

## **Reading from multiple documents: The role of text availability and question type**

*Raquel Cerdán, Ignacio Máñez, & Marian Serrano-Mendizábal*

### **Abstract**

The current information society requires people to handle information from different sources, which entails specific skills beyond those needed to comprehend and use information from a single text. Given the relevance of becoming a competent reader in functional reading scenarios, it is crucial to examine how contextual-, task-, and individual-related factors may influence multiple-documents question-answering activities, a key issue for assessment and instruction. In this study, we examined the influence of text availability and question type in students' question-answering performance and the moderator effect of students' reading skill. Seventy secondary school students read three expository texts about a history topic. Next, they performed a concurrent question-answering task consisting of a set of intra- and intertext questions they answered either with or without the texts available. Finally, participants performed a free-recall task after a 24-hour delay without access to the learning materials. Results showed higher performance scores in the intratext questions when the text was available than when it was not. However, the superiority of text availability disappeared in the delayed recall task. Additionally, skilled readers who had the texts available scored higher in the concurrent question-answering task and, moreover, recalled more ideas in the recall task. These findings shed light on how contextual- (i.e., text availability), task- (i.e., question type), and student-related (i.e., reading skill) factors may influence both adolescents' question answering from multiple documents and their recall. The implications of these findings for the science of reading and the educational practice are discussed.

## **Introduction**

In academic settings, students usually read to solve tasks. Answering questions is a common task used in these contexts, both for assessment and learning purposes, since they change and enhance retention of text information (e.g., Roediger & Karpicke, 2006). This task can vary on several dimensions, such as the availability of the text while answering the questions and the type of questions presented to students. The extant literature suggests that the cognitive processes involved in answering questions from written documents vary as a function of text availability (Author, 2017; Ozuru, Best, Bell, Witherspoon, & McNamara, 2007; Schroeder, 2011). In the same vein, previous research suggests that different processes are involved in answering questions of different type (Author, 2009; Author, 2008). Both dimensions, as well as their interactions, have an impact on students' question-answering performance (Author, 2009; Author, 2008; Author, 2017; Ozuru et al., 2007; Schroeder, 2011). These findings derive from studies where students answer questions using the information of a single text; however, research has not still explored the interplay of these dimensions when considering more complex reading scenarios that adolescents quite commonly encounter in their school assignments, such as reading information from more than one text on the same topic (i.e. multiple documents).

Even though most of the teachers involved in reading literacy would assume that students mastering fluent decoding and basic comprehension skills are capable of reading and using information from different types of documents, Rouet & Potocki (2018) question this assumption. Indeed, decoding and comprehension ability seem not to be sufficient for students to face new reading demands, such as solving a set of questions from more than one text. In these contexts, it seems also relevant to acquire strategies related to information search, evaluation, and integration within and across texts (Rouet & Potocki, 2018).

The reading situations which demand students to solve a set of tasks based on the inspection of single or multiple documents have been named purposeful or task-oriented reading scenarios (Rouet, 2006; Rouet & Britt, 2011; Vidal-Abarca, Mañá, & Gil, 2010). In such scenarios, students initiate the reading activity having one goal in mind, such as answering questions. Questions are then the means to set the goals for reading, so not all document/s or text segments might be relevant, but only those relevant to answer the questions. Indeed, reading in academic settings is typically

grounded in a task or purpose. However, canonical theories of reading comprehension explain mental model construction in terms of an interaction between the reader and a text, thus putting contextual- and task-related factors aside (Higgs, Magliano, Vidal-Abarca, Martínez, & McNamara, 2017). Contemporary perspectives of reading literacy, such as the *Reading as Problem Solving framework* (RESOLV, Rouet, Britt, & Durik, 2017) view reading as a dynamic interaction among the text, the reader, context, and task factors. These dynamic perspectives have motivated a body of research to better understand how different characteristics of the reader and text interact with different task-oriented reading situations to influence question-answering performance. One common task-oriented reading situation is reading to answer questions from multiple texts, with or without access to the documents. This is the specific situation under assessment in this study.

Given the demands raised by functional reading scenarios, especially with multiple texts, there is need for more research examining the conditions under which contextual factors such as text availability, task factors such as question type, and individual differences such as reading skill, may affect multiple documents comprehension. Moreover, considering functional reading skills in multiple documents is especially relevant for the progress of the science of reading. Assessment and training practices in reading, especially from multiple documents, should incorporate the components involved in functional reading (i.e., *impact of text availability for question-answering and question type*), in addition to those classically included so far, namely decoding and comprehension from single text (Rouet & Potocki, 2018). To our knowledge, there is not published research on the role of documents availability in task-oriented reading of multiple documents. Given the current relevance of the availability of documents and the role of individual differences in the assessment of students' performance (i.e., on-line distance learning), it seems crucial to disentangle the specific contribution of each of these variables for multiple documents reading scenarios.

In the next sections, we will first present an overview of the skills involved in purposeful reading of multiple documents. Then we will refer to the main results found for the effects of text availability on question-answering performance and its interplay with task and individual differences relevant for this study (i.e., type of question and reading skill). We will conclude the introduction with the main hypotheses and predictions for the study.

## **Purposeful reading of multiple documents**

Very frequently, students are now faced with functional reading situations both in paper and web-based environments, with access to multiple documents (Author, 2020; Braasch, 2020). Specific strategies such as deciding to access a document or not, evaluating the source and content of the document and integrating information within and across texts are crucial behaviors for successful reading performance in those environments (Author, 2018)

Reading multiple documents for performing academic tasks requires the students to deploy specific cognitive and metacognitive skills to build an integrated mental representation across multiple sources of information (Rouet et al., 2017). The processes of decoding and comprehension alone do not provide any mechanism to explain what proficient students would do in those environments and under purposeful reading demands. To account for the skilled reading of multiple documents, theories of reading have outlined the skills required for students to form goals and make the strategic decisions needed to fulfill them, such as MD-TRACE and RESOLV models, (Britt, Rouet, & Durik, 2018; McCrudden & Schraw, 2007; Rouet, 2006).

