Writing a synthesis versus reading: strategies involved and impact on comprehension

Núria Castells[1] · **Marta Minguela[1]** · **Esther Nadal[1]**

Abstract

Little evidence is available regarding the differential impact of reading versus reading and writing on multiple source comprehension. The present study aims to: (1) compare the inferential comprehension performance of students in reading versus reading/synthesis conditions; (2) explore the impact of performing the tasks on paper versus on screen with *Read&Answer* (R&A) software; and (3) explore the extent to which rereading, notetaking, and the quality of the written synthesis can explain student's comprehension scores. For the students in the synthesis condition, we also examined the relationship between the quality of the synthesis they produced and the comprehension they achieved. 155 psychology undergraduates were randomly assigned either to the reading (n = 78) or to the reading/synthesis condition (n = 77). From this sample, 79 participants carried out the task with the *Read&Answer* software, and 76 solved the task on paper. All the students took a prior knowledge questionnaire, and read three complementary texts about the conception of intelligence. Students in the reading condition answered an inferential comprehension test, whereas students in the synthesis condition were asked to write a synthesis before taking the same test. Results show no differences in comprehension between students in the four conditions (task and media). There was no significant association between rereading and task condition. However, students in the synthesis condition were more likely to take notes. We found that two of the categories for the quality of the synthesis, textual organization and accuracy of content had an impact on inferential comprehension for the participants who wrote it. The quality of the synthesis mediated between student's prior knowledge and inferential comprehension.

Keywords Multiple source comprehension · Synthesis text · Rereading · Note-taking · Inferential comprehension

✉ Núria Castells
nuria.castells@ub.edu

[1] Department of Cognition, Developmental and Educational Psychology, Faculty of Psychology, Universitat de Barcelona, Pg. de la Vall d'Hebrón, 171, 08035 Barcelona, Spain

Introduction

When students are doing course work and taking exams, they typically face the challenge of dealing with multiple sources that contain information that may be divergent or even partly contradictory. The ability to access information transmitted in multiple texts, the capacity to process it, integrate it, and adopt a critical perspective towards it constitutes a learning process that cannot be taken as a given when a student enters university. Several studies (Mateos & Solé, 2009; Miras et al., 2013; Spivey, 1997) stress the difficulty of learning from multiple sources, an activity typical of the university experience. Since students may approach learning from multiple texts in diverse ways, more research is needed to examine the different strategies they use and how these strategies help them to achieve this aim.

Previous research has shown that asking students to write using multiple sources enabled them to elaborate and organize the information they read in greater depth and enhanced their reading comprehension and final learning (Wiley & Voss, 1999). However, it is not clear how specific writing tasks and the strategies involved in them affect comprehension outcomes (Hebert et al., 2013). The present paper explores this issue by comparing the inferential comprehension outcomes of students who read several sources and students who also wrote a text after reading the sources, and by analyzing the strategies they used to deal with these two tasks. The use of the *Read&Answer* software allowed us to keep track of the strategies students applied.

Multiple sources comprehension

Comprehending multiple texts in depth is not an easy task (Britt et al., 2017; List & Alexander, 2019). Following on from Kintsch (1998), when dealing with a single text, students must progress towards different levels of representation: surface level (decode and understand the words and sentences), text-based (representation of the ideas explicitly mentioned in the text) and situation model (integrating information with prior knowledge and generating inferences). Inference is essential to establish the local or overall coherence of a text, through connecting several units of information (Basaraba et al., 2013). Additionally, when dealing with multiple documents, students must process and integrate information across texts to form a *documents model* (Britt et al., 1999), which is an integrated cognitive representation of the main content of the texts (called the Integrated Model), as well as a representation of source information (i.e., authorship, form, rhetorical goals…) together with the construction of links between texts as consistent or contradictory with one another (called the Intertext Model). For Barzilai et al. (2018), integration is central in multiple text comprehension because it involves articulating information from different sources in order to achieve aims such as understanding an issue or writing a synthesis. The capacity to infer information from the sources and across the sources seems to be crucial for integrating and achieving a comprehension of multiple texts (List & Alexander, 2019). The cross-textual linking, which may be either low-level (referring to explicit content

introduced in the sources) or high-level (categories or functional aspects of the source texts that are not always explicit and should be inferred), might be facilitated by using writing in diverse forms (e.g., note-taking, graphic organizers, summarizing, synthesizing—List & Alexander, 2019).

Synthesis writing

The idea that writing has a positive impact on learning has a long tradition, doubtlessly linked to the acknowledgement of the potentially epistemic nature of the activity (Galbraith & Baaijen, 2018). In specific circumstances, writing enables one to incorporate information, restructure knowledge and make conceptual changes (Scardamalia & Bereiter, 1987; Schumacher & Nash, 1991). Indeed, according to the functional view of reading-writing connections (Fitzgerald & Shanahan, 2020), comprehension can be enhanced by writing about texts while performing actions of different types (Applebee, 1984; Graham & Hebert, 2011; Hebert et al., 2013). First, writers must select the most relevant information from the texts, focusing on the content. Secondly, when writing, students need to organize ideas from the texts coherently, building connections between them. Thirdly, since the written document is permanent, it allows for reflection, in the sense that the writer can review, connect and construct new versions of the ideas in the text. Fourthly, the process requires the writer to take active decisions about the text's content and/or structure. Finally, since writers need to put text ideas into their own words, their understanding of these ideas may be increased.

At undergraduate level, students are habitually required to integrate information from multiple sources into a written text or document (i.e., dissertations, research work, essays, etc.). Despite differences in the types of documents students are assigned, from a cognitive perspective they share some basic aspects that make the task challenging. For Barzilai et al. (2018), and based on empirical studies, when people read-to-write from sources (Nelson & King, 2022), they deal with the following issues: understanding the content of the source texts and the aims of the authors; understanding the relations/connections that can be established between them; selecting the information relevant to achieving the aim; identifying a thread that helps to organize the selected information; taking decisions about the structure; linking the ideas in writing and revising during the process and at the end.

This explanation highlights the three main operations of synthesis writing: selecting, organizing and connecting the information (Barzilai et al., 2018; Spivey & King, 1989). Therefore, when they synthesize, students may perform various roles: e.g., as readers of source texts, writers of notes, readers of notes, writers of drafts and readers of their own text (Mateos & Solé, 2009). Several researchers have hypothesized that, since reading leads to learning and writing has epistemic potential (Nelson & King, 2022), the transitions between sources and the students' own text when students assume these changing roles may explain the transformation of their knowledge and their enhanced comprehension of the source texts (Moran & Billen, 2014; Nelson & King, 2022).

Writing and reading multiple sources and comprehension

Several meta-analyses have pointed to the fact that writing interventions which involve writing about a single text had an impact on comprehension outcomes (Bangert-Drowns et al., 2004; Graham & Hebert, 2011; Hebert et al., 2013). However, would reading-to-write from sources have the same impact on comprehension?

Research on the impact of synthesis writing on comprehension has not achieved conclusive results. Before revising the most relevant results from research, two precautions should be mentioned. First, the writing from sources tasks assigned to students are not always clearly conceptualized; that is, despite sometimes being called "synthesis", they may involve not just selecting, organizing and connecting information, but also providing one's own opinion on, or criticizing, a subject. Secondly, the tasks aimed at assessing comprehension after performing the writing from sources task may only assess comprehension and/or learning (when the task involves retention of source information). However, since reading might lead to learning, as noted by Nelson and King (2022), in the following review we have considered that tasks placed after the reading-to-write activity that assess learning are also assessing the comprehension, in a general sense, of the source information.

Bearing the aforementioned precautions in mind, we will review studies that have compared reading-to-write from sources with just reading, and which have included comprehension assessments. Wiley and Voss (1999) found that asking undergraduate students to write an argumentative synthesis after reading multiple history sources about nineteenth-century Ireland improved their performance on inference comprehension and analogy tasks more than writing narratives, summaries or explanations. The authors attributed these results to the fact that the argument text written by the students contained more connected, integrated and transformed ideas than the other types of texts written under the other research conditions.

Following on from the study by Wiley and Voss, Le Bigot and Rouet (2007) asked 52 university students to write either a summary (presenting the main ideas) or an argument (expressing an opinion) on social influence after reading multisource hypertexts. After producing their text, they were asked to complete a comprehension questionnaire. The results from this study differ from those of Wiley and Voss (1999). Although argument texts were found to show more causal/consequence connectives and more transformed ideas than summary texts (which contained a greater use of temporal connectives and more borrowed ideas from the source texts), they did not improve comprehension, as measured by the identification of main ideas and detection of local details. Nevertheless, writing a summary enhanced both kinds of comprehension.

