Improving Questioning–Answering Strategies in Learning from Multiple Complementary Texts: An Intervention Study

Núria Castells
Marta Minguela
Isabel Solé
Mariana Miras

Faculty of Psychology, Universitat de Barcelona (UB), Passeig de la Vall d'Hebron 171, Barcelona, 08035, Spain

Esther Nadal

Psychology and Education Science Studies, Open University of Catalonia, Rambla del Poblenou 156, Barcelona, 08018, Spain

Gert Rijlaarsdam

School of Child Development and Education, University of Amsterdam, Postbus 15776, Amsterdam, 1001 NG, The Netherlands

ABSTRACT

Studies have shown that inferential questions encourage a more in-depth understanding of texts and that students need to learn appropriate strategies for answering them, particularly when they deal with multiple texts. In this experimental study, the authors aimed to improve eighth-grade students' (13- to 14-years old) ability to answer intra- and intertextual inferential questions when they read one or multiple complementary texts. The intervention was implemented by a group of middle-school history teachers. Teachers in both the intervention and control groups (IG and CG, respectively) taught the same teaching unit using the same reading materials. However, teachers in the IG participated in 12 hours of professional development seminars on analysis of their classroom practice and how to improve their questioning strategies. Post-intervention results revealed that students in the IG were significantly better than those in the CG at answering intra- and intertextual inferential questions. This difference was maintained at follow-up (2 months after finishing the intervention). Students in the IG also performed better than those in the CG at a learning test. These results confirm the value of teaching students how to answer complex questions, especially when they refer to more than one text. The findings also support the value of the professional development program that enables teachers to reflect on their practice.

The need for readers to make inferences from multiple texts is more urgent than ever. In our internet and remote learning era, much of the students' learning still involves written texts (Olson, 2015). Students must develop strategies to understand information in a way that they can retrieve and transform it into applicable knowledge, making connections across texts and inferring new meaning from these texts.

The process of understanding a text involves identifying and establishing relationships between its components (e.g., words, sentences, paragraphs, and sections) and constructing what Kintsch (1988, 1998) refers to as a situation model. From this perspective, once a text reaches a certain level of complexity, understanding it becomes a question of degree. Depending on cognitive and linguistic factors, readers may understand the same text differently, and the same reader may achieve different levels of understanding of the same text. In other words, a reader's representation of the text's content may show different degrees of scope, integration, and coherence.

One common approach to assessing comprehension involves questions posed by a teacher after a reading task (Alvermann, Swafford, & Montero, 2004; Castells, 2019). These questions can take various forms

and address different levels of comprehension. Some questions may call for identification and literal (or nearly literal) retrieval of information from the text, requiring relatively simple or superficial understanding. In contrast, other questions may require an integrated and coherent representation of the text's content, which presumes a more in-depth understanding (Elleman, 2017).

In this study, we aimed at improving middle-school students' question-answering strategies about one or multiple complementary texts. Since individuals can access multiple documents to achieve their aims (Goldman & Scardamalia, 2013), students should be provided with strategies that help them construct relations between information in more than one text. Some intervention studies have centered on reading and question answering (Elleman, 2017). Our intervention focused on providing students with appropriate strategies for answering different kinds of questions, that is, it aimed to provide students with appropriate intentional goal-oriented procedures to construct the meaning of the texts (Afflerbach, Hurt, & Cho, 2020; Dinsmore, 2018). In particular, we wanted them to develop reading strategies to process, understand, and integrate information from complementary documents, where the information provided by one of the texts adds to the information from the other (Britt, Rouet, & Braasch, 2013; Firetto & Van Meter, 2018; List & Alexander, 2019; Perfetti, Rouet, & Britt, 1999; Rouet & Britt, 2011). According to Firetto's review on learning from complementary perspectives (2020), there has not been much research on multiple complementary texts because most research into intervention on multiple-text reading has focused on discrepant or conflicting-view texts (see also Barzilai, Zohar, & Mor-Hagani, 2018).

Three additional features characterize the intervention. First, it was implemented by history teachers, rather than the researchers. Through a professional development program that we designed, teachers acquired resources and instruments that helped them make use of different kinds of questions and taught their students question-answering strategies. Second, we assessed the level of learning achieved by the students after the completion of the teaching unit. Third, a follow-up test was used to gauge the long-term effects of the intervention. Below we outline the theoretical underpinnings of this study, and the studies that support those theories.

Research on Questioning

Studies on the teaching of reading have indicated that teachers at all educational levels use questioning as a regular part of their daily practice to verify understanding, recapitulate text content, or focus students' attention on particular issues (Armbruster, et al., 1991; Elleman, 2017; Hodges, 1980; Magnusson, Roe, & Blikstad-Balas, 2019;

Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009; Ness, 2011). Questions are also commonly asked before a reading task so that students can begin to recall what they already know about the topic, make predictions, and anticipate issues to be addressed (Castells, 2019; Magnusson et al., 2019).

The question-answer pattern may be used to promote comprehension of one or several texts and to foster students' understanding of the notions and concepts that underpin a given topic or subject area. Thus, a questioning strategy can be thought of as having two dimensions: one geared more toward assessment, with students' answers being taken as an indicator of what they have learned, and another aimed at guiding their understanding and learning. In this respect, what one finds in the texts is, largely, a result of what is being sought. As Beck, McKeown, and Gromoll (1989) noted, teachers' questioning strategies also serve as models that help students learn to ask appropriate questions themselves as they approach a text.

The types of questions that teachers may ask have been considered from a range of theoretical perspectives (Anderson & Krathwohl, 2001; Goldman & Durán, 1988; Graesser, Ozuru, & Sullins, 2010; McMaster et al., 2012; Mosenthal, 1996; Pearson & Johnson, 1978; Raphael & Au, 2005; Rouet, 2006; Sánchez, 2010). However, most researchers have identified two main types of questions: *literal* questions that primarily require students to *identify explicit data or information* in the fragments of a text and *inferential* questions that require students to *integrate the information and produce an interpretation* of a text. Some authors (Basaraba, Yovanoff, Alonzo, & Tindal, 2013; Cho, Afflerbach, & Han, 2018; Eason, Goldberg, Young, Geist, & Cutting, 2012; Graesser et al., 2010; OECD, 2019; Sánchez, 2010) have also included a third category, namely, questions that lead students to evaluate information and/or reflect critically on a text. This last type of questions will not be the subject of attention in this study.

The aforementioned question distinctions are consistent with the principle proposed by Kintsch (1988, 1998) that different levels of comprehension can be attained about a text and that these levels pose challenges of increasing cognitive complexity. Literal questions call for identifying, locating, and extracting explicit information. As these tasks involve recall and recognition (Basaraba et al., 2013; Eason et al., 2012), they generally imply minimum processing effort, although the degree of processing depends on variables such as the amount of information referred to in the question, the length and complexity of the source text, and the amount and relevance of similar information that it contains.

In contrast, answering inferential questions requires readers to go beyond the explicit information in the text and anticipate or infer (on the basis of their existing knowledge) information and meanings that are not explicit but can be deduced and are consistent with the text's

content. To answer such questions, readers must develop a situation model in which they select and link information that is spread throughout the text (i.e., text-based or text-connecting inferences) and/or combine existing knowledge with the information contained in the text (i.e., knowledge-based or gap-filling inferences) (Basaraba et al., 2013; Eason et al., 2012). Inference is essential to establish the local or overall coherence of a text. To perceive overall coherence, the reader must be able to connect several units of information (Basaraba et al., 2013). In intertextual inferences, students must integrate content from several texts. To achieve this, they must search for, select/evaluate, connect, and integrate diverse information (Barzilai et al., 2018; Rouet & Potocki, 2018).

Levels of Cognitive Complexity in Questions and Subsequent Learning

It is widely recognized that literal and inferential questions require the reader to engage at different levels of cognitive complexity. After reviewing various studies, Eason et al. (2012) concluded that literal questions are easier to answer than inferential ones. They carried out hierarchical regression analyses to determine the cognitive skills associated with children's performance on various kinds of questions. The ability to make inferences was found to be fundamental for answering all but literal questions, which only required basic linguistic skills. Spencer et al. (2019) found similar results when they included text complexity as a variable. Similarly, Muijselaar et al. (2017) found that "literal questions require a lower level of reading comprehension ability than inferential questions" (p. 80). Basaraba et al. (2013) reported a similar finding overall, but noted that variance also occurs within each type so that a certain literal question might be more complex than a particular inferential question. Thus, while they acknowledged the existence of different levels of text comprehension, they rejected the idea that the relationship between different levels is purely hierarchical.

The general recognition that students find literal questions easier to answer than inferential questions is reflected in the frequent use of the terms *low-level* and *high-level* for the two types of questions (Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gil, 2009; Rouet, 2006; Vidal-Abarca, Mengual, Sanjose, & Rouet, 1996). Low-level questions require students to engage only in superficial processing (e.g., identifying specific information in texts), while high-level questions require learners to engage in higher level processing, such as making inferences or drawing conclusions (List, Grossnickle, & Alexander, 2016). Furthermore, students tend to perform different actions when they are asked questions of different types. In Cerdán et al.'s (2009)

study, when students were confronted with low-level questions, they scanned several paragraphs to find the response. In contrast, students who were asked a high-level question paused longer on more paragraphs. In another study, Cerdán, Gilabert, and Vidal-Abarca (2011) found that poorer comprehenders tended to use a superficial strategy based on identifying the words overlapping between the question and the content to be extracted from the text. Although this strategy can be useful for answering some literal/low-level questions, it is likely to fail for most inferential/higher-level questions that require a deeper semantic strategy or approach (Cerdán et al., 2011). These actions and the processes involved in solving high-level questions may explain why making inferences might potentially increase content learning. Past research has also confirmed that inferential questions have greater potential to promote in-depth understanding (Graesser & Lehman, 2011; Halpain, Glover, & Harvey, 1985; Rosenshine, Meister, & Chapman, 1996; Taboada & Guthrie, 2006) and subsequent learning, as they promote integration of information through the construction of a situation model (Kintsch, 1988, 1998).