In multiple documents reading, readers need to link content presented in the single texts to organize an integrated mental representation of the documents, beyond the representations at the single-text level (Perfetti et al., 1999) and solve tasks that require them to form an integrated view on the topic (i.e., Author, 2008). According to the *documents model framework* (Perfetti et al., 1999), when readers read information from several documents on the same topic they might need to establish content links between the independent documents to build an integrated representation of the set of documents read (*documents model*). That is to say, the reading of multiple texts would entail the construction of mental representations of specific texts, the situations described in the texts, and the relations among them (Perfetti et al., 1999). Thus, when moving beyond a single text, readers should make an effort to establish connections between the ideas included in the separated texts and attempt to build an integrated mental representation (Britt, Perfetti, Sandak, & Rouet, 1999) This is the specific reading situation we address in this paper.

Consequently, dealing with more than one document poses important challenges on the reader, beyond those at the single text level. Rouet and Britt (2011) proposed a model which accounts for the processing demands involved in the comprehension of multiple documents (i.e., the *multiple-document task-based relevance assessment and*

*content extraction* or MD-TRACE). MD-TRACE (Rouet, 2006; Rouet & Britt, 2011) offers a detailed framework of how students deal with and learn from multiple texts (see also, RESOLV, Rouet et al., 2017; Britt et al. 2018). Readers, texts, and contextual factors influence how reading and learning from multiple documents takes place (Braasch, 2020). MD-TRACE would especially emphasize the text inspection processes and integration activity (within and between-texts). a framework which especially aligns to the main design (integrating complementary information) and goals of this paper.

In all, these models can help us better interpret how students behave when dealing with multiple documents, especially regarding factors such as availability of documents and type of integration demanded by the task (intra-texts or inter-texts). Whereas the effects of text availability have been extensively studied in single text reading (i.e., Author, 2017; Ozuru, Best, Bell, Witherspoon, & McNamara, 2007; Schroeder, 2011), less is known for multiple documents, neither its potential interaction with task type. Delving into how having the opportunity to revisit text content to answer a set of questions impacts question-answering performance in multiple documents is a relevant question for both research and practice. It might impact how question-answering assessment situations are designed in functional reading scenarios. We next disentangle the effects of text availability, and potential interactions with task and individual variables.

### **Text availability, task effects and influence of students' reading skills**

Empirical evidence suggests that, at least in single text, text availability influences readers' question-answering performance (Author, 2017; Ozuru et al., 2007; Schroeder, 2011). Moreover, these effects vary as a function of task-related factors (e.g., question format or text genre) and individual factors (e.g., prior knowledge or reading skill) (e.g., Ozuru et al., 2007; Schaffner & Schiefele, 2013; Schroeder, 2011). In general, keeping the texts available in single text question-answering situations seems to facilitate readers' performance, given the opportunities to revisit the text to locating question-relevant information when needed (Author, 2017; Ozuru et al., 2007; Schroeder, 2011). However, students must rely on their initially-built representation of the texts to meet the task demands when they are not given the chance to revisit the documents (unavailable-text situation).

Author (2017)'s studies found an interaction effect between text availability and question type on secondary school students' question-answering performance,

indicating that having the text available contributed to increasing performance on the literal questions, but not for inferential questions. Literal questions could be answered by extracting isolated pieces of information from the text, whereas inferential questions required readers to integrate several pieces of textual information (Author, 2008). The possibility to revisit the text in search for question-relevant information might have especially favored those tasks where the copy of verbatim parts of the text an effective answering strategy is (Author, 2009). When the task requires students to integrate and/or elaborate, being allowed to revisit the documents in the question-answering phase might make no substantial difference, as Author (2008) showed.

Previous findings also suggest that both prior knowledge and reading skill contribute to students' reading outcomes more significantly when the text is unavailable during the question-answering process (Ozuru et al., 2007; Schaffner & Schiefele, 2013), probably because readers have to deploy deeper integrative processes without knowing the task demands. In order to understand the underlying cognitive processes which might explain the differences between available and unavailable situation, Ozuru et al. (2007) proposed a timeline of the cognitive processes during the initial text reading and question-answering process. During text reading, readers' comprehension is primarily knowledge-driven. However, when question reading and answering, there are differences in readers' outcomes depending on text availability.

Specifically, readers who do not have access to the text must rely on their memory representation to meet the processing requirements of the questions. Thus, if the text is unavailable, readers' execution depends primarily on the mental representation derived from the connections established between their prior knowledge and the text content during initial text reading. There is wide evidence that less skilled readers have difficulties making inferences and connecting the text content with their prior knowledge (e.g., Cain & Oakhill, 1999; Oakhill, 1982, 1984), which results in an incomplete mental representation of the text that hinders their question-answering performance. Moreover, students with poor reading skills have significant difficulties handling the demands of question-answering scenarios (Author, 2009, 2011). They are specially prone to answering questions based on a literal copy of information from the texts, with reduced elaboration and a tendency to stick to literal words (copy-and-paste strategy) instead of analyzing the relationship between the wording of question prompts and the information in the texts based on ideas (Author, 2011). This tendency, observed in single text question-answering situations, might make poorer readers more vulnerable

to performing poorly when the documents are not available since they do not have the opportunity to revisit the text under the guide of the questions' demands.

To sum up, one of the main goals for this study is to test, by means of an experimental design, if these well-known effects in single-text regarding availability, and the influence of question type and students' reading skill also appear when facing more complex reading scenarios, that is, answering questions from multiple documents. We might not expect a significant influence of text availability for questions demanding content connections between texts. In such cases, students active processing of information to form inter-text connections (and individual differences in those processes) would play a more important role, than the possible re-access to the texts (Author, 2008; Perfetti et al., 1999).