Gil et al. (2010) also compared writing a summary to writing an argument from several texts on climate change. Unlike Le Bigot and Rouet (2007), they found that students who wrote a summary text produced more transformations, covered all the information in the text materials and integrated the information to a greater extent than students asked to write in the argument text condition. Furthermore, students in the summary group obtained higher scores on the measures that were developed to assess either superficial or inferential comprehension. Notably, in these two latter studies, the summary task was closer to a synthesis task than the argumentative

activity. For the summary task, Le Bigot and Rouet (2007) asked students to collect the main ideas from the texts, while Gil et al. (2010) asked for a report summarizing the causes of climate change presented in the texts. Therefore, students were able to integrate information from the texts without losing the intrinsic meaning of the information contained in them. In contrast, to write the argumentative proposal, in both studies students had to provide their own opinion, which requires them to adopt an external and critical stance and to distance themselves from the meaning of the texts. Cerdán and Vidal-Abarca (2008) decided to compare two clearly distinct tasks: writing an essay in order to answer an intertextual open question after reading three sources, and answering four intratextual open questions (the responses were placed in only one text). Students performed the reading and writing tasks using a special software program, *Read&Answer*, and were allowed to go back and forth from their products to the source texts and vice versa. A sample of the students performed the tasks while thinking aloud. After finishing the tasks, students completed a superficial comprehension task (sentence verification task), and a deep learning task (answering several open questions related to a problem situation). As predicted by the researchers, examination of the students' processes revealed that those who wrote the essay integrated more relevant units of information than those who answered the intratextual questions, given that the latter showed more single-unit processing. In terms of comprehension, writing the essay produced better results in the deep learning task than answering intratextual questions; in contrast, there were no differences between the two tasks in the superficial understanding test. Cerdán and Vidal-Abarca (2008) concluded that writing the response to intertextual questions fosters deeper learning than intratextual ones, and that the two kinds of activities did not differ at the level of superficial comprehension.

In short, task instructions seem to be relevant to the way students understand the specificities of the writing task (Barzilai et al., 2018; Spivey, 1997). Asking students to pay attention to relevant information from the texts and to integrate it (i.e., by writing a synthesis) might help them to process information in greater depth, through building more integrations between the information contained in the texts, than asking them to give their opinion on a subject (which does not necessarily require them to select, organize and connect information). In this regard, Miras et al. (2013) found that higher education students who wrote better syntheses than their peers after reading three texts responded better to different kinds of questions that involved retrieving, interpreting and reflecting on information. Since the processes conducted by students when writing a synthesis allow them to connect more intra- and intertextual information, they perform better on deep comprehension measures. However, to the best of our knowledge there is no evidence of the different impact of only reading multiple documents versus reading and writing a synthesis on comprehension.

One final concern related to the previous studies would be that all of them ask some or all of their participants to read on a computer, in some cases using special software such as *Read&Answer,* in order to collect data from them all (Cerdán & Vidal-Abarca, 2008), or from a sample of their participants (Gil et al., 2010). For years, researchers have asked participants to read on screen or on paper without specifically addressing the possible influence of the media on the level of

comprehension. However, some recent meta-analyses comparing students' comprehension outcomes when reading expository texts on paper or on screen have concluded that solving the tasks on screen negatively influences the level of comprehension finally achieved (Clinton, 2019; Delgado et al., 2018). It is important to add that these results were obtained under time constraints, but were not evident when participants had all the time they wanted to solve the tasks. Additionally, Gil et al. (2010) and Vidal-Abarca et al. (2011) have pointed out that using *Read&Answer* software does not negatively influence students' performance when compared with paper-and-pencil tasks.

In a different vein, with the exception of Cerdán and Vidal-Abarca (2008), the studies reviewed provide little information on the specific strategies involved in solving the tasks analyzed (e.g., rereading and note-taking). Is it the task in itself that leads to deep-level comprehension results? Or is it the *specific strategies* deployed by the students in order to solve the tasks that are responsible for the results?

Strategies to deal with multiple sources: rereading and note-taking

Just as the research suggests that essentially the same cognitive processes (generating ideas, questioning, hypothesizing, making inferences and supervising, etc.; see Tierney & Shanahan, 1996) come into play in both reading and writing, albeit with different weights, students may decide to use some or other of these processes or skills in a strategic way to solve the tasks they are assigned. According to Kirby (1988) and Dinsmore (2018), strategies are skills intentionally and consciously used to achieve an aim. Comprehension strategies refer to intentional attempts to control and modify meaning construction when reading a text (Afflerbach et al., 2008). Writing strategies involve the use of processes by the writer to improve the success of their writing (Baker & Boonkit, 2004). In this study, two strategies are prioritized: rereading and note-taking.

Rereading the text or parts of it is a strategy that has been highlighted as useful for achieving deep comprehension (Stine-Morrow et al., 2004). Goldman and Saul (1990), Hyönä et al. (2002) and Minguela et al. (2015) have described different overall reading processes that readers spontaneously engage in when approaching texts before performing a post-reading task, among which rereading seems to be key. In the Minguela et al. (2015) study, students who selectively reread some specific parts of the text obtained better results in a deep comprehension task containing questions that required integration of information across the text and/or reasoning beyond its literal content.

In the case of writing from various texts, researchers have also stressed the relevance of rereading to explain the differences in quality of the texts produced. Working from this perspective, McGinley (1992) was the first to mention the relevance of rereading (i.e., when students read either the source texts or their own products once again) for composing from sources. The importance of rereading documents in the synthesis task has been stressed by several research studies (see Lenski & Johns, 1997; Martínez et al., 2015; Solé et al., 2013; Vandermeulen et al., 2019). In addition to writing better products themselves, students who reread the texts (sometimes

together with performing other activities, such as writing a draft or taking notes) showed either better learning of the contents of the texts (in the intervention study conducted by Martínez et al., 2015) or better comprehension (Solé et al., 2013).

Other studies in the area of writing a synthesis from sources have examined another strategy that students can performed spontaneously: note-taking (Hagen et al., 2014). However, the role of taking notes in writing a synthesis is less clear than the role of rereading (Bednall & Kehoe, 2011; Dovey, 2010). In single text reading, Gil et al. (2008) compared students who were required to take notes to those who did not take notes; they found that, firstly, taking notes reduced the time students devoted to processing the information in the source, and secondly, students who took notes identified fewer inferential sentences than students who did not. This finding was also replicated by some studies that required students to read from multiple sources while allowing them to take notes. These studies found that students who mostly copy from the texts tended to obtain poorer comprehension results (Hagen et al., 2014) and wrote weak synthesis essays (Luo & Kiewra, 2019). However, results for this question are not conclusive, because Kobayashi (2009) found that when students were given a clear purpose (i.e., finding relations between texts), those who used external strategy tools (e.g., taking notes—Kobayashi, 2009), outperformed students who did not use tools in a test with intertextual questions.

To summarize, these studies suggest the importance of strategies when students read texts or write from sources. Producing a synthesis from different texts requires them to read and reread them in order to identify the relevant information and to elaborate and integrate it (Nelson, 2008; Solé et al., 2013). In this demanding task, taking notes may be of considerable help. However, the strategies involved in reading, and in reading and writing from multiple sources, require fuller examination, as does the impact of these practices on comprehension.

This paper aims to explore this issue in depth. Specifically, it pursues two sets of aims: one for a general sample of students, and another one for a subsample of these students that performed the task on a computer and using the *Read&Answer* (R&A) software.

Regarding the general sample, firstly, we aim to explore whether there were differences in the comprehension performance of participants depending on two variables (i.e., task—reading multiple texts and writing a synthesis vs. reading multiple texts only; and media—paper vs. screen with R&A software). Secondly, we aim to analyze whether there were differences in the quality of the synthesis written by participants in the reading/synthesis condition depending on the media used, and the relationship between the quality of their texts and the reading comprehension they achieved.

In line with previous research (Cerdán & Vidal-Abarca, 2008; Le Bigot & Rouet, 2007; Wiley & Voss, 1999), we expect students who write a synthesis to obtain better results in a deeper-level, inferential comprehension test of multiple sources than students who only read the texts. We also predict that the media used for performing the two tasks would not have any differential impact on students' comprehension results (Gil et al., 2010; Vidal-Abarca et al., 2011).

In the subsample that used R&A software, we assess the strategies used by students to solve reading/no synthesis and reading/synthesis tasks. In particular, for this

Table 1 Sample distributed in the four conditions according to media and task

Media	Task	
	Reading/no synthesis task	Reading/ synthesis task
Paper and pencil	38	38
R&A	40	39

subsample of participants, we aim to: (1) explore whether their use of rereading and note-taking was associated with the task they had been assigned; and (2) explore the relationship between the use of these strategies (and the quality of the synthesis written by those who did so) and students' comprehension performance.

Firstly, we expect that participants in the reading/synthesis condition would engage in more rereading and note-taking than those in the reading/no synthesis condition. Secondly, we expect that rereading and note-taking (provided that the notes are not mere copies of the source texts) would be associated with a higher level of inferential comprehension in each condition. For participants in the reading/ synthesis condition, we hypothesize that the use of these strategies would be associated with the quality of the synthesis written by the students in terms of selection and integration of relevant and accurate information and organization, if the notes taken and the synthesis text produced are not copied. This text quality would be associated with the level of inferential comprehension finally achieved.

Method

Participants

The sample was comprised of 155 first-year undergraduate psychology students at a state-run Spanish university who volunteered to take part in the study at the beginning of the academic year. The sample included 115 women and 40 men with an overall mean age of 19.64 years ($SD = 4.92$). All participants received all the details about the aims and tasks of the project. After giving their informed consent, they were randomly assigned to one of the four conditions depending on the task they were assigned and on the media they used to perform it (see Table 1): (1) reading/no synthesis, paper and pencil; (2) reading/synthesis, paper and pencil; (3) reading/no synthesis, *Read&Answer*; and (4) reading/synthesis, *Read&Answer*.

Thus, some participants (n = 79; age: $M = 20.19$; $SD = 6.16$) performed the assigned task (either reading/no synthesis or reading/synthesis) on a computer and using the *Read&Answer* (R&A) software. We used this software as a means of keeping track of the strategies they used to face the assigned task—specifically, in order to register their rereading without making them read aloud (reading/no synthesis condition: 62.5% women; reading/synthesis condition: 71.8% women).