Answering inferential questions correctly may sometimes depend more on the level of prior knowledge and/or the working memory capacity (Britt, Rouet, & Durik, 2017; Yuill, Oakhill, & Parkin, 1989) when compared to answering literal questions. Thus, students who have a higher level of prior knowledge and working memory capacity are more likely to answer inferential questions better. From an educational point of view, it therefore seems necessary to teach students how to answer this type of question and provide them with appropriate strategies. Nevertheless, analysis of questions formulated by teachers or contained in textbooks has shown that most of them are literal questions (Armbruster et al., 1991; Daines, 1986; Occelli & Valeiras, 2013; O'Flahavan, Hartman, & Pearson, 1989; Sáiz, 2011). Such questions may be answered without achieving an overall understanding of a text or even understanding the idea addressed in the question. As such, they have limited value in building knowledge or learning.

In these circumstances, students' learning from texts is at risk. First, they may only identify and reproduce information without achieving the overall understanding that is promoted when they are faced with inferential questions, for which they must link various pieces of information (through comparison or by identifying causal relationships) and make use of their existing knowledge. Moreover, they may come to believe that the sole purpose of asking questions about texts is to facilitate the reproduction of isolated items of information, whereas its true goal should be to help them modify their knowledge through in-depth processing of the information. The ability to make inferences and integrate information is a necessary skill for the successful completion of numerous tasks that involve reading to learn, for example, producing a

summary, writing an essay after reading one or more source texts that contain complementary or contradictory information, or choosing adequate source materials for certain purposes. Therefore, it is clearly important that students learn how to tackle questions that require them to make inferences in a conscious, structured way and to integrate information from different passages of the same text or from multiple texts. However, research conducted in natural settings has shown very few instances in which teachers explicitly teach their students to make inferences (Magnusson et al., 2019).

Multiple Complementary Text Comprehension

When a person is reading two or more texts, the task of connecting and relating information becomes even more complex (Miras, 2019; Britt & Rouet, 2012; Britt & Sommer, 2004; List & Alexander, 2019; Rouet & Britt, 2014; Rouet & Potocki, 2018). Most research on multiple-text reading has focused on conflicting-view documents (texts with inconsistent or contradictory information). However, in practice at the secondary educational level, learners are also assigned tasks that require integration of complementary information across texts (Firetto, 2020; Firetto & Van Meter, 2018). For instance, in middle school, students can be required to search for additional information to complete an assignment (e.g., writing an informative text or synthesis in history; Solé, Miras, Castells, Espino, & Minguela, 2013). Tasks that require reading of multiple documents to integrate the complementary information in them can be found in different educational levels and content areas, such as history (Solé, et al., 2013), social sciences (Nadal, Miras, Castells, & de la Paz, 2021), or sciences (Firetto & Van Meter, 2018). In these and other studies, different terms have been used to indicate or characterize the type of complementarity that can be established between the texts. List, Du, and Lee (2021), for instance, used texts with consistent information that were overlapping (information dealing with the same theme, but explained with different words in the texts) or distinct (information from one text adds to the information from the other text). Firetto and Van Meter (2018), on the other hand, conceptualized the connection between the information in their texts as componential because each text contributed unique information that fitted with the information from the other texts in the set (e.g., understanding the physiological systems in the human body). Nadal et al. (2021) presented the students with several pairs of texts that shared redundant information (almost the same information, but paraphrased differently between the texts) and complementary information. In other instances, the relations between the texts have been characterized as semantically congruent (e.g., Braasch, McCabe, & Daniel, 2016). Briefly, the type of connections between the information included in multiple complementary or congruent texts can range from redundant/overlapping (the same information that is just paraphrased or explained in a different way from one text to the other) to distinct/componential or complementary (when the information from one text adds to the information included in the other text).

Several models may explain the processes that learners carry out when they deal with multiple sources and the factors involved in such an endeavor (e.g., Multiple-Document Task-Based Relevance Assessment and Content Extraction [MD-TRACE], Rouet & Britt, 2011; REading as problem SOLVing [RESOLV] Britt et al., 2017); Integrated Framework of Multiple Texts [IF-MT], List, 2020; List & Alexander, 2019). The Integrated Framework of Multiple Texts (IF-MT), for instance, identifies three stages students go through when they use multiple texts: preparation, execution, and production. In the preparation stage, contextual (e.g., task goals and task structure) and individual variables (e.g., goals, task perception, prior knowledge and affects) interact and lead the student to adopt a *default stance* (List & Alexander, 2019) that shapes the execution stage. The execution stage includes the orchestration of behavior and cognitive and metacognitive strategies in order to solve the task. At the end of this stage and depending on the learners' goals for accessing the text, students develop affective and/or cognitive outcomes. The cognitive outcome usually consists of a mental representation of the multiple texts' content. Finally, during the production stage, learners transform the outcomes of the execution stage into external, often written, products (List & Alexander, 2019). Similarly, the Multiple-Document Task-Based Relevance Assessment and Content Extraction (MD-TRACE) model proposes that after developing a cognitive representation of the task demands and having determined what information is needed, students select, process, and integrate information across texts to form a documents model (Britt, Perfetti, Sandak, & Rouet, 1999; Perfetti et al., 1999). A documents model is an integrated, cognitive representation of the central topic of the texts (Integrated Model), as well as a representation of source information and the construction of relations among texts as consistent or contradictory with one another (Intertext Model; Perfetti et al., 1999). As in the IF-MT model (List & Alexander, 2019), we considered the construction of relations among the texts or cross-textual linking (List & Alexander, 2019) to be a four-step process that requires: (1) identifying that the texts may be related; (2) generating separate representations of the content of each text that is relatable to those of the others; (3) combining the separate representations into a singular statement that contains information from several texts; and (4) making explicit the type of relation between the multiple texts. As suggested by

List and Alexander (2019), this process would explain how readers create both low-level and high-level cross-textual links, with low-level cross-textual links defined as those based on the explicit content information present in the texts and high-level cross-textual links defined as those that connect thematic elements that are not explicitly stated in the texts.

Cross-textual inferences might be different for texts containing complementary information and those containing conflicting information (List & Alexander, 2019). According to List and Alexander (2019), when reading texts with complementary information and after detecting a relation between them, students may engage in synthesizing this information. This synthesis "refers to the process of combining and organizing like relations into a unified whole" (List & Alexander, 2019, p.11) and can include connecting multiple cross-textual links (one text agrees with another), summarizing multiple cross-textual links (both texts agree with one another), or identifying an additional relation (thematic unit) that integrates multiple cross-textual links.

Although intratextual and intertextual processing involves different strategies, List and Alexander (2019) proposed that some of the strategies that have been found to be useful for processing a single text can also enhance the comprehension of multiple texts. This is the case, for example, the strategies for making inferences from information that is not explicit in the text or self-explanation during reading. Other strategies that might be useful for the comprehension of both single and multiple texts involve macrostructure comprehension strategies, such as summarizing or identifying the main idea, which require the selection of relevant information and the analysis of text structure. However, to comprehend multiple texts and their intertextual links, readers need to use additional and specific types of strategies, such as organizational strategies (which aim to construct connections between the information contained in the different texts).

Other studies have also described specific strategies required for multiple text comprehension, such as comparing the content of the different texts, examining whether the information in the multiple texts is compatible or contradictory, and connecting, combining and integrating the information found in the multiple texts to construct a coherent representation of the texts (Barzilai et al., 2018; Rouet & Potocki, 2018).

Regarding the extent to which questioning might help to create a more integrated representation of multiple texts, Britt and Sommer (2004) indicated that asking students high-level questions (macro-questions) on each individual text results in more integrated representations of multiple documents. However, asking intertextual questions produces more integration than just asking intratextual questions (Cerdán & Vidal-Abarca., 2008). Nevertheless, students find it more difficult to answer intertextual questions (List et al., 2016). The complexity of integrating information from multiple texts is compensated by the potential learning benefits for the reader, as one has to infer relationships that do not appear in any of the individual source texts (Wiley, Steffens, Britt, & Griffin, 2014).

From Research to Teacher-Led Intervention

Ever since Durkin (1978) drew attention to the fact that very little explicit instruction on reading comprehension was offered in schools, many studies have sought to demonstrate that it is possible to improve students' comprehension abilities across all educational levels. For instance, in relation to questioning and answering strategies, Palincsar and Brown (1984) and Paris, Cross, and Lipson (1984) conducted interventions focused on formulating and answering questions and on making inferences, which helped students to deepen comprehension. Notably, question generation and answering are among the seven evidence-based strategies highlighted by the U.S. National Reading Panel (2000). In fact, numerous studies have examined the use of questions after reading, which have been found to have a favorable impact on comprehension (for a review, see McKeown, Beck, & Blake, 2009). A meta-analysis of experimental and quasi-experimental studies published between 1950 and 2014 (Elleman, 2017) noted the effectiveness of interventions designed to improve inference-making ability among students in Grades 2–9. Taken together, these studies demonstrate that it is possible to teach students strategies for processing textual information that enable them to go beyond mere reproduction to integration.

However, many of the interventions applied in the classroom context have been implemented by researchers who were already familiar with the strategies they used as teachers and had a clear idea of what they wished to achieve (González-Lamas, Mateos, & Cuevas, 2016; Mason, Dunn Davison, Hammer, Miller, & Glutting, 2013). In cases where classroom innovations were implemented by teachers, the intervention was often subjected to protocols designed primarily by researchers, with teachers performing basically an executive role (Bråten, Brante, & Strømsø, 2019; Cantrell, Almasi, Carter, Rintamaa, & Madden, 2010). These approaches have generated a valuable body of knowledge, showing that such interventions can improve the quality of teaching. Nevertheless, they have their limitations, notably questionable ecological validity (more so in some instances than others) and the ephemeral nature of the learning fostered by teachers.