Given the increasing relevance of functional reading scenarios, it is relevant both for research and practice to test the effects of specific contextual manipulations (such as having the possibility to revisit the texts to answer questions) on students' question-answering performance success with two types of questions requiring intra-text and inter-text integration. To our knowledge, there is no prior work testing this effect of text availability with multiple documents and its interactions with the type of task, despite its potential impact for research and the practice of reading. This is precisely the study we present next.

### **The current study**

According to the previous review, this research examines the influence of text availability and question type, as well as the role of reading skills, on question-answering performance from multiple complementary documents on the topic of History in Secondary Education. The research is timely and relevant for the science of reading, as it explores a specific text and task manipulation that may help us to better understand the cognitive processes underlying question answering from multiple documents in functional reading scenarios. We conducted a paper and pencil experiment with a total of 70 Secondary school students to explore two main research goals related to the following functional reading scenario.

Briefly, adolescent students would be presented with three complementary texts on a history topic and, after a first reading, they were asked to answer a set of questions (intra-text and inter-text questions, within-subjects manipulation) based on the documents. Half of the students would be given the chance to revisit the texts again to

answer the questions (available-texts condition), whereas the other half would not (unavailable-texts condition). Whereas the first condition would assemble current functional reading scenarios, where students may revisit documents to fulfill the task demands, the second situation would be more proximal to classical single-text reading assessments. Given the theoretical and practical consequences of these two reading scenarios, we proposed the following research goals and corresponding predictions.

The first goal was to examine the influence of text availability (available vs. unavailable while answering questions) and question type (intra-text and inter-texts questions) on students' question-answering performance and recall in multiple documents. We hypothesized that students' question-answering performance and delayed recall will vary as a function of text availability and question type, two dimensions that have an impact on the students' assessment. Two predictions were developed in this regard. Our first prediction was that students in the unavailable-text condition would achieve in general lower performance scores as opposed to those in the available-text condition. On the contrary, the possibility to revisit the texts when answering the questions would enhance students' question-answering scores, given their possibility to extract more ideas from the documents based on the challenges posed by the questions rather than their memory access (Britt et al., 2018; Vidal-Abarca et al., 2010). Further, we expected that text availability would interact with the type of questions and the level of ideas recalled. Thus, students would get higher scores for intra-text questions in the available-text condition, as observed in single-text studies (Author, 2017). Intra-text questions would be those tasks whose answer could be located within a single text and would require less degree of integration as opposed to inter-texts questions. The availability of documents would especially favor the former and would not make a significant difference for the latter. In the inter-texts questions latter type of tasks, students would be demanded to make inter-texts inferences. Having the possibility to revisit the texts while answering or not should not make a significant difference, as the difficulty of those tasks would yield in the interconnection of distant ideas (active processing on the part of the learner), and not in the immediate availability of verbatim information (Author, 2009). Thus, we might not expect text-availability differences for inter-text questions. In such cases, students active processing of information to form inter-text connections (and individual differences in those processes) should play a more important role (Author, 2008; Perfetti et al., 1999).

Regarding free recall, we have designed this measure as an overall estimate of recall of ideas present in the different texts (recalled at a literal or inferential level). However, no integration across texts was measured. As such overall estimate, we would expect only an impact of text availability at this level. Thus, we expect a greater representation of verbatim ideas (literal level) 24h after having answered the questions with the texts available. Based on the evidence that text reprocessing improves memory for the text (e.g., Amlund, Kardash, & Kulhavy), we would expect that text availability would facilitate students reinforcing their representation of literal ideas (text-base representation), given the opportunity to revisit the texts in a question-answering phase.

The second goal was to explore whether prior reading skill would moderate the relation between text availability and performance outcomes (i.e., question-answering performance and recall). For this purpose, we used a measure of reading skill in single text, as to our knowledge, there is not a standardized measure of reading comprehension for multiple documents. We hoped that reading skill in single text would be able to show us how students of different reading skill profile behave under task-oriented reading instructions (available-texts condition) and memory-based question-answering situation (unavailable-texts condition). We hypothesized that text availability will differently affect performance of students with different reading skills. Our prediction was that keeping the multiple documents unavailable would negatively affect students with lower prior reading skills, as this task (reading and integrating information from several sources) would be especially challenging for poor readers. Given less-skilled readers' poorer performance in general in question-answering situations in single text (i.e., Author, 2009, 2011), we expect also worse results in unavailable multiple-documents question-answering situations in comparison with high-skilled readers' performance. In contrast, the additional text-search revision opportunities through question-answering cycles with the documents available might compensate for their lack of reading skills (Rouet, 2006).

## **Method**

### **Participants**

A total of seventy Caucasian secondary school students were recruited, from an urban school in a southern European city, in their third year of secondary education (10<sup>th</sup> grade). The school belongs the public national system, in a city context. The

sample did not include participants with reading or learning difficulties and belonged to three equivalent classrooms. They performed the experiment as part of their regular classroom activity. Participation was approved by the school's board and parents' consent to participate in the study was collected. To preserve students' privacy, no personal data were collected. Instead, code identifiers were used to gather the different outcomes.

## Materials

**Standardized reading comprehension test.** (TPC: Martínez, Vidal-Abarca, Sellés, & Gilabert, 2008). It consists of two expository texts (469-549 words), about two topics not included in the Spanish school curriculum: "Penguins" and "Sioux". Each text contains 10 multiple-choice questions that address four specific comprehension processes (i.e., the formation of text ideas, anaphoric inferences, knowledge-based inferences, and macro-idea formation). Students completed it on a paper and pencil version, lasting about 50 minutes ( $\alpha=0.80$ ). This measure was used to determine participants' level of reading skill.

**Multiple documents and questions.** Students received a booklet which included three texts on the same historical topic and a set of questions to be answered based on the texts (7 *intra-text* and 4 *inter-text* questions). Whereas intra-text questions could be answered by identifying the relevant ideas in the single texts (located in one or two consecutive sentences, see Table 1), inter-text questions could only be answered by establishing interconnections of ideas across texts, so the amount of information to be connected was greater, as well as the location of this information was distributed across two or the three texts (see Table 2, for inter-text questions). So, the main difference between both types of tasks was the following: a) the location of the answer (intra-texts or inter-texts) and the number of pieces of information to be integrated (greater for inter-texts questions). According to the intra- and inter-texts levels of processing identified in multiple documents processing (Rouet, 2006), our tasks aimed at inducing those processes. Given that all students would be given the same set of texts including the two types of questions, students' processing at the single and inter-text level would be guaranteed.