Table 2 Means and standard deviations for the prior knowledge test on each condition (max. 15)

	Reading/no synthesis task		Reading/synthesis task	
	M	SD	M	SD
Task with paper and pencil	10.72	2.32	11.09	2.23
Task with R&A	11.08	2.21	10.69	2.57
Total	10.9	2.25	10.9	2.39

Finally, participants in the four conditions did not differ in their prior knowledge regarding the content of the texts they were asked to read to perform the task they were assigned (see Table 2) according to an ANOVA performed with task and media as factors. No main effects of the two factors were found either (task: $F(1,156)=0.001$, $p=0.990$, partial $\eta^2<0.001$; media: $F(1,156)=0.003$, $p=0.955$, partial $\eta 2<0.001$).

Materials

Texts

Three complementary texts about the current conception of intelligence, taken from reliable sources used in psychology, were adapted for the study. The first was an argumentative piece that makes a case for theories upholding the diverse character of intelligence as against unitary theories (973 words). The second was an expository piece that presented Gardner's theory of multiple intelligences (909 words). The third, which was also expository, presented Sternberg's concept of successful intelligence (611 words). The level of difficulty of the texts was considered suitable for higher education students by a panel of seven experts and according to the Flesch-Szigriszt Index (INFLESZ; adaptation by Barrio-Cantalejo et al., 2008) of readability. According to Barrio-Cantalejo et al. (2008), an INFLESZ value between 40 and 55 is considered appropriate for this age group, and the values obtained ranged from 43 to 47.

Understanding these three texts involved comparing, contrasting and integrating information. Three intertextual ideas that appeared in the texts were required for the correct completion of the task: (1) the contrast between intelligence as a unitary quality measured by testing and intelligence as a multiple, diverse competence; (2) the contrast between intelligence as an inherited trait and intelligence as a factor that can be modified by the influence of the environment; (3) the importance of maintaining a balance between intelligences and knowing how to use them appropriately (see "Appendix A" with two excerpts of the texts).

Prior knowledge questionnaire

A prior knowledge questionnaire about "intelligence" was produced, consisting of 15 statements whose reliability was based on CFA and corrected for test length

using the Spearman-Brown extension = 0.64, see "Appendix C"). The participants had to indicate whether the statements were true or false, and the total number of correct answers was recorded (see examples in "Appendix B").

Inferential comprehension test

We applied a verification task that consisted of 25 items that required students to make inferences of diverse complexity (intratextual and intertextual) (reliability based on CFA and corrected for test length using the Spearman–Brown extension = 0.83, see "Appendix C"). The intratextual inferential questions required students to infer information that was not explicit but could be deduced and was consistent with the text content, and/or to integrate information spread throughout the text (Graesser, et al., 2010). The intertextual inferential questions required students to link information contained in different source texts and/or combining prior knowledge with the information provided by the text (Castells et al., 2021). We decided to create a single test with deep-level inferential questions because this is the kind of task in which previous research has found the most significant differences (Cerdán & Vidal-Abarca, 2008; Gil et al., 2010; Wiley & Voss, 1999). The appropriateness of the questions asked in the test was assessed by a panel of seven experts. A pilot study was carried out to obtain additional information on appropriateness. As in the prior knowledge questionnaire, participants had to indicate whether the statements were true or false, and the total number of correct answers was recorded (see "Appendix B").

The reliability of the prior knowledge test measure was lower than desired. However, other research in this area has used similar reliability scores (see Gil et al., 2010; Bråten & Strømsø, 2009), because methodologists (Gronlund, 1985; Kerlinger & Lee, 2000) argue that a reliability estimate may or may not be acceptable depending on how the measure is used and what type of decision is based on the measurements.

Procedure

All participants performed the tasks in collective sessions lasting approximately 90 min. We collected data in a total of four sessions, one per condition (paper and pencil reading/no synthesis; paper and pencil reading/synthesis; R&A reading/no synthesis; R&A reading/synthesis). Each subsample performed the task in a different room supervised by a researcher, and once the students ended the task, they left the room. For participants in the paper and pencil conditions, the prior knowledge questionnaire was administered first. Once the students had finished it, they were given the three texts presented in the same handout in a counterbalanced form. The instructions for the reading task were: *Below you will find three texts with information on the concept of intelligence. Read them carefully, paying attention to the most important information they provide on the subject (intelligence) because afterwards you will have to answer questions related to the content of the texts. You may read*

them as many times as you wish, highlighting phrases, writing notes, etc.*, but you must answer the questions without the texts or your notes in front of you.*

The instructions for the synthesis task added the following requirement to the instructions above: *Write a text including and, above all, integrating, what is most important in the texts you have read on the topic (intelligence).*

Students were also informed that when answering the comprehension questions they would not be allowed access to any of the materials (e.g., the synthesis text or the notes, if they had taken any).

The participants in the *Read&Answer* (R&A) conditions (i.e., computer-based environment) performed the assigned task (reading/no synthesis or reading/synthesis) in front of a computer on separate R&A screens, following the same procedure.[1]

The *R&A* software (Vidal-Abarca et al., 2011) allows the collection of online data by registering some indicators of the reading process, e.g., the sequence of actions that were followed (sequence of segments of the text or specific question that participants read), the number of actions performed and the fragment of the text that was unmasked. *R&A* presents the texts and the questions or task on different screens in masked form. To unmask readers have to click on each segment of the text or task; only one segment is visible at a time. The participants can go back and forth through the text in any order, either on the same screen or between screens in longer texts (or when there is more than one source-text, as in this study) by clicking on the arrow icons. Likewise, they can go back and forth from the texts to the task screen by clicking on the question mark icon (see Fig. 1). The kind of masking that is used enables participants to see the layout of the text (i.e., the text structure, the form of paragraphs, headings, subheadings, etc.) even though the text is masked (as shown in Fig. 1).

The use of this software allows the collection of online data by recording all the participants' actions. It provides some indicators of the reading process, such as the segments of the text or questions that were unmasked, the sequence in which they were accessed, and the time spent on each action.

The segmentation of the text that the participant can make visible by clicking on it is decided by the researcher during the planning stage of the experiment. For instance, we chose to segment the texts as follows: Text 1 was divided into 16 fragments, of which seven dealt with the intertextual ideas; Text 2 was divided into 12 fragments, of which four dealt with the intertextual ideas, and Text 3 was divided into nine fragments, of which three dealt with the intertextual ideas.

---

[1] Participants in the paper and pencil conditions received all instructions and source texts on paper, performed the assigned task on paper, and were allowed to take notes on a separate sheet. Participants in the R&A conditions received all instructions and source-text on screen, performed the assigned task on screen, and were allowed to take notes on paper if they wished. We chose not to include an additional screen on R&A to take notes so as to avoid complicating the procedure with this software for participants and causing confusion between the notes screen and the synthesis screen, or to avoid the risk that some students might regard this as a requirement rather than as a strategy that they could decide to use or not. The average time taken to complete the tasks by the students in the R&A condition was 60 min for the reading/synthesis and 40 min for the reading/no synthesis condition.

Fig. 1 Read&Answer screen

For both tasks, and as in the paper conditions, the participants were allowed to spend as long as they wanted on reading and on producing their written texts. Only when they decided to answer the reading comprehension test were the texts, notes and the synthesis text (for students in the reading/synthesis conditions) made inaccessible.

Analysis of text quality

Following on from previous studies (Boscolo et al., 2007; Nadal et al., 2021), we analyzed the quality of the texts, considering the dimensions presented in Table 3. First, following on from Magliano et al. (1999), we segmented each synthesis into idea units. An idea unit contained a main verb. If an utterance had two verbs and one agent, it was treated as having two separate idea units (Gil et al., 2010). Then, following on from Luo and Kiewra (2019), the ideas were related to the source text they came from so as to obtain a picture of how students had organized and connected them in their own text. In this way, we were able to identify the number of main ideas that required intertextual connections (in a similar vein as Luo & Kiewra, 2019), which could be up to three. It also allowed us to identify the diverse types of organization (following on from Martínez et al., 2015; Nadal et al., 2021). Like Reynolds and Perin (2009) and Zhang (2013), we found mistakes or incomplete and incomprehensible ideas. In addition, we analyzed the

Table 3 Criteria for assessing the written synthesis with the scores corresponding to each category

Criteria/dimensions	Scores and categories
Relevant ideas (ideas that require integration of information, see *Texts* description) to answer the synthesis question	0—Includes more than 75% of irrelevant ideas. May include one relevant idea 1—Includes between 50–75% of irrelevant ideas. May include one relevant idea 2—Includes up to 50% of irrelevant information and at least two main ideas 3—Includes just the three relevant ideas
Accuracy of content	0—Presence of incomplete ideas and/or mistakes in the ideas (more than 50% of the ideas) that affect comprehension of the synthesis 1—Presence of some incomplete and incomprehensible ideas and/or mistakes in the ideas (between 1 and 50% of the ideas) that partially affect comprehension of the synthesis 2—No mistakes and/or incomplete and incomprehensible ideas
Textual organization	0—List of ideas from two or three texts 1—Juxtaposed summaries 2—Combination of fragments from the sources without a clear axis (what is more important about intelligence). Ideas from two of the texts are combined, but the synthesis text lacks coherence and cohesion 3—Organization of ideas from the three texts around a clear structural axis. Information from the three texts is combined in order to identify the most important data about intelligence, in a coherent and cohesive way

degree of copying from the source texts, since this has been shown to be a relevant factor for explaining comprehension results (Luo & Kiewra, 2019).