A related problem involves determining what qualifies as effective in-service professional development (PD). Anders, Hoffman, and Duffy (2000) indicated the characteristics of what they would consider high-quality in-service PD: availability of intensive support, encouragement

of teachers to systematically reflect on their own practice, use of debate and dialog, voluntary participation, and a collaborative spirit (between teachers and researchers). In a subsequent review, Dillon, O'Brien, Sato, and Kelly (2011) established similar criteria and suggested additional steps to achieve good-quality training: focus the PD on students' learning outcomes; base teachers' learning on their existing practice; ensure that the experience is lasting rather than anecdotal; and provide teachers with opportunities to work together on topics of importance to them. It has been shown that professional development programs that incorporate these criteria increase the use of the suggested practices among the teachers (Ingvarson, Meiers, & Beavis, 2005) and improve the students' learning (Powell, Diamond, Burchinal, & Koehler, 2010).

These and other contributions (e.g., Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009) have suggested that for a PD paradigm to achieve lasting change in teachers' classroom work, it must enable them to analyze and reflect on their individual practice using reliable theoretical instruments (Schön, 1987; see also Sato, Wei, & Darling-Hammond, 2008). Although the application of this paradigm has certain associated difficulties and costs, it is consistent with the socio-constructivist view of teaching and learning (Coll, 1990; Vygotsky, 1978) and with the idea of the teacher as a decision-making agent (Schön, 1987). This study is based on the aforementioned paradigm.

Study Aims

The aim of this study was to increase eighth-grade students' capacity to answer questions that go beyond explicit information and/or require the integration of information from one or more source texts (intra- and intertextual inferential questions, respectively), and to reliably assess the degree to which students have learned the content of texts that share complementary information. To achieve these objectives, we designed and implemented a professional development (PD) program for teachers in the intervention group that would help them to improve their existing questioning strategies. Teachers in the control group did not participate in the PD program until the data collection had been completed.

We sought to answer the following research questions on the intervention impact:

- Do students in the IG perform better than those in the CG when they are required to answer inferential (intra- or intertextual) questions at the end of the teaching unit and 2 months later?
- Do students in the IG perform better on a learning test covering the teaching unit?

We expected to find that students in the IG, whose teachers participated in the professional development program, would outperform those in the CG on intra- and intertextual questions that required making inferences, both immediately after the intervention (posttest) and 2 months later (follow-up). We also expected these students to exhibit better post-intervention learning outcomes as a result of this instruction, which would be measured at the end of the teaching unit as is usually done in the classroom.

Method

We designed an experimental intervention study with a pretest, posttest, and follow-up test administered to the intervention and control groups of students. The independent variable was teacher participation or nonparticipation in the PD program during implementation of the teaching unit. The dependent variables were students' answers to literal and inferential (intra- and intertextual) questions on course texts and their mastery of the content of the teaching unit subsequent to the intervention. We controlled for the following variables: students' general level of reading comprehension, their level of prior knowledge, and (through the pretest) their ability to answer different types of questions about expository texts before the intervention.

Participants

Participants in this study attended one of the nine schools that were involved. Based on data from the Ministry of Education in Catalonia, Spain, the participating schools could be considered comparable in terms of socioeconomic and educational status, as measured by parental education and occupation and by student performance on the periodic external assessments used by all schools in Catalonia. An additional school with the same characteristics was used to pilot the various instruments that would subsequently be used in the main study.

The participants were a total of 10 teachers and 369 eighth-grade students, aged 13–14. Following the requirements of the Bioethics Commission Board of the University of Barcelona, parents and students were informed of the aims of the research, and they provided written informed consent. Parents were also offered the opportunity to express their concerns and to request that their children's data not be included in the study. Scores from students diagnosed with special educational needs were excluded from the data analyses. Teachers were randomly assigned to the IG or CG. As some of the teachers in the IG taught more than one class of students (see Table 1), the number of students in that group was larger than that in the CG. Thus, the Intervention group (IG) was comprised

TABLE 1
Distribution of Participants (Schools, Teachers, and Students by Gender)

	Schools	Teachers	Group classes	Gender Female	Male	Total n° of students
Intervention group (n = 237)	School 1	Teacher 1	Class A	11	12	23
			Class B	12	6	18
			Class C	10	11	21
			Class D	10	11	21
	School 2	Teacher 2	Class A	10	9	19
	School 3	Teacher 3	Class A	14	10	24
			Class B	8	12	20
	School 4	Teacher 4	Class A	11	14	25
			Class B	9	14	23
	School 5	Teacher 5	Class A	15	8	23
			Class B	11	9	20
Control group (n = 132)	School 6	Teacher 6	Class A	6	10	16
			Class B	8	10	18
	School 7	Teacher 7	Class A	8	8	16
	School 8	Teacher 8	Class A	8	15	23
			Class B	7	10	17
	School 9	Teacher 9	Class A	10	8	18
		Teacher 10	Class B	12	12	24

of 5 schools, 5 teachers, and 237 students (51.1% female); the Control group (CG) included 4 schools, 5 teachers, and 132 students (44.7% female). The distribution of the participants is shown in Table 1.

To control the equivalence of the groups, we administered (1) the Test of Comprehension Strategies (Vidal-Abarca et al., 2007), (2) a prior knowledge test, and (3) a pretest measuring students' ability to answer different types of questions.

Statistical analysis revealed no significant differences between the IG and the CG on the measures of general reading comprehension, prior knowledge, or ability to answer literal, intra- and intertextual inferential comprehension questions (see the baseline data in the *Results* section).

As for the teachers, they all had at least 10 years of experience, and their participation in the study was voluntary. It is common for history teachers in our context to use a textbook that specifies the content and the tasks (mostly questions) for each teaching unit. Usually, students and their teacher read the content presented in the textbook. After reading, students are assigned different activities. These activities tend to be questions that students solve individually in class or as homework if there is not enough time to finish the lesson in the classroom. The answers are then checked by the teacher or corrected in class. The teachers who participated in our project used textbooks and followed this usual lesson structure before they started their professional development.

Student Measures and Materials

Reading Comprehension Test

Students' general level of reading comprehension was assessed by the Test of Comprehension Strategies (TEC; Vidal-Abarca et al., 2007). This norm-referenced reading test is designed for children aged 11–16. The questions from this test cover basic comprehension processes, that is, the formation of text ideas, anaphoric inferences, knowledge-based inferences, and macro-idea formation (Vidal-Abarca et al., 2007). Students are asked to read two different expository texts (one about penguins and another on the Sioux), and then answer ten multiple-choice questions on each text. This test has been validated and used as a measure of general comprehension (Cerdán, Gilabert, & Vidal-Abarca, 2011, 2013). Cronbach's α for the test's reliability was .80.

Prior Knowledge Test

We developed an *ad hoc* test that was evaluated by two external history teachers who assessed the language and potential difficulty of the items. After it had been piloted, it included 19 multiple-choice items. Four items referred to topics addressed by the pretest texts; 11 covered the content of the teaching unit; and four concerned the topics covered by the follow-up assessment texts (see Table 2 and Appendix A for examples of questions). Several of the questions for the pretest and posttest had also been used and validated in a previous study (Nadal, 2021). Cronbach's α for this test was .72.

Pretest (T1), Posttest (T2), and Follow-Up (T3) Comprehension Texts and Questions

Three pairs of texts, each with an associated set of open questions, were used to assess students' comprehension at three time points: prior to the intervention (T1, pretest of the topic: demographic changes in the European population during the 15th and 16th centuries), at the end of the intervention (T2, posttest: Economic activity and culture in al-Andalus), and 2 months later (T3, follow-up: The Catholic Kings). Each of the pairs of texts contained a core of distinct complementary information. The texts were adapted from the textbooks used by the teachers, and all the topics addressed by these texts were featured in the official curriculum (see Table 2). The level of difficulty of the texts was considered suitable for middle-school students, by their respective teachers as well as according to the Flesch-Szigriszt Index (INFLESZ; adaptation by Fernández-Huerta, 1959) of readability. An INFLESZ value above 55 is considered appropriate for this age group, and the values obtained ranged from 65 to 75.

To develop the tests associated with each pair of texts, the research team identified core information in the texts and then composed a series of literal and inferential questions regarding the key points. The literal questions required students to identify ideas within a single proposition in the text; the inferential questions required them to relate ideas located within different blocks of core information, either within the same text (intratextual inferential questions) or across both texts in a pair (intertextual inferential questions). The information to be integrated for the intertextual inferential questions was mostly distinct and complementary in nature (Nadal et al., 2021), since the students had to select the information presented in both texts, organize it, and integrate it in order to answer the questions (see the example in Appendix B). Intertextual inferential questions, therefore, required the students to create high-level cross-textual links (List & Alexander, 2019).

In addition, we calculated the reliability for each type of question per test moment using Cronbach's alpha corrected for test length using the Spearman–Brown extension: Cronbach's α *Mean* for literal questions = .66 (min: .60; max: .72); Cronbach's α *Mean* for intratextual questions = .68 (min: .60; max: .75); Cronbach's α *Mean* for intertextual questions = .77 (min: .69; max: .82). Levels of internal consistency reliability for the different questions are acceptable. Loewenthal (2001) suggests that reliability of .60 may be considered acceptable for scales with fewer than 10 items, which occurs in our three assessment moments. In Appendix C, we provide the results per test moment and the correlations for the different types of questions per test moment. One example of the texts used in the posttest together with one question of each type is provided in Appendix B.