The texts had been adapted from National Geographic and covered the historical topic: *The seventeenth century persecution of women as witches*. *Text 1* (362 words) presented the "*French case*"; *Text 2* (398 words) described the case of women

persecution as witches in *Navarra (Spain)*. Finally, *Text 3* (357 words) described the case in *Aragon (Spain)*. The ideas which were unique to each text, and those shared between the texts, were differentiated. The texts could be understood independently, though the combination of the three provided a better understanding of the topic (multiple documents reading).

Just after the set of texts, students were presented with a set of 7 *intra-text* questions which could be answered by identifying simple ideas within the independent texts (i.e., *When did the great era of the witch hunt take place?*). Students were also required to answer 4 *inter-text questions* demanding the interconnection of ideas between the texts (i.e., *According to the texts, what were the consequences of the witch-hunt?*). Thus, for inter-text questions, a complete answer could only be achieved by combining information across texts. A coding template was elaborated, and questions were coded against this template (see Tables 1 and 2, respectively) Correct responses were given one point each, and partial credit was also considered. Two researchers coded 20% of the answers together, until at least 90% of agreement was reached, and then the rest of the answers were coded independently. A percentage of success for all comprehension questions (and separately for intra and inter-text) was obtained.

Table 1. *Summary of intra-text questions and distribution across texts*

| Text, line | Intra-text question                                                                                         | Model of response                                                                                              |
|------------|-------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| T1, 2      | When did the great era of the witch-hunt take place?                                                        | In the first half of the 17th century.                                                                         |
| T1, 10-11  | What are the meetings between witches and wizards called and what was done in them?                         | They are called Lane de Aquellarre and the devil was worshipped.                                               |
| T1, 13     | According to Judge Pierre De Lancre, who caused the great disasters that occurred in France?                | The great disasters were the work of witches and wizards.                                                      |
| T2, 5      | After 80 alleged witches were burned by order of Judge De Lancre, where was the new outbreak of witchcraft? | The new outbreak of witchcraft was located in the area bordering France, more specifically in Zugarramurdi.    |
| T2, 18-19  | In the trials held in Zugarramurdi, what was needed to accuse someone of                                    | To accuse someone of witchcraft, no evidence was needed, a simple suspicion was enough, no defense existed and |

|           |                                                                                                                              |                                                                                                                                                                                  |
|-----------|------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           | being a witch or a wizard?                                                                                                   | confessions under torture were very frequent.                                                                                                                                    |
| T3, 19-20 | Who were the people affected by witchcraft in the Aragonese villages?                                                        | Those affected were mainly young, still unmarried women, who walked around as if they were disturbed and frightened, screaming as if they were drowning and unable to calm down. |
| T3, 26-27 | In the valley of Tena, in the Pyrenees of Huesca, what did the priests decide to do to expel the demons from the demonesses? | To expel the demons, the priests decided to organize processions and fasts and held collective exorcism ceremonies in the churches.                                              |

Table 2. *Summary of inter-text questions and distribution across texts*

| Text, line                                    | Inter-text question                                                                  | Model of response                                                                                                                                                                                                                                                  |
|-----------------------------------------------|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| T1, 4, 10-11.<br>T2, 2-3, 20.<br>T3, 2-4, 19. | According to the texts, what were the consequences of the witch-hunt?                | Spread of panic throughout Europe; persecutions; torture; death of many people.                                                                                                                                                                                    |
| T1, 5.<br>T2, 6.<br>T3, 7-8.                  | What are the common aspects in the cases of witchcraft mentioned in the three texts? | The persecution of witches and wizards took place all over Europe. Women burned at the stake; expansion of witches; witch and wizard hunts; trials and accusations of people without argument.                                                                     |
| T1, 22-23.<br>T2, 9-16, 18-19.                | How did the inquisitors conclude that witchcraft existed?                            | With a simple suspicion you could accuse someone. The inquisitors used the testimony of eyewitnesses or the confession of the accused.                                                                                                                             |
| T1, 10.<br>T2, 22-25.                         | Why did people accused of witchcraft denounce others for being witches and wizards?  | Witches were found in covens, therefore a witch had to know others. During interrogations under torture, the accused were asked for the names of witches and wizards. People were interrogated under torture until they confessed and then sentenced to be burned. |

**Free recall.** We included a measure of free recall, after the reading of the texts and answering of the questions, as an indicator of the students' mental representation from the question-answering task (Author, 2008). Students were told to write down all they remembered from the texts they had read, as a general open recall prompt. Thus, students would produce a written recall product of varying length. The summaries would be segmented into idea units (overall recall outcome), and those units would be

identified as coming literally from the texts (literal memory), or as an elaboration (inferential ideas). Thus, the students' level of ideas recalled after a 24h delay would range from literal to inferential ideas, depending on the way students had produced them in the recall output (literal overlap or greater level elaboration). For instance, a recall idea such as "*The witch hunt occurred in the first half of the seventieth century*" would be coded as literal output. Conversely, the recall product "The case in Aragon was somewhat smaller than the case in France" would be identified as inferential. Two coders rated together 20% of the essays, until at least 90% of agreement had been reached. Then, the rest of the protocols were coded independently. In each of the essays, we identified the number of literal and inferential ideas from the texts. It is important to mention that the free recall test did not attempt to measure learners' ability to integrate information across sources, as we acknowledge in the Discussion section. Rather, it explored how much information participants were able to recall after a 24h. delay, from each of the three texts. As such, recall was an overall estimate of students' retention of ideas (at a literal or inferential level) from any of the texts read in the experimental session.