Two independent raters, researchers who were not otherwise involved in the study, coded the first 20 texts. Two of the authors trained them in two sessions, providing them with the categories and the specific criteria, and using three written syntheses from the study that were not included in the random sampling. After that, the raters coded the first 20 texts (25% of all the texts, written either on paper or on screen). Interrater reliability was adequate for most dimensions (Relevant ideas, $K=0.67$; Accuracy of content, $K=0.81$; Textual organization, $K=0.84$). Disagreements were resolved through discussion and then the rest of the syntheses were distributed between the two raters for evaluation.

Additionally, the *percentage of verbatim copying from the three source texts in the synthesis* text was calculated using the WCopyfind tool and by applying the criterion that seven words copied literally (subject [article + noun], verb, predicate [article/preposition + noun + complement]) represented one instance of literal copying (following on Nadal et al., 2021).

Finally, the *percentage of copying from the notes in the synthesis* was calculated using the same tool and applying a criterion of four words (subject [noun],

verb, predicate [noun + complement]). In this case, we used this criterion because the notes were shorter than the three source texts (following on Van Weijen et al., 2018).

Analysis of the strategies in the R&A subsample

We analyzed the participants' use of rereading and note-taking. In particular:

Rereading strategy

Students were allowed to go back and forth in the texts as they wanted (during reading and/or, for those in the reading/synthesis condition, while they wrote their synthesis as well). Looking at the R&A output, all unmasking actions of less than 1 s that were performed repeatedly and consecutively on the same segment were deleted, as they reflected accidental double-clicks on a segment when participants intended to unmask it. After doing so, all the unmasking actions registered by the software that constituted a revisit of a segment that had already been previously unmasked were considered as "rereading". Regarding this strategy, we analyzed the *number of rereadings*, i.e., the number of times in which participants revisited some segment of the texts.

This variable was considered at two points of the process:

- *Number of rereadings during the initial reading of the text* Rereadings that occurred before facing the assigned task (either taking the comprehension test or writing the synthesis and then taking the reading test, depending on the condition). We focused on this initial rereading because it allowed us to compare rereading between participants in the two conditions (reading/no synthesis vs. reading/synthesis).
- *Number of rereadings during writing of the synthesis text* Additionally, for the students in the reading/synthesis condition, this variable was also considered by focusing on participants' rereading while they composed their text, which gave us a measure of how hybrid their synthesis writing was.

Note-taking strategy

Students were allowed to take notes during the reading/no synthesis and the reading/ synthesis task. Regarding this strategy, we analyzed the following variables:

- *Note-taking* The presence or absence of note-taking during the execution of the task was coded dichotomously.
- *Percentage of copy from the source texts contained in the notes* We used the WCopyfind tool, applying the criterion that four words copied literally (subject [noun], verb, predicate [noun + complement]) represented one instance of literal copying (following on Van Weijen et al., 2018).

Statistical analysis

After the analyses of the reading strategies and the written products, we performed different statistical analyses to meet our research aims.

Comprehension results for participants in the different conditions

Regarding the general sample (n = 155), and to explore whether there were differences in the comprehension performance of participants depending on the task and the media, we performed an ANCOVA analysis with inferential reading comprehension as the dependent variable (DV), tasks (reading/no synthesis vs. reading/synthesis) and media (paper and pencil vs. R&A) as factors, and prior knowledge as covariate.

Text quality in the different conditions and its relation to comprehension

In all participants assigned to the reading/synthesis task (n = 77) we tested whether there were differences in the textual organization, the accuracy of content, and the relevance of the ideas included in the synthesis they produced depending on the media used. To do so, we performed a Mann Whitney U test comparing the results obtained by participants using the R&A and those in the paper and pencil condition for each of the dimensions used to assess text quality (i.e., relevance of ideas, accuracy of content, and textual organization). Furthermore, and to test the relation between the different dimensions of text quality (textual organization, accuracy of content, and relevance of ideas) and inferential comprehension, we performed Spearman's correlation analysis. Non-parametric analyses were chosen for these data because the variables related to text quality are ordinal.

Strategy used to address different tasks in the R&A participants

We explored whether participants in the R&A subsample used rereading and note-taking differently depending on the task they had been assigned by means of chi-square analysis and mean comparison analyses (either Student's *t* test or Mann–Whitney U test, depending on the compliance of assumptions). Specifically, on the one hand, a chi-square analysis was performed to determine whether there was any association between the task (reading/no synthesis vs. reading/synthesis) and the presence or absence of note-taking. On the other hand, we compared participants in the reading/no synthesis task condition and participants in the reading/synthesis conditions on the remaining variables concerning the use of strategies (i.e., a Mann–Whitney U test was performed to compare the number of rereadings in the two groups of participants, due to non-normality of the DV; and

	Reading/no synthesis task		Reading/synthesis task	
	M	SD	M	SD
Task with paper and pencil	18.74	3.15	18.58	2.93
Task with R&A	18.65	2.95	17.9	2.86
Total	18.69	3.03	18.23	2.9

Table 4 Means and standard deviations for the inferential comprehension test on each condition (max. 25)

a Student's *t* test was done to compare the percentage of copying from the source texts contained in the notes).

Relationship between the use of strategies, the written products and students' inferential comprehension performance in the R&A participants

Finally, we performed bi-serial correlational analyses to explore the associations between the participants' use of strategies (and the quality of the synthesis written by those who did so) and their inferential comprehension test score. Based on the results of these correlations, we conducted regression and mediation analyses to identify possible impacts of independent variables on inferential comprehension.

Results

Comprehension results for participants in the different conditions

Descriptive statistics for the inferential comprehension test score obtained by participants depending on the tasks they had been assigned to, and the media with which they performed the task are shown in Table 4. We conducted analyses of covariance (ANCOVA), with the task (reading/synthesis vs. reading/no synthesis) and the media (paper and pencil/R&A) as the factors, prior knowledge as a covariate, and the inferential comprehension test score as the dependent variable.

The covariate, prior knowledge, was not significantly related to the comprehension test score, $F(1, 149) = 2.195$, $p = 0.141$, partial $\eta^2 = 0.015$. No statistically significant main effects were found either for the task, $F(1, 149) = 1.183$, $p = 0.279$, partial $\eta^2 = 0.008$, or for the media, $F(1, 149) = 0.452$, $p = 0.502$, partial $\eta^2 = 0.003$. Finally, no interaction was found between task and the media for the comprehension test score, $F(1, 149) = 0.470$, $p = 0.494$, partial $\eta^2 = 0.003$.

This result suggests that the media in which participants performed the task did not affect the comprehension results they obtained.

Text quality in the different conditions and its relation to comprehension

As explained in the *Method* section, participants in the reading/synthesis conditions wrote a synthesis after reading the source texts, and completed the inferential
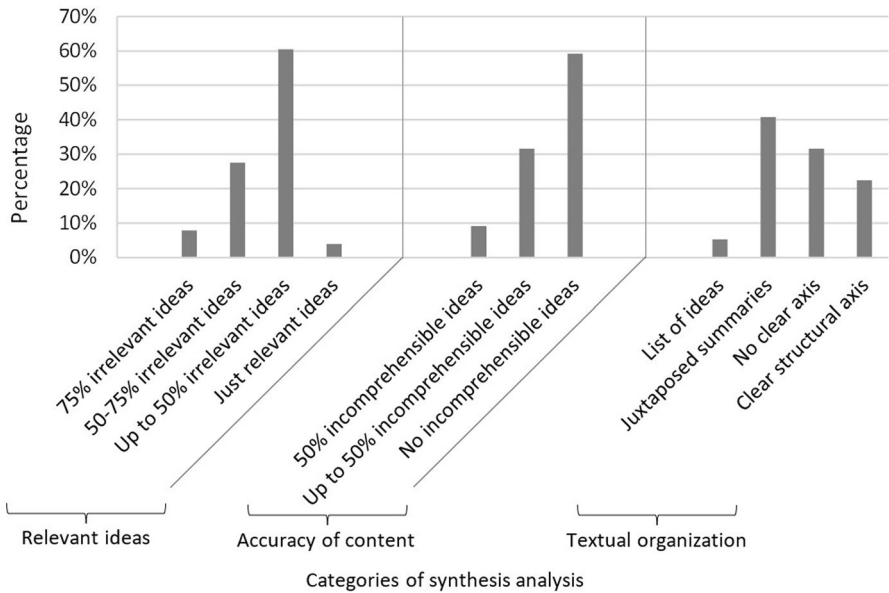
Fig. 2 Percentage of participants in each of the categories for the synthesis analysis

comprehension test once they had finished this text. Figure 2 presents the results of the analysis of the quality of their synthesis, according to the different categories established for doing so. Students tended to write short texts in terms of the number of words ($M = 285.28$; $SD = 106.81$), without mistakes or incomprehensible ideas in most cases (59.5%), but few included only relevant ideas (8.1%); forty per cent produced juxtaposed summaries, while only 23% wrote a text with a clear structural axis).

The three dimensions considered for analyzing text quality significantly correlated with inferential comprehension (relevance of ideas: $rho = 0.227$, $p = 0.049$; accuracy of content: $rho = 0.257$, $p = 0.025$; textual organization: $rho = 0.285$, $p = 0.013$).