Coding of the Students' Responses

Students' responses on the pretest, posttest, and follow-up tests were content analyzed and checked for interrater

TABLE 2

Characteristics of the Texts and the Types of Questions Associated with each Task

Task	Content	Texts	Number of words	Number and type of questions associated
Pretest[a]	Demographic changes in the European population during the 15th and 16th centuries and their impact on agriculture	Pretest TA: Causes of demographic growth during the 15th and 16th centuries in Europe Pretest TB: The impact of demographic growth in Europe	616 672	8 questions: 2 literal 4 intratextual inferential 2 intertextual inferential
Posttest[b]	Characteristics of Al-Andalus society. Economic activity and cultural exchange among social groups	Posttest TA: Social groups and ways of life in Al-Andalus society Posttest TB: Economic activity in Al-Andalus	639 448	8 questions: 2 literal 4 intratextual inferential 2 intertextual inferential
Follow-up	The Hispanic monarchy: the Catholic Kings	Follow-up TA: Isabella and Ferdinand, the Catholic Kings Follow-up TB: The Iberian Peninsula in the 15th century: the expansion of the Hispanic monarchy	471 541	7 questions: 3 literal 2 intratextual inferential 2 intertextual inferential

Note. [a]The texts used for the pretest and follow-up assessments were adapted from a study by Nadal (2021). [b]The texts used for the posttest assessment brought to a close the teaching unit implemented by teachers in both the IG and CG.

reliability. For each test, the first step involved establishing the criteria for evaluating answers to each question. Answers to literal questions were scored as 0 if incorrect or 1 if correct. For the intertextual inferential questions, we adapted the SOLO taxonomy (Biggs & Collins, 1982), to rank students' written responses according to the number of correct relevant pieces of information they included, and the completeness of the information that was integrated. We adjusted the SOLO scale from 0 to 1. A score of 0 corresponded to students who provided erroneous information. A score of 0.25 corresponded to students who provided incomplete information from a single text. A score of 0.5 corresponded to responses that included complete information from one text. A score of 0.75 corresponded to responses that included information from the two texts, but in an incomplete manner. A score of 1 corresponded to responses that included complete information from the two texts.

For the intratextual inferential responses, we followed a similar pattern since students were supposed to integrate information within a single text. Therefore, the scoring procedure was as follows: 0 for an erroneous response; 0.25 for an incomplete response that focused on a single piece of intratextual information; 0.5 for a response that focused on a single complete piece of intratextual information; 0.75 for a response that included incomplete information from several intratextual pieces of information; and 1 for the complete response that included information from several intratextual pieces of information.

Next, six researchers analyzed the same 20 sets of test responses, which were randomly selected to check the degree of agreement between pairs of raters. Any disagreements were resolved by consensus. The kappa indices for interrater agreement were adequate, as they ranged between .71 and 1. Once any disagreements had been resolved, the test responses of all participating students were distributed among the six researchers for evaluation.

Learning Test

This *ad hoc* test comprised 16 multiple-choice items designed to assess students' level of learning in the core knowledge areas addressed by the teaching unit, considering only the content that had been implemented by the teachers using the texts adapted by us. It was first given to two external history teachers who assessed the language and potential difficulty of the questions and it was then piloted. The test, which had a reliability of Cronbach's $\alpha = .65$, was administered within 2–3 days of the end of the teaching unit and the administration of the posttest (see Appendix D for a sample of questions).

Professional Development Program Materials

For the professional development (PD) led by one of the authors, we developed two sets of materials: (a) four pairs of texts presenting complementary content from a history teaching unit (see Table 3), which all the participating teachers would teach in their classrooms, and hence it provided the focus for the PD seminars; and (b) three short dossiers containing ideas and suggestions for classroom practice regarding reading and answering questions from texts.

Teaching Unit

The teaching unit for the intervention was chosen with the teachers so that they would feel comfortable about participating in the professional development and implement the content according to the planned schedule. The teaching unit that they chose was entitled "The Origin and Spread of Islam: Ways of Life and Confluence of Cultures in al-Andalus." The texts, which were expository in nature, were developed with reference to the history curriculum and textbooks used by eighth-grade students. Pairs of texts were created in a way that each pair related to one of the subtopics into which the teaching unit was divided.

All the texts that teachers and students worked with had the same format as those typical of eighth-grade

TABLE 3
Characteristics of the Teaching Unit Texts Used in the Professional Development Seminars

Subtopic of the teaching unit	Content of the texts	Number of words
Origin and principles of Islam: the historical context of its emergence	T1. The emergence of Islam: historical context T2. Origin and beliefs of the Islamic religion	881 496
Spread of Islam in the Iberian Peninsula: Al-Andalus	T3. The spread of Islam T4. Islam in the Iberian Peninsula	442 703
Cultural contributions of Islam: the legacy of Al-Andalus	T5. Culture in the Arab world T6. The cultural legacy in the Iberian Peninsula	655 649
Characteristics of Al-Andalus society: economic activity and cultural exchange	Posttest TA: Social groups and ways of life in Al-Andalus society Posttest TB: Economic activity in Al-Andalus	639 448

textbooks. The teachers who participated in the professional development seminars considered the material's difficulty level suitable for their students. Additionally, the use of INFLESZ indicated that the texts were appropriate for students aged 13–14 (the values ranged from 63 to 74, above the recommended minimum of 55).

Dossiers

The three dossiers (8–9 pages each) given to the teachers were used as the basis for discussion in the training sessions. Each dossier addressed a specific topic: (1) reading strategies, (2) types of questions that can be formulated to promote students' comprehension of a text, and (3) the general and specific characteristics of the process of reading from multiple texts. All the dossiers began with a brief theoretical introduction, followed by recommendations about how to apply the relevant concepts to classroom practice.

Professional Development Seminars

As a group, the teachers in the IG attended 12 hours of face-to-face professional development seminars over 6 weeks (one 2-hour seminar session per week).

The main aims of the PD program were to raise teachers' awareness of how different types of questions can help students achieve different levels of comprehension and to provide deeper information about what comprehension and learning from texts imply. The program was based on an analysis of teachers' actual classroom practice and was designed to aid and scaffold their implementation of the intervention. Since we aimed for teachers to gradually release responsibility (Fisher & Frey, 2013; Pearson & Gallagher, 1983) when they taught question answering strategies, we also tried to follow this model in the professional development program. The activities we planned were sequenced in a way that the responsibility could be transferred to the teachers with decreasing levels of scaffolding (e.g., modeling how to read and seek for the relevant information from a text; guided practice with the dossiers; collaborative activities; and finally individual activities).

Thus, we started the PD seminar sessions by assessing the teachers' prior knowledge on questioning (*pool of initial questions*). In the following sessions, we discussed and agreed with the teachers specific ways of presenting the pairs of texts, promoting their students' comprehension of them, and helping their students give more adequate and complete answers to both literal and inferential questions. We gave the teachers examples, helped them analyze, and revise their initial formulated questions and encouraged them to develop their own inferential questions (both intra- and intertextual) regarding the texts that their students would be reading. After that, the teachers agreed on the questions that were going to be used in the classroom so that all the teachers in the IG presented the same

written questions to their students for each pair of texts. In this context, we discussed and modeled different strategies with the teachers that might be useful for the tasks they would be setting, such as:

- Having the students begin by carefully and silently reading the texts.
- Clarifying concepts and responding to students' queries about the meaning(s) of the texts.
- Encouraging students to analyze and verify their understanding of each question before answering—attending to the wording of the questions.
- Helping students to identify the ideas that needed to be reflected in an answer, which are more complex for the inferential than for the literal questions.
- Encouraging students to make predictions based on the text.
- Helping students to make inferences based on the information contained in one or both texts of a pair by selecting information, connecting information (between paragraphs in one text or between paragraphs from multiple texts) and organizing the information (selected from more than one text).
- Motivating students to write clear, full, and well-structured answers to each question.
- Helping students to paraphrase the information contained in the text(s).
- Modeling all the previous strategies to make the thinking of the teacher explicit when different types of questions are answered.

This allowed the teachers to plan a detailed script to be followed during the implementation of each pair of texts. This plan included starting the sessions by stating the purposes of the activity, followed by activating prior knowledge among the students about the texts that were going to be read and encouraging the students to make some predictions. The students would then read silently and identify unknown words. This would be followed by reading aloud in class and resolving any queries, with the teachers asking some questions orally. After clarifying the information, the students would start reading the second text of a pair following the same procedure. In the next session, the teacher would start by asking what had been done in the previous session and then rereading the text aloud to clarify the meanings. At the end of this activity, the teacher would present all the questions and, for the first pair of texts, model how the students should proceed to answer some of them (see Table E1 in Appendix E). This would allow the students to answer the other questions in pairs. In the same session, the students might review some of the answers, with the teachers helping them select, connect, and organize the information from one or more texts. This sequence

would be similar for the other pairs of texts, which would allow more independent work by the students as they make progress in answering the questions by themselves.

After the fourth seminar session and alongside the PD program, each teacher in the IG began teaching the content of the teaching unit at their own pace. They used the materials developed for this purpose and applied the ideas and strategies (focused primarily on the formulation and answering of comprehension questions) that had been discussed and agreed between the PD program participants and the researcher leading the PD program. The latter ensured that the proposed questions included intra- and intertextual questions. A detailed description of the PD program and a summary of it are provided in Appendix E.

Fidelity of the Intervention

Pool of Teachers' Initial and Final Questions

The participating teachers were asked to indicate the tasks and/or questions they planned to give their students (pool of teachers' initial questions) regarding the first pair of texts in the teaching unit (T1 and T2 in Table 3). This inquiry allowed us to ensure that teachers in both groups had comparable approaches prior to the intervention. By the end of the teaching unit, teachers were also asked to design the tasks and/or questions that they would assign their students (pool of teachers' final questions) on the final pair of texts (TA and TB in Table 3). In this way, we could check that the teachers on the CG continued to propose questions of a similar complexity before they worked on the pairs of texts, compared to the teachers participating in the PD program (results from the analysis of teachers' pool of questions are provided in Appendix F).