## **Procedure**

The experiment was performed on paper and pencil in three different sessions of approximately 50 minutes each, and across three non-consecutive days. In *session 1* participants completed the reading comprehension test, by reading two texts and answering 10 multiple-choice comprehension questions per text. The texts were available when answering the questions. *Session 2* was dedicated to the main experimental session. Participants performed the activity as part of their regular class activity. All students were first given the set of texts and were told to read them carefully, as they would need to perform a set of questions afterwards. They were informed that after the reading of the texts some of them would keep the texts, and others would not, to answer a set of comprehension questions.

Students were also prompted to read the documents and raise their hand once they were ready to answer the questions. Once they raised their hand, the experimenter gave them another booklet with the eleven questions and took the documents out or not depending on the experimental group assigned (unavailable vs available). Students were also informed that they would perform a final task the next day to demonstrate what they had learnt. Finally, in the *third session* participants performed the free recall task.

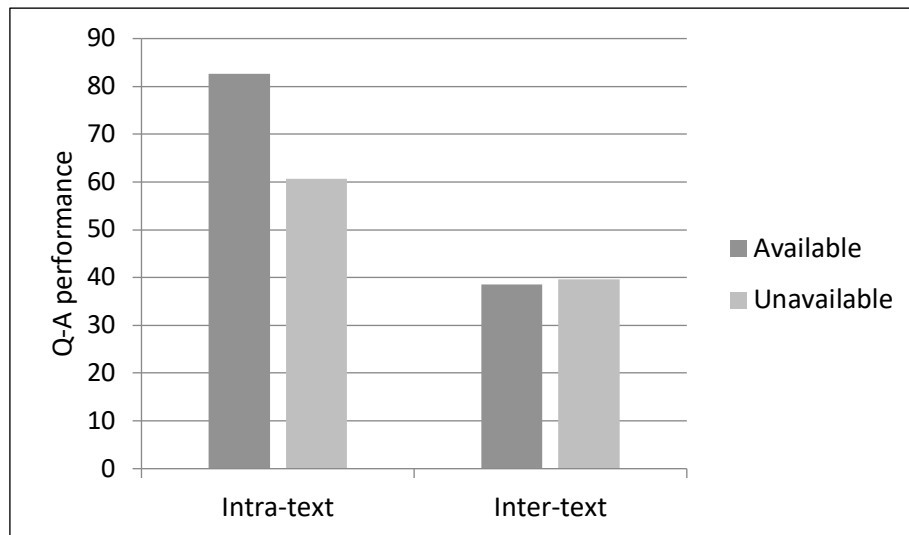
## Analyses and Results

### Effects of text availability and question type on multiple documents question-answering performance and recall

The first aim was to determine to what extent text availability and question type influenced students' question-answering from multiple documents and final recall. To address this aim we conducted an ANOVA with questions performance as dependent variable, and text availability (unavailable, available) and question type (intra-text, inter-text) as independent variables. Text availability was introduced in the analysis as a between-subjects variable and question type as a within-subjects variable.

Results showed a main effect of text availability,  $F(1, 68) = 5.07, p = .028$ , *partial*  $\eta^2 = 0.07$ , with a moderate effect size, indicating that students who had the text available ( $M = 60.58, SD = 20.97$ ) outperformed those who had the text unavailable ( $M = 50.15, SD = 17.74$ ), as we had predicted. Likewise, a main effect of question type was also significant,  $F(1, 68) = 126.36, p < .001$ , *partial*  $\eta^2 = 0.65$ , with a large effect size, suggesting that students tend to provide more correct responses for the intra-text questions ( $M = 71.33, SD = 24.64$ ) as opposed to the inter-text questions ( $M = 39.11, SD = 23.30$ ). These effects were qualified by a significant interaction between text availability and question type (Figure 1),  $F(1, 68) = 15.54, p < .001$ , *partial*  $\eta^2 = 0.19$ , with a large effect size.

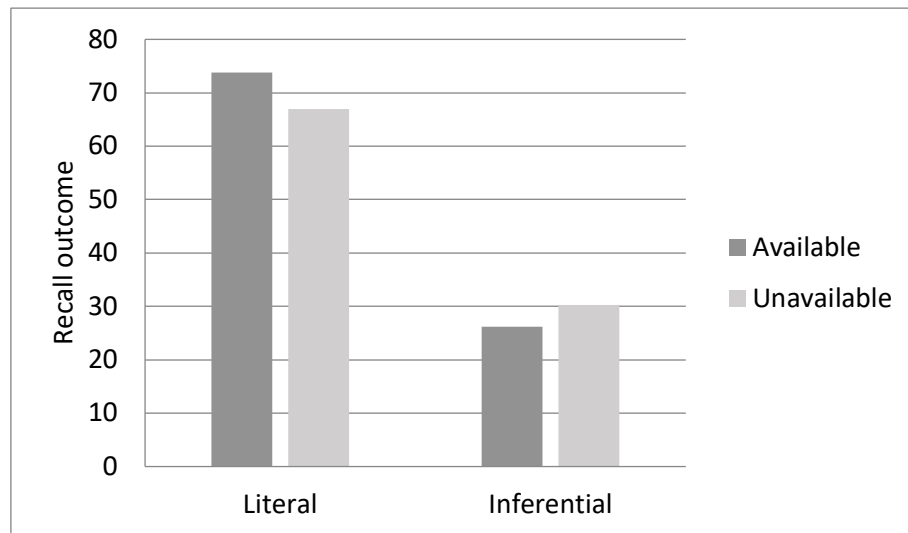
Following our prediction for type of task, results showed that students got higher performance scores for the intra-text questions when the documents were available ( $M = 82.56, SD = 21.91$ ) rather than when the texts were unavailable ( $M = 60.71, SD = 22.49$ ), but no differences were found for the inter-text questions ( $M = 38.60, SD = 25.26$ ;  $M = 39.58, SD = 21.65$ , for the available- and unavailable-text groups, respectively). In other words, having the text available boosted students' performance when answering intra-text questions, but students still struggled to build inter-text inferences even when the multiple documents were at hand. So, the availability of documents would not be a guarantee of success, when the tasks demand higher degrees of elaboration (i.e. inter-text integration).