Mann–Whitney U tests to compare participants performing this task on different media showed no significant differences between them in the different dimensions considered for analyzing text quality (Relevant ideas: $U = 595.5$, $p = 0.278$; Accuracy of content: $U = 504.5$; $p = 0.07$; Textual organization: $U = 581.5$; $p = 0.247$). Thus, considering both this result and the ones presented in the previous section, in the following sections we will focus on the subsample of participants who performed the task with the R&A software.

Strategies used to address different tasks in the R&A participants

Note-taking was used by 40.51% of participants in the R&A subsample. The notes tended to be schematic, with a mean length of 210.38 words for those who did the

Table 5 Frequencies and percentages for the presence/absence of note-taking for each task

Note-taking	Reading/no synthesis task		Reading/synthesis task	
	n	%	n	%
Presence of notes	11	34.4	21	65.5
Absence of notes	29	61.7	18	38.3

Table 6 Bi-serial correlations for participants in the reading condition

	1	2	3	4
1. Prior knowledge				
2. Inferential comprehension	0.213			
3. No. of rereadings	0.007	−0.103		
4. Note-taking	−0.075	0.036	0.164	
5. Percentage of copy in notes	−0.254	0.269	−0.309	X^1

X^1 Since the percentage of copy of the notes was assessed only for the students who took notes, the value of the note-taking variable was constant

reading/synthesis task ($SD = 134.47$), and of 107 words for those who were assigned the reading/no synthesis task ($SD = 59.16$).

A chi-square analysis was performed to determine whether there was any association between the task and the presence or absence of note-taking. The analysis revealed an association between these two variables [$\chi^2(1) = 5.688$, $p = 0.017$, *Cramer's V* $= 0.268$]: as shown in Table 5, 65.6% of participants who decided to take notes had been told they had to write a synthesis later; conversely, 61.7% of the participants who did not take notes were in the reading/no synthesis condition. No expected values were below 5.

Student's *t* test to compare participants performing the different tasks found no significant differences between them, either in the percentage of copying of source texts information in the notes, $t(30) = 0.458$, $p = 0.650$ (reading/no synthesis: $M = 23.09$, $SD = 13.51$; reading/synthesis: $M = 25.81$, $SD = 17.03$) or in the number of rereadings during the initial reading of the text, $U = 592.5$, $p = 0.830$ (reading/no synthesis: $M = 10.65$, $SD = 14.13$; reading/synthesis: $M = 10.12$, $SD = 14.06$).

Relationship between the use of strategies, the written products and students' inferential comprehension performance among R&A participants

Reading/no synthesis condition

Table 6 displays the results of bi-serial correlations between the strategies carried out by participants (number of rereadings during the initial reading of the text, and presence/absence of note-taking), the percentage of copy in the notes that participants took, and the results obtained on different tests (prior knowledge and inferential comprehension).
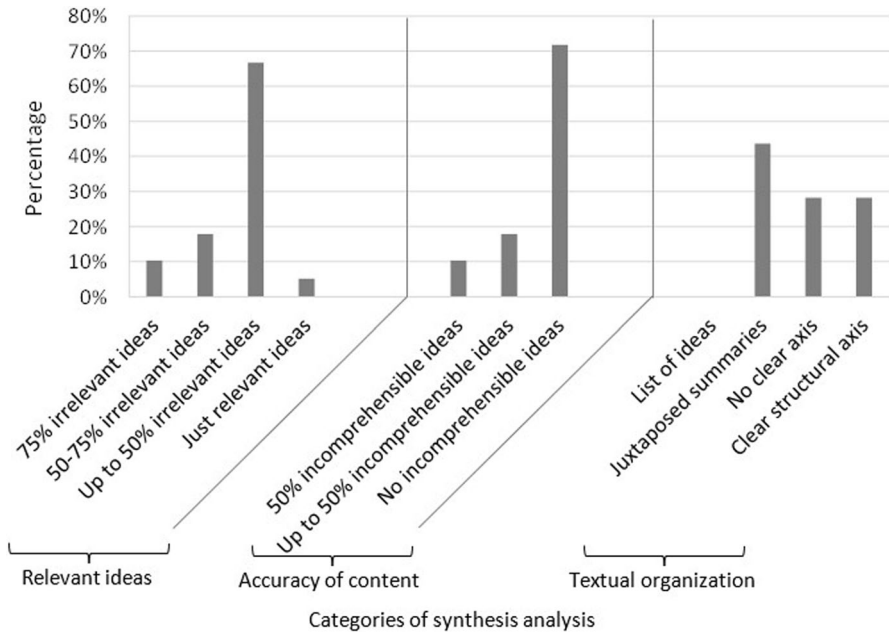
Fig. 3 Percentage of participants in each of the categories for the synthesis analysis

Table 6 shows that none of the considered variables correlated with the inferential comprehension test scores obtained by students in the reading/no synthesis condition.

Reading/synthesis condition

For participants in the reading/synthesis condition, results for the different categories used to analyze the quality of the synthesis can be observed in Fig. 3. Students tended to write short texts in terms of the number of words ($M = 285.28$; $SD = 106.81$), without mistakes or incomprehensible ideas in most cases (71.8%), but few included only relevant ideas (5.1%) and a clear structural axis (28.2%).

For this subsample of participants, bi-serial correlations were calculated between the variables referring to the strategies used by participants (i.e., number of rereadings during the initial reading, number of rereadings during writing of the synthesis, presence/absence of note-taking, percentage of copy from the source texts contained in the notes); the variables related to the written products [the three categories of synthesis analysis—relevant ideas, accuracy of content, and textual organization—the percentage of copying from the source texts ($M = 4.13$, $SD = 7.2$), and the percentage of copying from the notes ($M = 24$, $SD = 25.71$)]; and the prior knowledge and the inferential comprehension test scores (see Table 7).

Looking at the strategies used by the reading/synthesis condition participants, we observed that none of them correlates with inferential comprehension.

Table 7 Bi-serial correlations for participants in the reading/synthesis condition

	1	2	3	4	5	6	7	8	9	10
1. Prior knowledge										
2. Inferential comprehension	0.089									
3. No. of rereadings initial reading	0.108	−0.092								
4. Note-taking	−0.153	−0.179	0.177							
5. Percentage of copy in notes	−0.163	−0.006	0.305	X^1						
6. No. of rereadings during writing	−0.111	−0.101	−0.159	−0.531**	−0.122					
7. Relevant ideas	0.097	0.345*	−0.135	−283	0.117	0.174				
8. Accuracy of content	0.371*	0.347*	0.145	−0.226	0.380	−0.128	0.212			
9. Textual organization	0.354*	0.331*	−0.314	−0.480**	−0.447*	0.038	0.296	−0.061		
10. % of copy of the source texts in the synthesis	−0.126	−0.243	−0.472**	0.437**	0.824***	−0.257	0.028	0.135	−0.547***	
11. % of copy of the notes in the synthesis	−0.180	0.078	0.188	X^1	0.633**	0.351	0.406	0.311	−0.477*	0.794***

$*p<0.05$, $**p<0.01$, $***p<0.001$

X^1 Since the percentage of copy of the notes was assessed only for the students who took notes, the value of the note-taking variable was constant

The only variables that correlate with the inferential comprehension scores are the categories of quality of the synthesis. To identify whether the three categories for analyzing the quality of the text (textual organization, accuracy of content and relevance of ideas) had a similar impact on inferential comprehension, we performed a stepwise regression analysis. The analysis of standard residuals showed that the data contained no outliers (Std. Residual Min $= -2.22$, Std. Residual Max $= 1.37$). Tests to see whether the data met the assumption of collinearity indicated that multicollinearity was not a concern (Tolerance $= 0.996$, VIF $= 1.004$). The data met the assumption of independent errors (Durbin-Watson value $= 1.927$). The first model of the stepwise regression included the accuracy of content, and explained 9% of the variance ($R^2 = 0.097$, $F(1, 37) = 5.081$, $p = 0.03$), while the second model included text structure, excluded relevant ideas, and explained 20% of the variance in inferential comprehension: $R^2 = 0.203$, $F(1, 36) = 5.924$, $p = 0.02$ (accuracy of content: $\beta = 0.369$, $t(38) = 2.542$, $p = 0.015$; textual organization: $\beta = 0.353$, $t(38) = 2.434$, $p = 0.02$).

Several categories of the quality of the synthesis correlated with the strategies assessed. Thus, we decided to conduct additional analyses in order to identify other potential relations. As shown in Table 7, textual organization negatively correlated with note-taking ($p = 0.002$, meaning that note-takers wrote syntheses with a worse textual organization) and with the percentage of copy from the source text in the synthesis ($p < 0.001$, indicating that those who copied the source text most wrote syntheses with a worse textual organization). In order to look more closely at these results for textual organization, we performed a linear regression with textual organization as the dependent variable, and two independent variables: percentage of copy from the source texts in the synthesis and note-taking (a categorical variable: not taking notes was considered the reference group). We did not include other variables that correlated with the percentage of copy from the source text in the synthesis (specifically, the percentage of copy from the source texts in the notes, and the percentage of copy from the notes in the synthesis text), because the information for these two variables is only accessible for those who took notes and the variable note-taking becomes constant. An analysis of standard residuals showed that the data contained no outliers (Std. Residual Min $= -1.91$, Std. Residual Max $= 1.71$). Tests to see whether the data met the assumption of collinearity indicated that multicollinearity was not a concern (Tolerance $= 0.809$, VIF $= 1.236$). The data met the assumption of independent errors (Durbin–Watson value $= 2.145$). The regression showed that these variables explained a significant proportion of variance in textual organization, $R^2 = 0.371$, $F(2, 36) = 10.60$, $p < 0.001$ (percentage of copy from source texts: $\beta = -0.298$, $t(36) = -2.836$, $p = 0.007$; note-taking: $\beta = -0.417$, $t(36) = -2.025$, $p = 0.049$).