Teacher Logs and Checklist of Activities

To ensure that teachers in the IG actually implemented the questions covered during the professional development program, and to keep a record of the questions used by teachers in the CG, we created a checklist (Judd, Smith, & Kidder, 1991), on which teachers were asked to indicate the following information for each class taught as part of the teaching unit: the text(s) they worked with; the activities used and questions asked, along with their order; and the time spent on each question or activity. They also had to describe the sequence of work carried out and provide what they considered were three significant examples of how their students had resolved the given problems (see Appendix G).

Our analysis of these data and the examples of the activities solved by the students in each group confirmed that the teachers in the CG continued to propose similar activities and questions to those that they had indicated initially (i.e., literal and intratextual inferential questions). Furthermore, this enabled us to verify that teachers in the

IG were incorporating and posing the questions agreed to in the PD program (see Appendix G).

Observations

In addition to the other two instruments, to assess the extent to which the IG teachers' lesson was in line with what had been agreed with them, and to verify what the CG teachers were doing with the texts, one class by each teacher was observed (10 observations in total, approximately 50–60 minutes each). Observations were conducted on the second or third session of the implementation of the one and two pair of texts. This allowed us to see whether the teachers were considering the two texts together and posing questions that related to information contained in them, or if they were treating the texts as single documents. A checklist similar to the one filled in by the teachers was used (Judd et al., 1991), and included the same items (texts read during the session; activities and questions and the order in which they had been asked; time spent). We also added the reading strategies taught during the session. The observations revealed that there were no substantial differences between what the teachers explained in the checklist and logs, and what was observed subsequently.

During the implementation period in the CG, no specific instance of teaching about question answering occurred and we could observe the use of literal and intratextual inferential questions exclusively. Although some of the teachers referred to the content of text 1 before or when text 2 was being read, questions and activities were only about text 2. According to the checklist, in most cases, the questions and activities on text 1 had been implemented in a previous session. Thus, although the teachers in the CGs worked with the same pair of texts in the same format than those in the IGs, the instructional activities in their classes did not focus on working with them as multiple complementary texts, and there were no instances of intertextual inferential questions. Additionally, although teachers in the CGs sometimes checked the answers to questions with their students and asked them to explain the answers, there were few instances in which they showed the importance of understanding the question, how to select the information in the text or how to paraphrase it. In contrast, teachers in the IG followed the guidelines that had been agreed (see the Professional development seminars section). Observations showed that they concentrated on question-answering strategies after reading both texts. The teachers tended to discuss and model how students had to proceed to answer the intra- and intertextual inferential questions correctly, that is, how to select information and reflect on information that was complementary between the two texts, and how to connect it. The teachers also gave the students prompts to organize the information contained on the texts.

Data Collection Procedure

After designing the aforementioned materials, we piloted them at a school with characteristics similar to those of the schools in the main study, and adjusted them accordingly. At the beginning of the academic year, once the teachers had been randomly assigned to the control or the intervention group, we administered the pretest measures to their students.

Concomitantly, we asked teachers to indicate the activities and/or questions that they planned to use in relation to the first pair of texts in the teaching unit (initial pool of teachers' questions). Teachers in the CG performed this task during a first meeting with the researchers. This meeting was designed to get to know the teachers, to find out some basic information about them, to organize the observations and show them the pairs of texts they would work on with their students. Teachers were also shown how to fill in the checklist. Teachers in the IG were asked to do the pool of initial teachers' questions during the first seminar session. The PD program for the IG teachers consisted of six seminar sessions (see *Professional development seminars* section, and Appendix E for a summary). During the implementation of the intervention, teachers in both groups met with their students twice a week for approximately 50–60 minutes each class session. Teachers in the CG, who did not participate in the PD, delivered the teaching unit using the same reading materials and were instructed to work with the texts using the activities they would normally employ in their classrooms. These teachers did not have access to the dossiers, or the questions covered in the PD program. Teachers in the IG delivered the teaching unit while attending the PD seminars (see Figure 1). All the participating teachers filled out the checklist each time they used the texts in the teaching unit.

The teachers in the intervention group and the teachers in the control group implemented the subject of the intervention in 10 lesson sessions (55 minutes each lesson approximately). Once the teachers had finished working with texts 5 and 6 of the teaching unit they informed us, and the research team visited the school and administered the posttest to all participating students. At this point, all teachers were required to indicate the activities they would propose to their students related to the pair of posttest texts (TA & TB) (pool of final teachers' questions). Two months later, the students took the follow-up tests. Upon completion of the data collection, the teachers in the CG received the 12-hour PD program.

The collection and handling of data met all the ethical requirements imposed by the university.

Results

Baseline Data

We performed non-parametric Mann–Whitney U tests to analyze differences between the IG and the CG because the pretest scores deviated from normality. The results obtained by students in the CG and IG prior to the intervention are presented in Table 5. The Mann–Whitney U test found no significant differences between the two groups on any of these variables.

The results in Table 4 indicate that, as expected, students in both groups had a low level of prior knowledge about the content of the teaching unit they would be learning during the intervention period. Additionally, the two groups did not significantly differ on the reading comprehension test. The means were slightly below what one might expect for eighth graders, as they were in approximately the 40th percentile (Vidal-Abarca et al., 2007).

The pretest results showed that the students could answer most of the literal questions correctly but struggled with the inferential questions, especially the intertextual

FIGURE 1
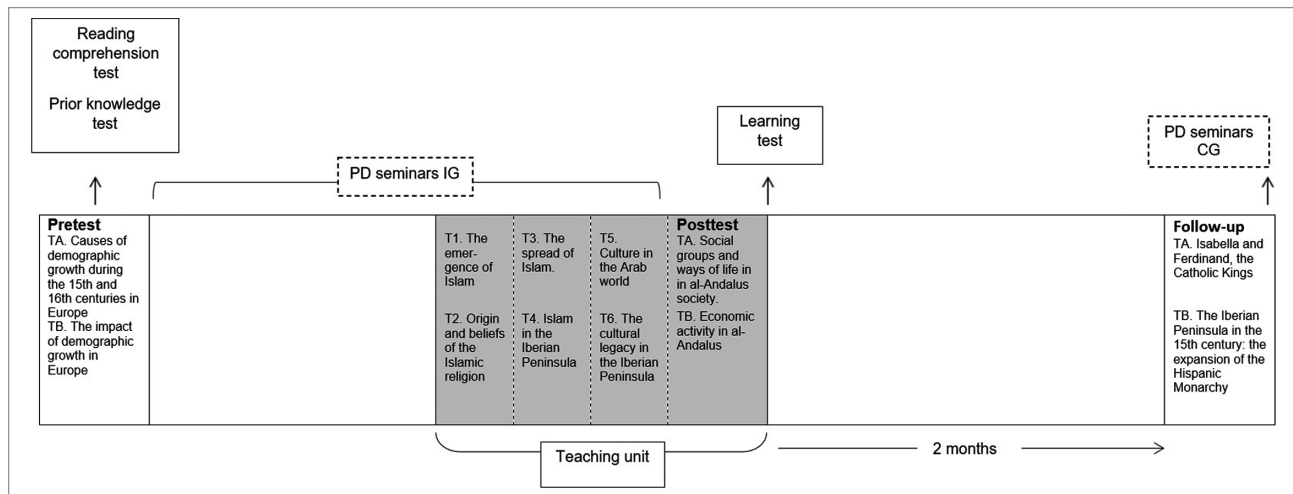**Summary of the Data Collection Procedure**

TABLE 4
Means, Standard Deviations, and Mann—Whitney U test Results for Pretest Comparisons Between Students in the Intervention and Control Groups

Variables	Control group M (SD)	Intervention group M (SD)	Comparison of means
Reading comprehension test	11.96 (4.04)	12.04 (3.95)	U = 15605, p = .970
Prior knowledge test	36.60 (12.37)	34.31 (12.93)	U = 14048.5, p = .102
Literal questions	75 (29.95)	77.85 (30.62)	U = 14718, p = .276
Intratextual inf. Questions	32.10 (27.29)	33.07 (25.63)	U = 15164, p = .622
Intertextual inf. Questions	20.08 (23.92)	22.73 (23.46)	U = 14315.5, p = .164

Note. Values for the reading comprehension test are based on raw scores (maximum = 20). The remaining results are given as percentages.

TABLE 5
Pearson Correlations between Personal Variables and Comprehension Question Results on Posttest (T2) and Follow-up Test (T3)

	T2			T3		
	Literal questions	Intratextual questions	Intertextual questions	Literal questions	Intratextual questions	Intertextual questions
Reading comprehension test	.267**	.337**	.360**	.211**	.211**	.319**
Prior knowledge test	.164**	.165**	.207**	.012	.149**	.213**

Note. **p < .01

ones. The CG and IG did not significantly differ in their ability to answer any of these three types of questions (see Table 4).

Impact of the Intervention on Comprehension Results

Before focusing on students' performance after the intervention, we conducted correlation analyses (see Table 5) to identify any relationships between the personal variables (levels of reading comprehension and prior knowledge) and the results achieved on any of the question types on the posttest (T2) and follow-up test (T3).

As demonstrated in Table 5, the personal variables were significantly correlated with the comprehension question results in most cases (and there was a correlation among the two personal variables; r = .187, p < .01). Therefore, we decided to include both the variables in the subsequent analyses.

Because the participants were nested within classes, a mixed-model analysis was chosen so that the variance within and between participants could be estimated simultaneously (Braaksma, Rijlaarsdam, & van den Bergh, 2018; Field, 2013). For each of the comprehension measures obtained after the intervention, several models were calculated. In what is known as the empty model (Model 0), we estimated the impact of the random factor (pertaining to a specific group-class) and we calculated the Intraclass Correlation Coefficient (ICC). Personal variables (reading comprehension and prior knowledge) were added as co-variables in the first and second models, respectively.