**Figure 1.** Multiple-Documents Question-Answering Performance as a Function of Text Availability (Available or Unavailable) and Question Type (Intra- or Intertext)

We also ran an ANOVA for the students' recall performance. Text availability was introduced in the analysis as a between-subjects variable and level of ideas recalled (literal, inferential) as a within-subjects variable. Results showed a main effect of level of ideas recalled,  $F(1, 68) = 133.29, p < .001, \text{partial } \eta^2 = 0.66$ , with a large effect size, indicating that students recalled more literal ideas ( $M = 70.27, SD = 17.41$ ) than inferential ideas ( $M = 28.30, SD = 15.56$ ). No main effects of text availability were found,  $F(1, 68) = 0.94, p = .34$ .

Contrary to our hypothesis that having the texts available during the question-answering task would boost students' recall of literal ideas, no interaction effect between text availability and level of ideas recalled came out (Figure 2),  $F(1, 68) = 2.23, p = .14$ . Keeping the texts available did not significantly enhance students' recall of literal ideas ( $M = 73.79, SD = 14.90$ ) over the unavailable-text group ( $M = 66.94, SD = 19.10$ ). No differences were found for the inferential ideas recalled either ( $M = 26.22, SD = 14.90$ ;  $M = 30.28, SD = 16.12$ , for the available- and unavailable-text groups, respectively). Thus, the availability of texts seems to be relevant when considering immediate question-answering performance, but it seems not to play a role for delayed recall.



**Figure 2.** Final Recall Outcomes as a Function of Text Availability (Available, Unavailable) and Level of Ideas (Literal, Inferential)

**Does reading skill moderate the effect of text availability on student's multiple documents question-answering performance and recall?**

The second aim was to examine whether the effect of text availability on students' performance outcomes (i.e., question-answering performance and recall) varies as a function of students' reading skill. Hence, we conducted two simple linear moderation analyses that included students' reading skill as a moderator of the relationship between text availability and performance outcomes. Reading skill was introduced as a continuous variable in percentiles according to the standardized test employed, so that higher scores on this measure indicate greater reading skill while answering questions from an available text. For the *multiple-documents question-answering score*, the model accounted for 43% of the variance in the question-answering performance. Results are reported in Table 3. Whereas both text availability and students' reading skill significantly predicted students' outcomes, moderation analysis revealed a non-significant interaction effect between text availability and reading skill. Contrary to our initial prediction, students' reading skill and text availability did not interact when predicting the question-answering outcomes.

Table 3. *Results of the Moderated Regression Analysis for the Multiple-Documents Question-Answering Performance (N = 70)*

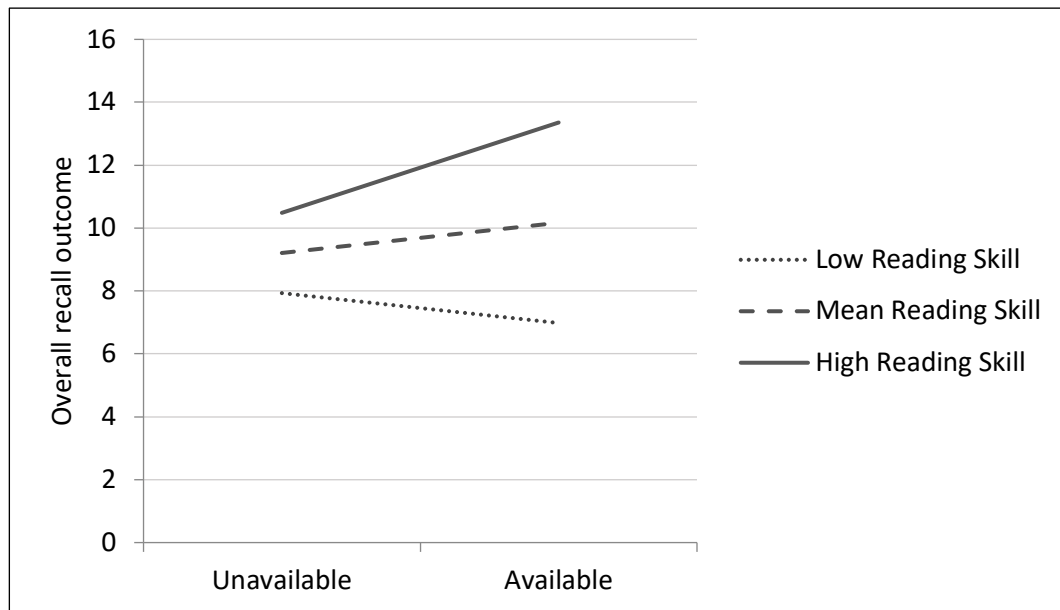
| Parameters               | $\beta$ | SE $\beta$ | $t$   | $p$   | 95% CI |       |
|--------------------------|---------|------------|-------|-------|--------|-------|
|                          |         |            |       |       | Lower  | Upper |
| Constant                 | 54.11   | 1.84       | 29.43 | <.001 | 50.44  | 57.78 |
| Text Availability        | 8.15    | 3.68       | 2.22  | .030  | .80    | 15.49 |
| Reading Skill            | .42     | .07        | 6.37  | <.001 | .29    | .55   |
| Text Av. x Reading Skill | .16     | .13        | 1.20  | .233  | -.10   | .42   |

Another simple linear moderation analysis was run for the *overall recall*. In this case, the model accounted for 31% of the variance in performance. Results are reported in Table 4. Moderation was revealed by a significant interaction between text availability and reading skill, which means that the relationship between text availability and overall recall outcomes is moderated by students' reading skill. However, contrary to our prediction (i.e., keeping the text unavailable would negatively affect students with lower reading skill), we found that students with high reading skill recalled more ideas after performing the task with the documents available. Additionally, reading skill significantly predicted recall outcomes. However, text availability did not significantly predict recall outcomes by itself.