Another result worth noting was that rereading during the initial reading correlated with the percentage of copy from the source text. To test whether this last variable could be mediating the effect of rereading during initial reading on textual organization, we conducted a mediation analysis (Hayes, 2013). The number of rereadings during the initial reading was the independent variable, the percentage of copy from the source texts was the mediator, and textual organization was the dependent variable. No significant indirect effect of the number of rereadings during

the initial reading was found on textual organization through the percentage of copy from the source texts, $ab = -0.0153$, BCa CI $[-0.04, 0.0022]$.

Finally, in view of the results shown in Table 7, we performed a parallel multiple mediation analysis, to test whether the categories for text quality textual organization and accuracy of content could be mediating the effect of prior knowledge on inferential comprehension. The indirect effect of prior knowledge through textual organization is estimated as $ab = 0.176$, BCa CI $[0.036, 0.37]$. A second indirect effect of prior knowledge on inferential comprehension is modeled through accuracy of content, estimated as $ab = 0.192$, BCa CI $[0.054, 0.36]$. As a whole, 28.7% of the variance in inferential comprehension is explained by both proposed mediators and prior knowledge, $p = 0.0017$.

Discussion

In this study, we aimed to establish whether asking students to write about the information available in three sources influenced their level of inferential reading comprehension compared to students who only read the three documents. We also assessed the potential impact of different media to solve the tasks, and the relationship between the categories used to analyze the quality of the synthesis (i.e., relevance of ideas; accuracy of content; and textual organization) and inferential reading comprehension. Additionally, in a subsample using R&A software, we sought to identify which of the strategies (i.e., rereading and note-taking) the students used to complete each task were associated with their inferential reading comprehension results.

Focusing on the general sample, since some of our students performed the tasks using a special software program (*Read&Answer*), we compared the results obtained using this program with those obtained performing the task on paper. In line with other studies (Cerdán et al., 2009; Gil et al., 2010), and also with our expectations, we did not find significant differences between students who completed the task in the two different media. In addition, although several meta-analyses (Clinton, 2019; Delgado et al., 2018) have shown that reading on screen can decrease the level of comprehension, this effect was not visible in our study, possibly because we did not set time restrictions to accomplish the tasks. Focusing on the task variable, the students who had the opportunity to produce a written product of an integrative type did not obtain better comprehension results than those whose task consisted exclusively of reading. This unexpected result is at odds with the findings of other studies, such as Graham and Hebert (2011), and Hebert et al. (2013). However, it is important to note that those two meta-analyses compared writing interventions and their impact on comprehension, while in our study we did not instruct our students how to write the synthesis. As Graham and Hebert (2011) mentioned, it is possible that the impact of writing on comprehension is less visible among higher education students carrying out more complex tasks, such as synthesizing from multiple documents.

A possible explanation for this result, which contradicts our initial hypothesis, may be found not so much in the type of task the students were assigned, but in the strategies they chose in order to carry out the tasks, and the quality of the written

product in the reading/synthesis condition. Starting with the last variable, the categories considered to assess the quality of the written product significantly correlated with the level of inferential reading comprehension, meaning that students who wrote a better synthesis, in terms of textual organization, accuracy of content and relevance of the ideas included, also obtained better results in the comprehension test whether they performed the task on paper or with R&A, a result that has also been observed in prior research (Martínez et al., 2015). In fact, the accuracy of content and the textual organization explained 20% of the variance of inferential comprehension for the students who wrote the synthesis. This result points to the importance of correctly selecting and organizing the information as a means to deepen and understand the sources, in line with other research findings (Solé et al., 2013).

Regarding the strategies deployed by participants in the reading/no synthesis versus the reading/synthesis condition, the results show that in both situations, independently of the task assigned and contrary to our expectations, students mostly reread the text several times. Nevertheless, and in line with our expectations, the students in the reading/synthesis condition tended to take notes more than their peers in the reading/no synthesis condition. Since these students were required to write a text that involved integrating information from several sources, it is logical that they used note-taking to a higher extent, as we expected.

Although neither of these strategies (rereading and note-taking) was related to inferential comprehension for the students in both conditions, for the reading/synthesis condition, note-taking and the percentage of copy from the source texts in the synthesis explained 37% of the variance of the textual organization, meaning that those students who took notes mostly copied from the source texts and achieved low textual organization scores. Conversely, students who did not take notes and did not copy from the source texts were able to provide a better structure to their synthesis. In agreement with prior research (Gil et al., 2008; Hagen et al., 2014), students' notes in this condition were mostly copied fragments from the source texts, and note-takers apparently trusted these notes to write their synthesis instead of going back to the source texts, which hypothetically may have led them to write lower-quality syntheses and to obtain worse inferential comprehension results than students who did not take notes. Following on from Kobayashi (2009) and Luo and Kiewra (2019), it may be the case that, when students have not been taught how to use this tool strategically, they simply use it routinely, focusing mainly on superficial information from the texts and copying information instead of elaborating upon it. In contrast to other studies of text comprehension (Gil et al., 2010; Kintsch, 1998; Le Bigot & Rouet, 2007), we did not find a relationship between prior knowledge and the inferential comprehension results in the correlation analyses. Although it might be thought that the questionnaire lacked the sensitivity required to assess prior knowledge, one of the results seems to point to a more complex explanation. In fact, for the students in the reading/synthesis condition, prior knowledge correlated with two of the categories of the quality of the synthesis (text structure and accuracy of the content), meaning that the students who had a higher degree of prior knowledge were also able to better organize and include accurate information in their syntheses. For this reason, we decided to perform mediation analyses of prior knowledge on inferential comprehension scores, through the two previous synthesis categories.

The results showed significant indirect effects of prior knowledge on inferential comprehension through text organization and accuracy of content. In sum, although background knowledge did not play a direct role in comprehending the texts in depth, it was relevant to helping students to produce better syntheses.

A final comment should be made regarding the category of the synthesis quality "relevant ideas". Although this category correlated significantly with the inferential comprehension, it did not correlate with prior knowledge and did not show a clear impact on inferential comprehension when introduced in the regression analysis. This result may be related to the fact that identifying the three main ideas shared by the texts did not prevent students from just copying them in their text without expanding their understanding of the sources' content.

Several limitations of the present study must be pointed out. The first is to do with the inferential comprehension test. Although it might be argued that using other comprehension scales (e.g., literal comprehension), might have produced additional results, higher education students are expected to achieve deep inferential comprehension when studying from several texts. Therefore, it seems important to focus on this level of comprehension, in which most studies on understanding multiple sources have found significant differences.

A second possible limitation is the "artificial" nature of the software program (*Read&Answer*). Although we did not find differences between the paper condition and the R&A condition in the inferential comprehension results, we cannot guarantee that the use of this software did not have an impact on the strategies students performed when solving the tasks. Although studies that have used this tool suggest that it does not have a relevant impact on participants' cognitive processes (see Cerdán et al., 2009; Gil et al., 2010), and it has been validated against eye-tracking and paper and pencil testing (Vidal-Abarca et al., 2011), its effect may depend on the complexity of the task. A third important limitation is the possibility that the reading/synthesis condition may have appeared artificial to the students. It is conceivable that at least some of the students would not have spontaneously written a synthesis as a way of improving their comprehension of the texts. As this was an assigned task, neither decided by them, nor assessed by their professors, nor linked to a subject, its potential may be diminished because it was produced routinely and not conducted strategically, as shown by the students' reliance on the notes which were mostly copied from the texts.

Despite its limitations, however, our study contributes to the understanding of the relationship of synthesis writing (and the strategies undertaken to perform it) on inferential comprehension. Our results suggest the importance not only of using specific strategies to carry out reading and/or writing tasks (e.g., rereading during writing, or note-taking) but of using them strategically to fulfil the aims and to adapt to the demands of the task. Thus, for example, students tended to take notes to a higher extent when they were required to write a synthesis, a practice that seems eminently suitable in view of the complexity of the task; however, as we argued above, the key issue is how these notes are used. On the other hand, the quality of the written products seems to be crucial to improving reading comprehension. Thus, it is not the task itself, but the way students meet its requirements, that determines the quality of the result. Our results for synthesis quality confirm that students' ability to integrate

information from different sources cannot be taken for granted, and emphasize once again the need for specific writing instruction in higher education.

Appendix A

Segments of the source texts in which the intertextual idea of the "contrast between intelligence as a unitary quality measured by testing, and intelligence as a multiple, diverse competence" can be found

Text 1: multiple intelligences

"The idea that people are, globally, more or less intelligent than others, and that this general intelligence is what conditions or determines learning, is an essential part of the "common sense" conception of intelligence and its relationship with school learning. The strength of this idea has led to—and has been, at the same time, reinforced by—a simplistic interpretation of the information provided by intelligence tests, and in particular by the so-called "Intelligence Quotient" (IQ) that many of these tests make it possible to calculate.

From this interpretation, IQ has ceased to be understood as an indicator of intellectual capacity and is considered as the very substance of that ability: according to this, you are intelligent if you have a high score on intelligence tests, and you have a high score on tests because you are intelligent. With this, what at first was nothing but a more or less convenient artifact to simplify the measure of intelligence becomes the essence of intelligence itself, and what had been measured from a set of different aspects or factors ends up being conceived as a single, unitary and uniform entity.