In the third model, we included the time period when the test occurred (i.e., T1, T2, or T3) as a factor. In the fourth model, we added the condition (IG versus CG). In the fifth model, we considered the interaction between time and condition; in the sixth model, the reading comprehension test was added to the interaction; and in the seventh model, prior knowledge was added. Moreover, in both the sixth and seventh models, the personal variables (general comprehension ability and prior knowledge) were added to the interaction. The difference in fit between these (nested) models can be tested by −2loglikelihood, as the difference −2loglikelihood in nested models is $\chi2$-distributed (with the difference in the number of estimated parameters represented as degrees of freedom). To determine the effect of the intervention on T2 and T3, T1 was used as the baseline. Effect sizes were determined by calculating Cohen's d, taking into account the estimated means and standard deviations for each significant model.

Table 6 reports the seven models and analyses the effect of the intervention on literal question performance on the posttest and follow-up test.

As shown in Table 6, the best-fit model for the literal questions was the one that added time to the analysis of

TABLE 6
Comparison of Models with Student performance on Literal Questions as a Dependent Variable at T2 and T3

Model	−2log likelihood	Models	χ^2	df	P
		Comparison			
		T2			
0. Intercept + random component (participants) ICC = .11	7.022.95				
1. M0 + Comprehension (C)	6999.17	0 vs. 1	23.778	1	< .001
2. M1 + Prior Knowledge (PK)	6993.68	1 vs. 2	5.493	1	.019
3. M2 + Time	6985.66	2 vs. 3	8.021	1	.005
4. M3 + Condition[a]	6982.66	3 vs. 4	2.996	1	.083
5. M4 + interaction Time*Condition	6982.49	4 vs. 5	.175	1	.676
6. M5 + interaction Time*Condition*C	6975.89	5 vs. 6	6.597	3	.086
7. M6 + interaction Time*Condition*C* PK	6975.55	6 vs. 7	.337	1	.562
		T3			
0. Intercept + random component (participants) ICC = .15	7053.90				
1. M 0 + Comprehension (C)	7036.40	0 vs. 1	17.502	1	< .001
2. M1 + Prior Knowledge (PK)	7036.24	1 vs. 2	0.159	1	.690
3. M2 + Time	7032.30	2 vs. 3	3.942	1	.047
4. M3 + Condition	7024.88	3 vs. 4	7.423	1	.006
5. M4 + interaction Time*Condition	7021.35	4 vs. 5	3.521	1	.061
6. M5 + interaction Time*Condition*C	7018.61	5 vs. 6	2.740	3	.433
7. M6 + interaction Time*Condition*C* PK	7017.86	6 vs. 7	0.754	1	.385

Note. [a]The control group was the reference group, labeled as 0; the intervention group was labeled as 1.

the posttest (Model 3: $\beta = -6.06$; $SE = 2.13$; t (445.79) = -2.84; $p = .005$). Thus, the students achieved better results on the posttest (T1: $EM = 77.25$; $SE = 1.50$; T2: $EM = 83.31$; $SE = 1.50$) as an effect of time but not as an effect of the intervention (see posttest Model 4). The effect size (ES), computed by calculating the difference between T2-T1, was $d = .23$, showing a small impact of the effect of time. For the follow-up test, in contrast (see Table 9), adding the condition (intervention versus control) improved the model fit (Model 4: $\beta = -6.39$; $SE = 2.33$; t (373.44) = -2.73; $p = .006$). To compute the ES, we calculated the difference between the comprehension results of the IG and the CG at follow-up. The result shows a small effect of the intervention (condition) for the literal questions ($d = .14$).

Table 7 shows that Model 4 had the best fit in explaining student performance on intratextual inferential questions in the posttest. Thus, students in the IG obtained higher results than those in the CG ($\beta = -4.488$; $SE = 1.98$; t (368.54) = -2.26; $p = .024$, irrespective of the measurement occasion; $d = .33$). This effect size shows that the

intervention was more effective in improving students' responses to intratextual questions than regular teaching.

Regarding the follow-up test, the interaction between condition and time of assessment was significant (see Model 5 in Table 7; $\beta = 10.37$; $SE = 4.11$; t (366.84) = 2.52; $p = .012$). This result shows that students in the IG performed better than students in the CG (see Table 9). In terms of effect sizes, the effect of time in IG ($d = .98$) is larger than in CG ($d = .58$), demonstrating a moderate impact of the intervention for the intratextual questions in the follow-up.

For intertextual inferential questions, the inclusion of the interaction between time and condition significantly improved the model fit on both the posttest and the follow-up test (see Model 5, Table 8).

These interactions show that the IG outperformed the CG on T2 ($\beta = 7.67$; $SE = 2.89$; t (369.07) = 2.65; $p = .008$) and T3 ($\beta = 6.94$; $SE = 3.42$; t (368.17) = 2.03; $p = .043$). On T2, the effect of time in IG ($d = .53$) is larger than in CG ($d = .20$), showing the impact of the intervention. On T3, the effect of time in the IG ($d = .65$) is larger than that in the

TABLE 7
Comparison of Models with Student Performance on Intratextual Inferential Questions as a Dependent Variable at T2 and T3

Model	−2log likelihood	Comparison			
		Models	χ^2	df	P
T2					
0. Intercept + random component (participants) ICC = .20	6932.635				
1. M0 + Comprehension (C)	6844.636	0 vs. 1	87.999	2	< .001
2. M1 + Prior Knowledge (PK)	6804.601	1 vs. 2	40.035	1	< .001
3. M2 + Time	6748.441	2 vs. 3	56.160	1	< .001
4. M3 + Condition[a]	6743.346	3 vs. 4	5.095	1	.024
5. M4 + interaction Time*Condition	6739.642	4 vs. 5	3.704	1	.054
6. M5 + interaction Time*Condition*C	6735.465	5 vs. 6	4.177	3	.243
7. M6 + interaction Time*Condition*C* PK	6733.72	6 vs. 7	1.745	1	.187
T3					
0. Intercept + random component (participants) ICC = .02	7283.787				
1. M0 + Comprehension (C)	7228.621	0 vs. 1	55.166	1	< .001
2. M1 + Prior Knowledge (PK)	7130.952	1 vs. 2	97.669	1	< .001
3. M2 + Time	7069.598	2 vs. 3	61.354	1	< .001
4. M3 + Condition	7062.177	3 vs. 4	7.421	1	.006
5. M4 + interaction Time*Condition	7056.156	4 vs. 5	6.021	1	.014
6. M5 + interaction Time*Condition*C	7051.736	5 vs. 6	4.420	3	.220
7. M6 + interaction Time*Condition*C* PK	7051.504	6 vs. 7	0.232	1	.630

Note. [a]The control group was the reference group, labeled as 0; the intervention group was labeled as 1.

CG (d = .37), implying that the impact of the intervention was maintained.

The estimated means and the standard error for the questions that demonstrated an impact of the condition (either alone or combined with another variable) are summarized in Table 9.

Finally, performance on the reading comprehension and on the prior knowledge tests were significant predictors of performance on all types of questions (see Tables 6, 7, and 8), but in no case was there an interaction with condition. This result indicates that the effect of condition is generalizable across all the levels of prior knowledge and general reading comprehension.

Impact of the Intervention on Learning

To analyze the impact of the intervention on students' learning, we performed a mixed-model analysis, with group-class as a random factor in Model 0, personal variables as co-variables in Models 1 and 2, and condition (CG versus IG) in Model 3. Finally, Models 4 and 5 deal with the interaction between personal variables and condition (Table 10).

The model fit was significantly better when the condition was added (Model 3: β = −4.56; t (369) = −3.24; p = .001; d = .35). Therefore, students in the IG achieved better scores on the learning test (M = 73.97, SE = .84) than those in the CG (M = 69.41, SE = 1.12).

Performance on the reading comprehension and prior knowledge tests were significant predictors of the learning test results, but there was no interaction with condition. This again indicates that the effect of condition is generalizable across levels of prior knowledge and general reading comprehension.

Discussion and Conclusions

In this study, we aimed to improve students' reading comprehension skills and questions answering through a

TABLE 8
Comparison of Models with Student Performance on Intertextual Inferential Questions as a Dependent Variable at T2 and T3

Model	−2log likelihood	Comparison			
		Models	χ^2	df	*P*
		T2			
0. Intercept + random component (participants) ICC = .34	6815.978				
1. M0 + Comprehension (C)	6715.396	0 vs. 1	100.582	2	< .001
2. M1 + Prior Knowledge (PK)	6695.079	1 vs. 2	20.317	1	< 0.001
3. M2 + Time	6663.150	2 vs. 3	31.929	1	< 0.001
4. M3 + Condition[a]	6651.707	3 vs. 4	11.443	1	0.001
5. M4 + interaction Time*Condition	6644.755	4 vs. 5	6.952	1	0.008
6. M5 + interaction Time*Condition*C	6641.842	5 vs. 6	2.913	3	0.405
7. M6 + interaction Time*Condition*C* PK	6639.892	6 vs. 7	1.950	1	0.163
		T3			
0. Intercept + random component (participants) ICC = .23	7015.662				
1. M0 + Comprehension (C)	6925.968	0 vs. 1	89.694	1	< .001
2. M1 + Prior Knowledge (PK)	6883.834	1 vs. 2	42.134	1	< .001
3. M2 + Time	6850.019	2 vs. 3	33.815	1	< .001
4. M3 + Condition	6841.837	3 vs. 4	8.182	1	.004
5. M4 + interaction Time*Condition	6835.744	4 vs. 5	6.093	1	.014
6. M5 + interaction Time*Condition*C	6834.757	5 vs. 6	0.987	3	.804
7. M6 + interaction Time*Condition*C* PK	6834.006	6 vs. 7	0.751	1	.386

Note. [a]The control group was the reference group, labeled as 0; the intervention group was labeled as 1.