To interpret the moderation effect we examined the simple slopes (Figure 3), which indicate that there was a strong relation between text availability and recall outcomes for students with high reading skill,  $\beta = 2.87$ ,  $t(66) = 2.24$ ,  $p = .029$ . However, there was a non-significant relation at the average value of reading skill,  $\beta = .96$ ,  $t(66) = 1.06$ ,  $p = .293$ . Likewise, there was a non-significant relation between text availability and recall for students with low reading skill,  $\beta = -.95$ ,  $t(66) = -0.74$ ,  $p = .462$ . Thus, it seems that students with higher reading skill built a coherent mental representation of the documents and benefited from the possibility to revisit the text in successive question-answering cycles (i.e., available-texts condition), so that they recalled more knowledge after a 24h. delay.

Table 4. *Results of the Moderated Regression Analysis for the Overall Recall (N = 70)*

| Parameters               | $\beta$ | SE $\beta$ | $t$   | $p$   | 95% CI |       |
|--------------------------|---------|------------|-------|-------|--------|-------|
|                          |         |            |       |       | Lower  | Upper |
| Constant                 | 9.67    | .45        | 21.37 | <.001 | 8.77   | 10.58 |
| Text Availability        | .96     | .91        | 1.06  | .293  | -.85   | 2.77  |
| Reading Skill            | .08     | .02        | 4.84  | <.001 | .05    | .11   |
| Text Av. x Reading Skill | .07     | .03        | 2.10  | .0399 | .003   | .13   |



**Figure 3.** Moderation Effect of Reading Skill Between Text Availability and Overall Recall Outcomes

### Conclusions and Discussion

This study was designed with two goals in mind. First, to examine the influence of text availability and question type on students' question-answering performance from multiple documents. Second, to explore whether students' reading skill moderates the relationship between text availability and performance outcomes (i.e., question-answering performance and recall). We compared two question-answering situations commonly found in education, that is, answering questions from multiple available documents and answering questions with no re-access to the texts. Whereas the first condition would assemble current functional reading scenarios (Britt et al., 2018), the second situation would be more proximal to classical single-text reading assessments (Zoller, Ben-Chaim, D., & Kamm, 1997). The consideration of these two reading approaches in a multiple documents reading scenario is a relevant topic for the science of reading and has potential instructional implications, as we discuss next.

First, as we had predicted, question-answering performance was superior when the text was available than when it was not. This was consistent with previous research in single text (i.e., Author, 2017; Ozuru et al., 2007; Schroeder, 2011). Additionally, we found a significant interaction between text availability and question type, indicating that students obtained higher performance scores in the intra-text questions when the

text was available than when it was not. This also confirms our predictions and aligns with previous research in single-text reading (Author, 2017). So, it seems that availability especially favors the answering of intra-text questions, but it would not exert a facilitating effect when the task requires more cognitive effort due to the need to integrate several pieces of information across different texts. This latter effect has consequences for distance and virtual education and is quite relevant in our current context. In other words, having access to documents would not significantly matter if the task is demanding enough in terms of information to be inter-connected (especially from multiple texts).

Further, the superiority of text availability disappeared in the delayed recall task, which aligns with previous research in single-text reading (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Agarwal & Roediger, 2011). The different pattern of results in the immediate and delayed performance may be explained by the role of retrieval difficulty in promoting long-term retention (Bjork & Bjork, 1992). According to these authors, conditions that most rapidly increase the accessibility of the mental representation (retrieval strength) differ from the conditions that increase the connections of the memory representation with related knowledge (storage strength).

That is, the available-texts condition increases the momentary accessibility of information, making the students interpret this as information stored in memory; however, the retrieval of that information produces small increments in storage strength. In the present experiment, having the texts available when answering the questions might have increased the accessibility of the mental representation, but not the associations of the information with the students' related knowledge, which resulted in short-term benefits that disappeared in a delayed task. This is particularly important with respect to student academic achievement because learning is seldom tested immediately after study.

Second, skilled readers who had the texts available got, in general, higher scores in the question-answering task. Further, and more interesting, results showed that the relationship between text availability and overall recall outcomes was moderated by the students' reading skill, which means that text availability played a different role as a function of students' initial reading status. In other words, skilled readers who had the texts available were able to recall more ideas in the independent recall task after delay as compared with skilled readers who did not have the texts available. Further, and contrary to our initial prediction that less-skilled readers' outcomes would be hindered

by having the texts unavailable, results indicate that text availability did not influence less-skilled readers' outcomes significantly. In fact, less-skilled readers did not perform worse when they did not have the texts available than the less-skilled readers who did have the texts available did. Reading skill seemed to play a role in the available condition, facilitating the greater recall of ideas for better readers because they are likely building stronger mental models. According to the RESOLV model (Britt et al., 2018), functional reading is defined as a problem-solving scenario through successive text inspection and question-answering cycles. According to our results, skilled readers benefited more from the possibly to revisit the texts and reinforced their memory representation, as evidenced in the recall protocols. However, contrary to our prediction, this interaction effect was not found in question answering performance, since both high- and low-skilled readers benefited from having the texts available. It must be acknowledged that the lack of interaction effect may be explained by the low statistical power of this analysis. Future research should employ larger samples in order to explore such interaction effect. That said, it is also important to acknowledge that reading skill was defined by means of a single-text reading comprehension measurement. There is not, to our knowledge, a standardized measure of multiple documents reading. Future research should consider using reading indicators which better align to functional reading competencies.