As compelling as this unitary and global view of intelligence may seem, there are abundant scientific arguments against this idea. The most forceful come from various authors and recent theories on intellectual behavior, which have as a common element the questioning of traditional intelligence tests as an appropriate method of measuring and studying it. Despite the differences among them, these authors agree that traditional intelligence tests only measure a small part of people's intellectual abilities: a set of logical-mathematical capacities, associated with scientific thinking and the more academic school content. These authors and theories propose the existence of a much broader spectrum of intellectual capacities, which would constitute, in their own right, "intelligences", as important and worthy of attention as the academic intelligence to which all intelligence tests refer."

Text 2: two essential claims about intelligence

"The theory of multiple intelligences makes two complementary claims. First, the theory is a complete explanation of human cognition: it presents intelligences as a new definition of the nature of human beings from a cognitive point of view. Whereas Socrates saw man as a rational animal and Freud emphasised

the irrationality of the human being, I have (with due caution) described the human being as an organism possessing a basic set of seven, eight or a dozen intelligences.

(…) The second assertion—that we all have a unique combination of intelligences—leads to the most important consequence of this theory for the next millennium. We can choose to ignore this uniqueness, we can choose to minimise it, or we can choose to enjoy it. I believe that the great challenge of deploying human resources is to find the best way to take advantage of the uniqueness we have been given as a species: that of having multiple intelligences."

Text 3: successful intelligence

"Successful intelligence is, according to Sternberg (1997), that which is really important in life, that which is used to achieve important goals and that shown by those who have succeeded in life, either according to their personal patterns, or according to those of others. This intelligence has little to do with the intelligence measured by traditional tests and IQ scores. According to Sternberg, these tests refer only to a small and not very important (though academically overrated) part of a much broader and more complex intellectual spectrum, and essentially measure "inert intelligence", that is, potentialities that do not necessarily lead to a movement or a directed action, which one does not have to know how to use in order to produce real changes in life, for oneself or for others. According to Sternberg, the notion that there is a general intelligence factor that can be measured with IQ is false, and is based on the fact that all traditional intelligence tests measure essentially the same narrow range of skills.

For Sternberg, successful intelligence involves three aspects: an analytical aspect, a creative aspect and a practical aspect. The first is used to solve problems; the second, to decide which problems to solve; and the third, to put the solutions into practice. Conventional intelligence tests measure only the analytical aspect of intelligence, and not even completely. These three aspects are considered relatively independent of each other, and in fact each of them is conceptualized as a specific intelligence. In doing so, Sternberg points to the multiple, non-unitary character of intelligence."

Appendix B

Sample of prior knowledge test questions (true/false)

1. Intelligence is a general ability that each person has to a certain degree. T/F

2. What a person learns modifies their intelligence. T/F
3. Success in life is a more reliable indicator of a person's intelligence than his/her own IQ. T/F

Sample of inferential reading comprehension questions (true/false)

1. From the pluralist theories of intelligence, the balance and relationship between the different types of intelligence of a person is more important than the fact that he/she may be more or less gifted in any one of them. T/F
2. Despite defending the plural nature of people's intellectual abilities, the theories of Gardner and Sternberg accept that IQ reflects the overall degree of these abilities. T/F

Appendix C

Reliability calculation process

According to Lord and Novick (2008), if the true-score variance of the items vary, Cronbach's alpha is an underestimate of the true reliability to and unknown height. Therefore it is preferable to use and estimate of the reliability that is not affected by differences in true-score variance between items (H. Van den Bergh, personal communication, May 19, 2022). For this reason, one-factor CFA models were conducted with the Prior knowledge questionnaire (see Fig. 4), and with the Inferential comprehension test (see Fig. 5), and, since the models fitted the data (Prior knowledge questionnaire: RMSEA (95% CI) = 0.063 (0.044; 0.082); $X^2 = 145.385$; df = 90; $p < 0.001$; $x^2/df = 1.615$; Inferential comprehension test: RMSEA (95% CI) = 0.04471 (0.03041; 0.05715); $X^2 = 358.539$; df = 275; $p < 0.001$; $x^2/df = 1.304$), the regression of the items on the factor was interpreted as the proportion of true-score variance. Taking the average proportion of true-score variance, the Spearman–Brown extension for lengthening a test was calculated.
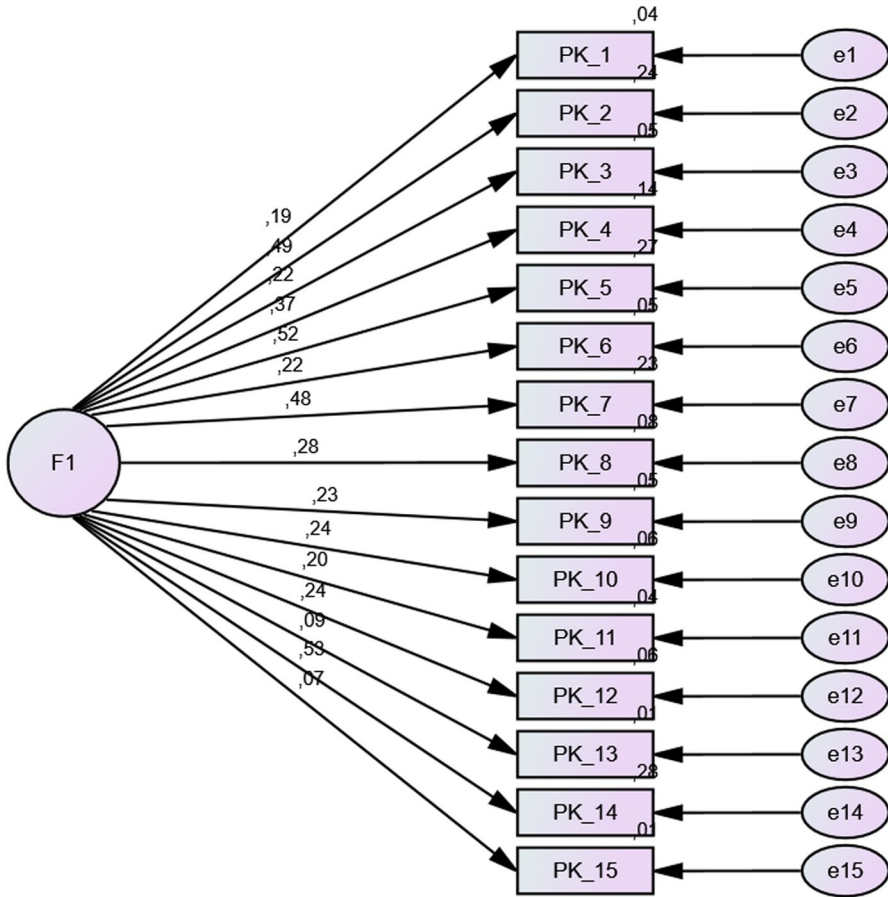
Fig. 4 Inferential question items and their loadings (regression values of the items on the factor are presented in the middle of the arrows)
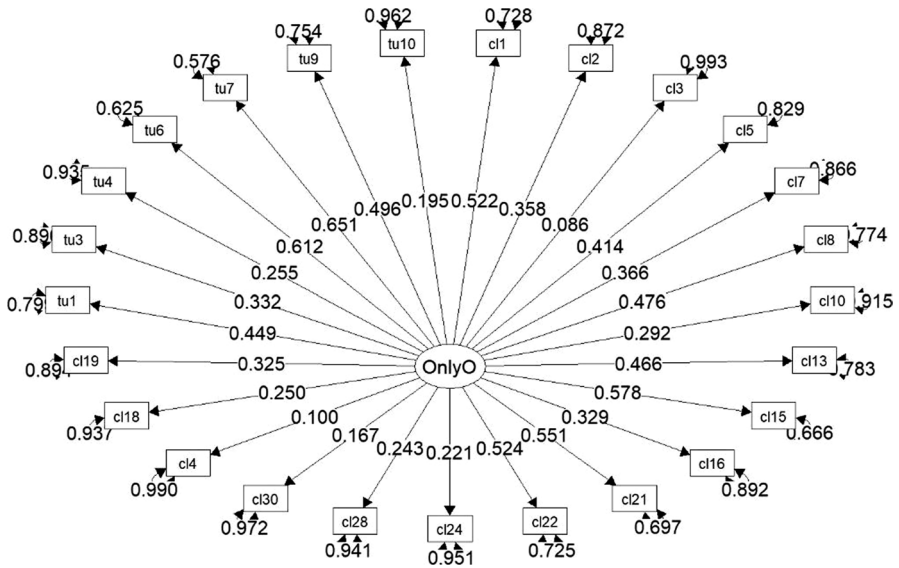
Fig. 5 Inferential question items and their loadings (regression values of the items on the factor are presented in the middle of the arrows)

Data availability Availability of data and material under request.

Declarations

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

Ethical approval This manuscript has not been published elsewhere and is not under consideration by another journal.