TABLE 9
Estimated Means and Standard Deviations for Cases when the Condition Affected Posttest or Follow-up Scores

Type of question	Test	Model	Control group	Intervention group
Literal	Follow-up	4 (Condition)	75.25 (30.75)	79.96 (24.85)
Intratextual	Post	4 (Condition)	33.17 (15.05)	46.61 (20.32)
Inferential	Follow-up	5 (Time*Condition)	49.37 (31.48)	61.01 (31.40)
Intertextual	Post	5 (Time*Condition)	24.73 (22.63)	35.18 (22.93)
Inferential	Follow-up	5 (Time*Condition)	29.53 (27.22)	39.17 (27.24)

professional development program for teachers. To evaluate the program's impact, we analyzed students' ability to answer inferential questions in one or multiple complementary texts. We compared their performance with that of a control group of students whose teachers had not participated in the professional development program. We also compared the learning results obtained by both groups.

Impact of the Intervention on Comprehension Results

Our first research question was whether the students in the IG would outperform those in the CG when they were asked to answer inferential questions following the completion of the teaching unit. The results revealed that on both posttest and follow-up test, students in the

TABLE 10
Comparison of Models with Student Results on the Learning Test as a Dependent Variable

Model	−2log likelihood	Comparison			
		Models	χ^2	df	*P*
0. Intercept + random component (participants) ICC = .33	3063.802				
1. M0 + Comprehension (C)	2968.590	0 vs. 1	95.212	2	< .001
2. M1 + Prior Knowledge (PK)	2946.314	1 vs. 2	22.276	1	< .001
3. M2 + Condition[a]	2935.910	2 vs. 3	10.404	1	.001
4. M3 + interaction Condition*C	2934.76	3 vs. 4	1.150	1	.284
5. M4 + interaction Condition*C*PK	2931.153	4 vs. 5	3.607	1	.165

Note. [a]The control group was the reference group, labeled as 0; the intervention group was labeled as 1.

IG performed significantly better than their peers in the CG at answering intra- and intertextual inferential questions. In the intratextual questions on the follow-up test and the intertextual questions on both the tests, the condition, together with time (i.e., the moment at which the assessment was performed), explained the IG's superior achievement. Thus, our research confirms that students' answers to comprehension questions improve in comparison to the CG when they receive explicit instructions on how to process textual information in greater depth. This finding is consistent with the research by Paris et al. (1984) and the studies reviewed by Elleman (2017). The results of our study demonstrate that it is possible to improve students' ability to answer questions by promoting their teachers' professional development and helping them incorporate new strategies for grasping the meaning of texts.

Although understanding and integrating information from multiple texts is generally an objective reflected in the Curriculum for Secondary Education, many students find it challenging to work with multiple texts (List, 2020; Rouet, 2006). Consequently, various researchers have suggested that they need adequate, explicit instructions to benefit from this kind of learning experience (Bråten et al., 2019; Bråten, Ferguson, Anmarkrud, & Strømsø, 2013; Britt & Rouet, 2012; Rouet, 2006). Many of these studies have used texts that contain opposing viewpoints or arguments on the same topic (Bråten et al., 2019; Rouet & Britt, 2014). By contrast, the pairs of texts used in our study contained distinct complementary information (List et al., 2021; Nadal et al., 2021), which is something relatively unusual in this field of research. This scenario of complementary, rather than directly opposed texts, deserves further investigation (Barzilai et al., 2018; Cerdán & Vidal-Abarca, 2008; Firetto, 2020), considering that teachers at middle-school level assign tasks that involve integrating information from texts with consistent, supplementary information (Solé et al., 2013).

In addition to the professional development, teachers who participated in the study recognized the potential of inferential questions to promote deeper understanding. However, they also mentioned that they were relatively unaccustomed to teaching students how to understand and integrate information from multiple texts in the classroom, even if they may have usually given homework assignments that required their students to read more than one source text and produce written answers based on those texts. The professional development program enabled them to identify differences between types of questions, reflect on the levels of comprehension that such questions encourage, and increase the number of inferential questions that they would use with their students in their regular classroom practice, as shown by the results of the analyses of the pool of questions.

One finding about literal questions merits comment, even though we did not specifically aim to improve students' performance on these questions. The IG performed significantly better on literal questions than their control peers on the follow-up assessment two months later. Although we can offer no firm explanation for this finding, some studies that sought to improve students' inferential comprehension also achieved a clear improvement in their ability to answer literal questions. This effect has been observed when skilled and less skilled readers are compared (Elleman, 2017), and has been attributed to the explicit instruction that students received on how to locate relevant information and integrate it with their existing knowledge to answer inferential questions. Therefore, it is plausible that this teaching strategy has a positive impact on students' ability to access the information required to answer literal questions. Alternatively, the students in the IG may have consolidated their newly developed skills more strongly, and these skills would have been maintained more fully during the period after the intervention.

Impact of the Intervention on Learning

Our second research question asked whether the IG would perform better on a learning test covering the material in the teaching unit. The results were in line with our hypothesis. As noted in the introduction, inferential questions are usually associated with more in-depth understanding of content (Graesser & Lehman, 2011; Taboada & Guthrie, 2006). Furthermore, the need to integrate information from two texts may potentially facilitate greater learning because the reader must construct high-level cross-textual links that are not explicitly present in the source material (List & Alexander, 2019; Rouet, 2006; Wiley et al., 2014). The students in our IG were instructed to pay close attention to the wording of questions, to read the texts carefully before formulating their answers, and to ensure that their answers were complete and well-organized. As they were specifically shown how to answer different types of inferential questions, we expected them to show greater learning at the end of the intervention. The results confirmed this expectation and indicated that the intervention promoted more solid learning among students in the IG than among their peers in the CG. These results are in line with previous research using different types of questions (e.g., Graesser & Lehman, 2011; Taboada & Guthrie, 2006).

Personal Variables

Furthermore, we found that personal variables that were shown to be relevant in comprehending and learning from texts—such as the initial level of comprehension and prior knowledge about the subject—had an impact on students' ability to answer all types of questions (except literal questions in the follow-up test) and the learning test. These results provide additional evidence of the importance of these two variables in promoting learning from texts when the tasks involve diverse levels of comprehension (Alexander & Jetton, 2000; Gil, Bråten, Vidal-Abarca, & Strømsø, 2010; Kintsch, 1988, 1998; Ozuru, Dempsey, & McNamara, 2009; Taboada & Guthrie, 2006). These variables were included in the models to assess whether they interacted with other variables such as condition. The fact that we did not find any interaction between prior knowledge, level of comprehension, and condition shows that the intervention had a similar effect on all students, regardless of their personal differences in reading comprehension ability. Hence, our study achieved one of the challenges that Stadtler, Bromme, and Rouet (2018) identified: to embed the instruction of multiple text reading skills in school education and, at the same time, strengthen students' basic reading abilities.

Scope and Limitations

Presumably, the broad, sustainable results that students achieved might be attributable to specific aspects of the intervention. First, the intervention was comprehensive in the sense that it targeted reading strategies quite broadly, from understanding the text or texts to understanding the questions fully. It also focused on the processes required to answer the questions by searching for, selecting, connecting, and organizing the information. Second, these processes, which are involved in comprehending multiple documents (Barzilai et al., 2018; Rouet & Potocki, 2018), were taught by classroom teachers and integrated into authentic curriculum-based activities that were relevant to the students. Finally, the professional development program, based on a gradual release of responsibility (Fisher & Frey, 2013; Pearson & Gallagher, 1983), allowed teachers to experience ways of scaffolding their students and modeling strategies to answer questions. This made the teachers more autonomous when they implemented the tasks.

However, an assessment of the scope of our results reveals a number of aspects that need to be considered. When we designed our study, we assumed that an intervention focusing on teachers' classroom practice could influence students' learning outcomes. From a conceptual perspective, one might deduce that we applied a pure *process-outcome* paradigm, wherein something that teachers did would produce a change in student learning. However, we could argue that although input from teachers is an essential ingredient, how students respond to this input is also a key factor in determining whether teaching leads to better and longer lasting learning outcomes. From this perspective, the design of our study should be considered as a *process* (more explicit teaching and guidance)–*process* (more strategic and rule-based learning)–*outcome* paradigm.

From a methodological perspective, for some of the variables related to the types of questions, the internal consistency reliability was relatively low, although sufficient for scales with fewer than 10 items (Loewenthal, 2001). Bearing this limitation in mind, in terms of the studied variables, apart from prior knowledge and the general level of reading comprehension, we could have also considered many other factors that are involved in solving tasks with multiple documents. For instance, we could have included other personal and contextual variables that, as the RESOLV model illustrates (Britt et al., 2017), also have an impact (e.g., task representation, working memory capacity, emotions, etc.) on the way students solve a task. However, including other variables would have required a different methodological approach, which would have also had repercussions on the ecological validity.

The complexity of this intervention makes it impossible to attribute its effects to a single or specific components. Notwithstanding this limitation, the fact that teachers implement the strategies in real classroom contexts seems ecologically and pedagogically more meaningful than isolating specific components of the intervention and trying them separately. Additionally, our study shows

that it is possible for research findings to reach classrooms and to be successfully applied by teachers themselves through their participation in PD programs designed for such purposes. As the professional development program involved a reflection on and analysis of classroom practice, the teachers could draw up their plans considering insights from the program. We believe that the following characteristic features of high-quality professional development, in addition to the aforementioned gradual release of responsibility, were crucial components of the results (Anders et al., 2000; Dillon et al., 2011): the professional development focused on students' learning outcomes; teachers were given intensive support and opportunities to work together; the professional development promoted systematic reflection on classroom practice and modeled how to teach strategies; and it fostered sustained and voluntary collaboration between teachers and researchers on a content area of relevance to them.