In all, this set of findings shed light on how contextual- and task-related factors, as text availability and question type respectively, and student-related factors, as reading skill, may influence adolescents' question-answering from multiple documents. These results are relevant for both the science of reading and the professional practice. In terms of research in single- and multiple-documents reading, it contributes to understanding how students respond to very frequent learning and assessment school tasks, such as answering different types of questions with multiple documents at hand or not. Our results are also relevant for assessment and instructional design, as they inform how the access to documents interacts with task type, and students would not especially benefit from this access when the task is sufficiently demanding. In fact, our results suggest that providing access to texts privileges students with better reading skills. Therefore, teachers should carefully consider whether the texts will be available or not while their students answer questions from multiple documents, as the odds of recalling what they read may vary as a function of their reading skill level. This is an issue which

could matter for teachers and educators, compelled to designing on-line courses and distance assessments in the current pandemic crisis.

This is a timely topic for the science of reading, as the ability to integrate information from different sources becomes a critical skill in our today's societies, especially for adolescent readers (Author, 2020; Braasch, 2020). Beyond decoding and comprehension skills in single text, practitioners should also consider incorporating into explicit reading instruction a set of functional reading skills that account for the uses of documents in complex reading scenarios (Rouet & Potocki, 2018). Regarding the practice of reading and the proposal of evidence-based recommendations for reading instructors, our study might contribute to designing assessment instruments and reading interventions which incorporate essential components considered in functional reading scenarios, such as the ability to integrate information within and across texts, the consideration of source information, and the regulation of students' decisions to search or not an available text/set of texts (Britt et al., 2018; Rouet & Potocki, 2018).

Our study was conducted with a sample of adolescents on paper and pencil, and students read three complementary texts on a history topic. This research question-answering tasks aimed at measuring intra-text and inter-text integration, though we did not capture integrative processes in the recall measure, which is a challenge for future studies. Future research could also consider more complex and realistic scenarios, such as navigating multiple documents on the web on a conflicting topic. In addition, on-line measures of task resolution and interaction with the texts could also be obtained and could provide a better overview of the strategies involved when answering questions from multiple documents, with or without the texts available. Moving towards more realistic scenarios might help converge the interests and practices of the scientific community (reading research), policy makers and teachers. We also hope this research contributes to this aim.

## References

- Agarwal, P. K., & Roediger III, H. L. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory*, 19(8), 836–852. doi:10.1080/09658211.2011.613840
- Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger III, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology*, 22(7), 861–876. doi:10.1002/acp.1391

- Amlund, J. T., Kardash, C. A. M., & Kulhavy, R. W. (1986). Repetitive reading and recall of expository text. *Reading Research Quarterly*, 21(1), 49–58. doi: 10.2307/747959
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (pp. 35–67). Hillsdale, NJ: Erlbaum.
- Braasch, J.L.G. (2020). Advances in research on internal and external factors that guide adolescents' reading and learning on the Internet. *Journal for the Study of Education and Development*, 43(1), 210–241. doi:10.1080/02103702.2019.1690851
- Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J.-F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209–233). Mahwah, NJ: Lawrence Erlbaum.
- Britt, M. A., Rouet, J.-F., & Durik, A. (2018). *Literacy beyond text comprehension: A theory of purposeful reading*. New York, NY: Routledge.
- Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, 11(5-6), 489–503. doi:10.1023/A:1008084120205
- Cerdán, R., Candel, C., & Leppink, J. (2018). Cognitive load and learning in the study of multiple documents. *Frontiers in Education*, 3, Article 59. <https://doi.org/10.3389/feduc.2018.00059>
- 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>

Ferrer, A., Vidal-Abarca, E., Serrano, M.A., & 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>

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>

Graesser, A.C., & Franklin, S.P. (1990). QUEST: A model of question-answering. *Discourse Processes*, 13(3), 279–303. <https://doi.org/10.1080/01638539009544760>

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. doi:10.1080/0163853X.2015.1100572

Macedo-Rouet, M., Kammerer, Y., & Cerdán, R. (2020). Introduction to the special issue ‘How adolescents read and learn on the web: Internal and external factors’. *Journal for the Study of Education and Development*, 43(1), 9–18. <https://doi.org/10.1080/02103702.2020.1723987>

Martínez, T., Vidal-Abarca, E., Sellés, P., & Gilabert, R. (2008). Evaluación de las estrategias y procesos de comprensión: El Test de Procesos de Comprensión. *Journal for the Study of Education and Development*, 31(3), 319–332. doi:10.1174/021037008785702956

McCrudden, M.T., & Schraw, G. (2007) Relevance and Goal-Focusing in Text Processing. *Educational Psychology Review*, 19, 113–139. doi:10.1007/s10648-006-9010-7

Oakhill, J. V. (1982). Constructive processes in skilled and less-skilled comprehenders’ memory for sentences. *British Journal of Psychology*, 73, 13–20. doi:10.1111/j.2044-8295.1982.tb01785.x

Oakhill, J. V. (1984). Inferential and memory skills in children’s comprehension of stories. *British Journal of Educational Psychology*, 54, 31–39. doi:10.1111/j.2044-8279.1984.tb00842.x

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.  
doi:10.1080/07370000701632371

Perfetti, C. A., Rouet, J. F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99-122). Mahwah, NJ: Lawrence Erlbaum.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, 17(3), 249-255.  
doi:10.1111/j.1467-9280.2006.01693.x

Rouet, J-F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mahwah, NJ: Lawrence Erlbaum.

Rouet, J. F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, 52(3), 200–215.  
doi:10.1080/00461520.2017.1329015

Rouet, J-F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Greenwich, CT: Information Age Publishing.

Rouet, J. F., & Potocki, A. (2018). From reading comprehension to document literacy: learning to search for, evaluate and integrate information across texts. *Journal for the Study of Education and Development* 41(3), 415–446.  
doi:10.1080/02103702.2018.1480313

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

Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, 103(4), 877–896. doi:10.1037/a0023731

Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology*, 102(4), 817–826. doi:10.1037/a0020062

Zoller, V., Ben-Chaim, D., & Kamm, S. D. (1997). Examination type preferences of college science students and their faculty in Israel and USA: A comparative study.

*School Science and Mathematics*, 97(1), 3–12. doi:10.1111/j.1949-8594.1997.tb17334.x