References

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*(5), 364–373. https://doi.org/10.1598/RT.61.5.1

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research, 54*(4), 577–596. https://doi.org/10.3102/00346543054004577

Baker, W., & Boonkit, K. (2004). Learning strategies in reading and writing: EAP contexts. *RELC Journal, 35*(3), 299–328. https://doi.org/10.1177/0033688205052143

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*(1), 29–58. https://doi.org/10.3102/00346543074001029

Barrio-Cantalejo, I. M., Simón-Lorda, P., Melguizo, M., Escalona, I., Mirajúan, M. I., & Hernando, P. (2008). Validation of the INFLESZ scale to evaluate readability of texts aimed at the patient. *Anales Del Sistema Sanitario De Navarra, 31*(2), 135–152. https://doi.org/10.4321/s1137-66272008000300004

Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting integration of multiple texts: A review of instructional approaches and practices. *Educational Psychology Review, 30*(3), 1–27. https://doi.org/10.1007/s10648-018-9436-8

Basaraba, D., Yovanoff, P., Alonzo, J., & Tindal, G. (2013). Examining the structure of reading comprehension: Do literal, inferential, and evaluative comprehension truly exist? *Reading and Writing, 26*, 349–379. https://doi.org/10.1007/s11145-012-9372-9

Bednall, T. C., & Kehoe, E. J. (2011). Effects of self-regulatory instructional aids on self-directed study. *Instructional Science, 39*(2), 205–226. https://doi.org/10.1007/s11251-009-9125-6

Boscolo, P., Arfé, B., & Quarisa, M. (2007). Improving the quality of students' academic writing: An intervention study. *Studies in Higher Education, 32*(4), 419–438. https://doi.org/10.1080/03075070701476092

Bråten, I., & Strømsø, H. I. (2009). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes, 47*(1), 1–31. https://doi.org/10.1080/01638530902959646

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). Erlbaum.

Britt, M. A., Rouet, J. F., & Durik, A. M. (2017). *Literacy beyond text comprehension: A theory of purposeful reading*. Taylor & Francis.

Castells, N., Minguela, M., Solé, M., Miras, M., Nadal, E., & Rijlaarsdam, G. (2021). Improving questioning-answering strategies in learning from multiple complementary texts: An intervention study. *Reading Research Quarterly, 57*(3), 879–912. https://doi.org/10.1002/rrq.451

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., Salmerón, L., Martínez, T., & Gilabert, R. (2009). Read&Answer: A tool to capture on-line processing of electronic texts. *The Ergonomics Open Journal, 2*, 133–140.

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading, 42*(2), 288–325. https://doi.org/10.1111/1467-9817.12269

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review, 25*, 23–38. https://doi.org/10.1016/j.edurev.2018.09.003

Dinsmore, D. L. (2018). *Strategic processing in education*. Routledge.

Dovey, T. (2010). Facilitating writing from sources: A focus on both process and product. *Journal of English for Academic Purposes, 9*(1), 45–60. https://doi.org/10.1016/j.jeap.2009.11.005

Fitzgerald, J., & Shanahan, T. (2020). Reading and writing relationships and their development. *Educational Psychologist, 35*(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5

Galbraith, D., & Baaijen, V. M. (2018). The work of writing: Raiding the inarticulate. *Educational Psychologist, 53*(4), 238–257. https://doi.org/10.1080/00461520.2018.1505515

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology, 35*(3), 157–173. https://doi.org/10.1016/j.cedpsych.2009.11.002

Gil, L., Vidal-Abarca, E., & Martínez, T. (2008). Efficacy of note-taking to integrate information from multiple documents. *Infancia y Aprendizaje, 31*(2), 259–272. https://doi.org/10.1174/021037008784132905

Goldman, S. R., & Saul, E. U. (1990). Flexibility in text processing: A strategy competition model. *Learning and Individual Differences, 2*(2), 181–219. https://doi.org/10.1016/1041-6080(90)90022-9

Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81*(4), 710–744. https://doi.org/10.17763/haer.81.4.t2k0m13756113566

Graesser, A., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M. G. McKeown & L. Kucan (Eds.), *Bringing research to life* (pp. 112–141). Guilford Press.

Gronlund, N. E. (1985). *Measurement and evaluation in teaching*. MacMillan.

Hagen, Å. M., Braasch, J. L., & Bråten, I. (2014). Relationships between spontaneous note-taking, self-reported strategies and comprehension when reading multiple texts in different task conditions. *Journal of Research in Reading, 37*(1), 141–157. https://doi.org/10.1111/j.1467-9817.2012.01536.x

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression based approach*. Guilford Publications.

Hebert, M., Simpson, A., & Graham, S. (2013). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing, 26*(1), 111–138. https://doi.org/10.1007/s11145-012-9386-3

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

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. Harcourt College Publishers.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kirby, J. R. (1988). Style, strategy, and skill in reading. In R. R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 229–274). Springer.

Kobayashi, K. (2009). Comprehension of relations among controversial texts: Effects of external strategy use. *Instructional Science, 37*(4), 311–324. https://doi.org/10.1007/s11251-007-9041-6

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

Lenski, S. D., & Johns, J. L. (1997). Patterns of reading-to-write. *Reading Research and Instruction, 37*(1), 15–38. https://doi.org/10.1080/19388079709558252

List, A., & Alexander, P. A. (2019). Toward an integrated framework of multiple text use. *Educational Psychologist, 54*, 20–39. https://doi.org/10.1080/00461520.2018.1505514

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.

Luo, L., & Kiewra, K. A. (2019). Soaring to successful synthesis writing: An investigation of SOAR strategies for college students writing from multiple sources. *Journal of Writing Research, 11*(1), 163–209. https://doi.org/10.17239/jowr-2019.11.01.06

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology, 91*(4), 615–629. https://doi.org/10.1037/0022-0663.91.4.615

Martínez, I., Mateos, M., Martín, E., & Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: Effects on an instructional programme on learning, reading and writing processes, and text quality. *Journal of Writing Research, 7*, 275–302. https://doi.org/10.17239/jowr-2015.07.02.03

Mateos, M., & Solé, I. (2009). Synthesising information from various texts: A study of procedures and products at different educational levels. *European Journal of Psychology of Education, 24*(4), 435–451. https://doi.org/10.1007/BF03178760

McGinley, W. (1992). The role of reading and writing while composing from multiple sources. *Reading Research Quarterly, 27*(3), 227–248. https://doi.org/10.2307/747793

Minguela, M., Solé, I., & Pieschl, S. (2015). Flexible self-regulated reading as a cue for deep comprehension: Evidence from online and offline measures. *Reading & Writing, 28*(5), 721–744. https://doi.org/10.1007/s11145-015-9547-2

Miras, M., Solé, I., & Castells, N. (2013). Creencias sobre lectura y escritura, producción de síntesis escritas y resultados de aprendizaje [Reading and writing beliefs, written synthesis production and learning results]. *Revista Mexicana De Investigación Educativa, 18*(57), 437–459.

Moran, R., & Billen, M. (2014). The reading and writing connection: Merging two reciprocal content areas. *Georgia Educational Researcher*. https://doi.org/10.20429/ger.2014.110108

Nadal, E., Miras, M., Castells, N., & de la Paz, S. (2021). Intervención en escritura de síntesis a partir de fuentes: Impacto de la comprensión [Intervention in writing a synthesis based on sources: Impact of comprehension]. *Revista Mexicana De Investigación Educativa, 26*(88), 95–122.

Nelson, N. (2008). The reading-writing nexus in discourse research. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 435–450). Erlbaum.

Nelson, N., & King, J. R. (2022). Discourse synthesis: Textual transformations in writing from sources. *Reading and Writing*. https://doi.org/10.1007/s11145-021-10243-5

Reynolds, G., & Perin, D. (2009). A comparison of text structure and self-regulated writing strategies for composing from sources by middle school students. *Reading Psychology, 30*(3), 265–300. https://doi.org/10.1080/02702710802411547

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Cambridge monographs and texts in applied psycholinguistics. Advances in applied psycholinguistics: Reading, writing and language learning* (pp. 142–175). Cambridge University Press.

Schumacher, G. M., & Nash, J. G. (1991). Conceptualizing and measuring knowledge change due to writing. *Research in the Teaching of English, 25*(1), 67–96.

Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating information: An analysis of the processes involved and the products generated in a written synthesis task. *Written Communication, 30*(1), 63–90. https://doi.org/10.1177/0741088312466532

Spivey, N. (1997). *The constructivist metaphor: Reading, writing and the making of meaning*. Academic Press.

Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly, 24*(1), 7–26.

Sternberg, R. J. (1997). *Successful intelligence: How practical and creative intelligence determine success in life*. Plume.

Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences inrereading. *Memory and Cognition, 32*(5), 696–710. https://doi.org/10.3758/BF03195860

Tierney, R. J., & Shanahan, T. (1996). Research on the reading-writing relationship: Interactions, transactions, and outcomes. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (pp. 246–280). Erlbaum.

Vandermeulen, N., Van den Broek, B., Van Steendam, E., & Rijlaarsdam, G. (2019). In search of an effective source use pattern for writing argumentative and informative synthesis texts. *Reading and Writing, 33*(2), 239–266. https://doi.org/10.1007/s11145-019-09958-3

Van Weijen, D., Rijlaarsdam, G., & Van Den Bergh, H. (2018). Source use and argumentation behavior in L1 and L2 writing: A within-writer comparison. *Reading and Writing, 32*(6), 1635–1655. https://doi.org/10.1007/s11145-018-9842-9

Vidal-Abarca, E., Martinez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., Mañá, A., Llorens, A. C., & Ferris, R. (2011). Recording online processes in task-oriented reading with Read&Answer. *Behavior Research Methods, 43*(1), 179–192. https://doi.org/10.3758/s13428-010-0032-1

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 301–311. https://doi.org/10.1037//0022-0663.91.2.301

Zhang, C. (2013). Effect of instruction on ESL students' synthesis writing. *Journal of Second Language Writing, 22*(1), 51–67. https://doi.org/10.1016/j.jslw.2012.12.001