Distributing vs. blocking learning questions in a web-based learning environment. *Journal of Educational Computing Research*, 51(4), 397–416. <https://doi.org/10.2190/EC.51.4.b>
- Kassambara, A. (2020). *rstatix: Pipe-friendly framework for basic statistical tests*. R package version 0.7.0. <https://CRAN.R-project.org/package=rstatix>.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Lapan, R., & Reynolds, R. E. (1994). The selective attention strategy as a time-dependent phenomenon. *Contemporary Educational Psychology*, 19(4), 379–398. <https://doi.org/10.1006/ceps.1994.1028>
- Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, 94(4), 778–784. <https://doi.org/10.1037/0022-0663.94.4.778>
- Mair, P., & Wilcox, R. R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52, 464–488. <https://doi.org/10.3758/s13428-019-01246-w>
- Mañá, A., Vidal-Abarca, E., Domínguez, C., Gil, L., & Cerdán, R. (2009). Papel de los procesos metacognitivos en una tarea de pregunta-respuesta con textos escritos [Role of metacognitive processes in a question-answering task with written texts]. *Infancia y Aprendizaje*, 32(4), 553–565. <https://doi.org/10.1174/021037009789610412>
- Máñez, I. (2020). ¿Influye la retroalimentación correctiva en el uso de la retroalimentación elaborada en un entorno digital? [Does corrective feedback influence the use of elaborated feedback in a digital environment?] *Psicología Educativa*, 26(1), 57–65. <https://doi.org/10.5093/psed2019a14>
- Máñez, I., Vidal-Abarca, E., & Magliano, J. P. (2022). Comprehension processes on question-answering activities: A think-aloud study. *Electronic Journal of Research in Educational Psychology*, 20(1), 1–26.
- McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, 19(2), 113–139. <https://doi.org/10.1007/s10648-006-9010-7>
- McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, 38(1), 1–30. https://doi.org/10.1207/s15326950dp3801_1
- Ness, M. (2011). Explicit reading comprehension instruction in elementary classrooms: Teacher use of reading comprehension strategies. *Journal of Research in Childhood Education*, 25(1), 98–117. <https://doi.org/10.1080/02568543.2010.531076>
- Olson, G. M., Duffy, S. A., & Mack, R. L. (1985). Questions-asking as a component of text comprehension. In A. C. Graesser & J. B. Black (Eds.), *The psychology of questions* (pp. 219–226). Erlbaum.
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399–438. <https://doi.org/10.1080/07370000701632371>
- Peverly, S. T., & Wood, R. (2001). The effects of adjunct questions and feedback on improving the reading comprehension skills of learning-disabled adolescents. *Contemporary Educational Psychology*, 26(1), 25–43. <https://doi.org/10.1006/ceps.1999.1025>
- Phillips, F., Lobdell, B., & Neigum, J. (2020). Does the effectiveness of interspersed and blocked questions vary across readers? *Issues in Accounting Education*, 35(1), 1–12. <https://doi.org/10.2308/iace-52630>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Reynolds, R. E., Standiford, S. N., & Anderson, R. C. (1979). Distribution of reading time when questions are asked about a restricted category of text information. *Journal of Educational Psychology*, 71(2), 183–190. <https://doi.org/10.1037/0022-0663.71.2.183>
- Rickards, J. P., & Di Vesta, F. J. (1974). Type and frequency of questions in processing textual material. *Journal of Educational Psychology*, 66(3), 354–362. <https://doi.org/10.1037/h0036349>

- Roediger, H. L., III., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Rouet, J.-F., & Britt, M. A. (2014). Multimedia learning from multiple documents. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 813–841). Cambridge University Press.
- Rouet, J. F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, 52(3), 200–215. <https://doi.org/10.1080/00461520.2017.1329015>
- Schumacher, G. M., Moses, J. D., & Young, D. (1983). Students' studying processes on course related texts: The impact of inserted questions. *Journal of Literacy Research*, 15(2), 19–36. <https://doi.org/10.1080/10862968309547481>
- Tawfik, A. A., Graesser, A., Gatewood, J., & Gishbaugher, J. (2020). Role of questions in inquiry-based instruction: Towards a design taxonomy for question-asking and implications for design. *Educational Technology Research and Development*, 68, 653–678. <https://doi.org/10.1007/s11423-020-09738-9>
- Torchiano, M. (2020). Effsize: Efficient effect size computation. Zenodo. <https://doi.org/10.5281/zenodo.1480624>
- Uner, O., & Roediger, H. L., III. (2018). The effect of question placement on learning from textbook chapters. *Journal of Applied Research in Memory and Cognition*, 7(1), 116–122. <https://doi.org/10.1016/j.jarmac.2017.09.002>
- van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 123–139). Information Age.
- van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54(5–6), 360–372. <https://doi.org/10.1080/0163853X.2017.1306677>
- van den Broek, P., Risdén, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–373). Lawrence Erlbaum.
- van den Broek, P., Tzeng, Y., Risdén, K., Trabasso, T., & Basche, P. (2001). Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology*, 93(3), 521–529. <https://doi.org/10.1037/0022-0663.93.3.521>
- van Oostendorp, H., & Goldman, S. R. (1999). *The construction of mental representations during reading*. Lawrence Erlbaum Associates Publishers.
- Vidal-Abarca, E., Martínez, T., Serrano, M. A., Gil, L., Mañá, A., Máñez, I., & Candel, C. (2018, July 17–19). *Read&Learn: A research tool to record online processing while learning* [Poster session]. 28th Annual Meeting of the Society for Text & Discourse. <https://easychair.org/smart-program/STD2018/index.html>.
- Vidal-Abarca, E., Martínez, T., Serrano, M. A., Gil, L., Mañá, A., Máñez, I., & Candel, C. (2018, July 17–19). *Read&Learn: A research tool to record online processing while learning* [Poster session]. 28th Annual Meeting of the Society for Text & Discourse. <https://easychair.org/smart-program/STD2018/index.html>.
- Weinstein, Y., Nunes, L. D., & Karpicke, J. D. (2016). On the placement of practice questions during study. *Journal of Experimental Psychology: Applied*, 22(1), 72–84. <https://doi.org/10.1037/xap0000071>
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), 1–20.
- Wilcox, R. (2013). *Introduction to robust estimation and hypothesis testing* (3rd ed.). Elsevier.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.