Educational Implications

The results of this study have several implications for educational practice. First, they confirm that students benefit from explicit instruction aimed at teaching them how to read texts in a way that promotes understanding and learning, rather than merely identifying and repeating information, and that this instruction can be delivered in the context where it matters the most, that is, within the subject area to which a text pertains (Shanahan & Shanahan, 2008).

Second, it is important to work on strategies such as those required for answering inferential questions that may appear basic or to have already been addressed as part of primary school education. Students are often assigned tasks requiring in-depth processing of texts, and teachers may assume that their students already have the skills required to do so. However, international assessments and research have demonstrated that even with regard to relatively simple strategies such as answering different types of questions, there remains considerable room for improvement. Our study reinforces this conclusion.

Third, for students to become lifelong learners in formal and informal contexts and to function adequately in society, they need to learn to process information from multiple sources and transform it into knowledge (Barzilai et al., 2018). Our results highlight the fact that the skills required for answering inferential questions from multiple texts are not the same as those needed to answer inferential questions from a single text. Therefore, specific instruction on this more complex reading task is required. Giving students meaningful intertextual tasks and teaching them ways of approaching the tasks are essential in enabling them to become independent learners.

Finally, none of these improvements are possible without the support of teachers who firmly believe that effective instruction in a given subject area implies not only the transmission of concepts, theories, and methods but also the teaching of specific ways of accessing information, including the ability to process written information. Although many teachers do share this belief, it is not always reflected in their classroom practice. Involving teachers in reflective professional development programs that aim to analyze and improve their practice is a costly and demanding process that requires sustained support to achieve change. However, this endeavor is a crucial part of equipping teachers to provide students with the education they deserve.

Conflict of Interest Disclosure

The authors declare that they have no conflict of interest.

Ethical Approval Statement

All procedures followed were in accordance with the ethical standards of the institutional committee on human experimentation (University of Barcelona, Spain) and with the 1964 Helsinki declaration and its later amendments. Informed consent from legally authorized parents/or guardians and from the same students was obtained for being included in the study.

Permission to Reproduce Material from other Sources

The authors declare that they do not reproduce material from other sources in this work.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author [N. Castells]. The data are not publicly available due to their containing information that could compromise research participant consent.

NOTE

REFERENCES

Afflerbach, P., Hurt, M., & Cho, B. Y. (2020). Reading comprehension strategy instruction. In D. L. Dinsmore, L. K. Fryer, & M. M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 99–118). Routledge.

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 285–310). Lawrence Erlbaum Associates.

Alvermann, D. E., Swafford, J., & Montero, K. M. (2004). *Content area literacy instruction for the elementary grades.* Allyn & Bacon.

Anders, P. L., Hoffman, J. V., & Duffy, G. G. (2000). Teaching teachers to teach reading: Paradigm shifts, persistent problems, and challenges. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 719–742). Lawrence Erlbaum Associates.

Anderson, L., & Krathwohl, D. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.

Armbruster, B. B., Anderson, T. H., Armstrong, J. O., Wise, M. A., Janish, C., & Meyer, L. A. (1991). Reading and questioning in content area lessons. *Journal of Literacy Research*, *23*(1), 35–59. https://doi.org/10.1080/10862969109547726

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*, 973–999. 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

Beck, I. L., McKeown, M. G., & Gromoll, E. W. (1989). Learning from social studies texts. *Cognition and Instruction*, *6*, 99–158. https://doi.org/10.1207/s1532690xci0602_1

Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy.* Academic Press.

Braaksma, M. A. H., Rijlaarsdam, G., & van den Bergh, H. (2018). Effects of hypertext writing and observational learning on content knowledge acquisition, self-efficacy, and text quality: Two experimental studies exploring aptitude treatment interactions. *Journal of Writing Research*, *9*, 259–300. https://doi.org/10.17239/jowr-2018.09.03.02

Braasch, J. L., McCabe, R. M., & Daniel, F. (2016). Content integration across multiple documents reduces memory for sources. *Reading and Writing*, *29*, 1571–1598. https://doi.org/10.1007/s11145-015-9609-5

Bråten, I., Brante, E. W., & Strømsø, H. I. (2019). Teaching sourcing in upper secondary school: A comprehensive sourcing intervention with follow-up data. *Reading Research Quarterly*, *54*, 481–505. https://doi.org/10.1002/rrq.253

Bråten, I., Ferguson, L. E., Anmarkrud, Ø., & Strømsø, H. I. (2013). Prediction of learning and comprehension when adolescents read multiple texts: The roles of word-level processing, strategic approach, and reading motivation. *Reading and Writing*, *26*, 321–348. https://doi.org/10.1007/s11145-012-9371-x

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. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276–314). Cambridge University Press. https://doi.org/10.1017/CBO9781139048224.017

Britt, M. A., Rouet, J. F., & Braasch, J. L. (2013). Documents as entities. In M. A. Britt, S. R. Goldman, & J. F Rouet (Eds.), *Reading: From words to multiple texts* (pp. 160–79). Routledge.

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

Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology*, *25*, 313–339. https://doi.org/10.1080/02702710490522658

Cantrell, S. C., Almasi, J. F., Carter, J. C., Rintamaa, M., & Madden, A. (2010). The impact of a strategy-based intervention on the comprehension and strategy use of struggling adolescent readers. *Journal of Educational Psychology*, *102*, 257–280. https://doi.org/10.1037/a0018212

Castells, N. (2019). Leer para aprender a partir de un texto [Reading to learn from one text]. In I. Solé (Coord.), *comprender y aprender. Propuestas para favorecer el aprendizaje a partir de textos [Reading, comprehending and learning. Proposals to promote learning from texts]* (pp. 73–96). Horsori.

Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences*, *21*, 201–205. https://doi.org/10.1016/j.lindif.2010.11.007

Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2013). Self-generated explanations on the question demands are not always helpful. *The Spanish Journal of Psychology*, *16*, E26. https://doi.org/10.1017/sjp.2013.45

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*, 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*, 13–27. https://doi.org/10.1016/j.learninstruc.2007.12.003

Cho, B. Y., Afflerbach, P., & Han, H. (2018). Strategic processing in accessing, comprehending, and using multiple sources online. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 133–150). Routledge.

Coll, C. (1990). Un marco de referencia psicológico para la educación escolar: La concepción constructivista del aprendizaje y la enseñanza [A psychological framework for school education]. In J. Palacios, A. Marchesi, & C. Coll (Comps.), *Desarrollo psicológico y educación. Psicología de la Educación [Psychologycal development and education. Educational Psychology]* (pp. 435–454). Alianza.

Daines, D. (1986). Are teachers asking higher-level questions? *Education*, *106*, 368–374.

Dillon, D. R., O'Brien, D. G., Sato, M., & Kelly, C. M. (2011). Professional development and teacher education for reading instruction. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. B. Afflerbach (Eds.), *Handbook of reading research* (pp. 629–660). Routledge.

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

Durkin, D. (1978). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, *14*, 481–533.

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology*, *104*, 515–528. https://doi.org/10.1037/a0027182

Elleman, A. M. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review. *Journal of Educational Psychology*, *109*, 761–781. https://doi.org/10.1037/edu0000180

Fernández-Huerta, J. (1959). Medidas sencillas de lecturabilidad [Simple measures for readability]. *Consigna*, *214*, 29–32.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. SAGE Publications.

Firetto, C. M. (2020). Learning from multiple complementary perspectives. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (pp. 223–244). Routledge.

Firetto, C. M., & Van Meter, P. N. (2018). Inspiring integration in college students reading multiple biology texts. *Learning and Individual Differences*, *65*, 123–134. https://doi.org/10.1016/j.lindif.2018.05.011

Fisher, D., & Frey, N. (2013). *Better learning through structured teaching: A framework for the gradual release of responsibility*. ASCD.

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*, 157–173. https://doi.org/10.1016/j.cedpsych.2009.11.002

Goldman, S. R., & Durán, R. P. (1988). Answering questions from oceanography texts: Learner, task and text characteristics. *Discourse Processes*, *11*, 373–412. https://doi.org/10.1080/01638538809544710

Goldman, S. R., & Scardamalia, M. (2013). Managing, understanding, applying, and creating knowledge in the information age: Next-generation challenges and opportunities. *Cognition and Instruction*, *31*, 255–269. https://doi.org/10.1080/10824669.2013.773217

González-Lamas, J., Mateos, M., & Cuevas, I. (2016). Arguing from sources: Design and evaluation of a program to improve written argumentation and its impact according to students' writing beliefs. *Infancia y Aprendizaje*, *39*, 49–83. https://doi.org/10.1080/02103702.2015.1111606

Graesser, A., & Lehman, B. (2011). Questions drive comprehension of text and multimedia. In M. McCrudden, J. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 54–74). Information Age.

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.

Halpain, D. R., Glover, J. A., & Harvey, A. L. (1985). Differential effects of higher and lower order questions: Attention hypotheses. *Journal of Educational Psychology*, *77*, 703–715. https://doi.org/10.1037/0022-0663.77.6.703

Hodges, C. A. (1980). Commentary: Toward a broader definition of comprehension instruction. *Reading Research Quarterly*, *15*, 299–306.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy. *Education Policy Analysis Archives*, *13*(10), 1–28. https://doi.org/10.14507/epaa.v13n10.2005

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations*. Holt, Rinehart & Winston.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182. https://doi.org/10.1037/0033-295X.95.2.163

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

List, A. (2020). Six questions regarding strategy use when learning from multiple texts. In D. L. Dinsmore, L. K. Fryer, & M. M. Parkinson (Eds.), *Handbook of strategies and strategic processing* (pp. 119–140). Routledge.

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

List, A., Du, H., & Lee, H. Y. (2021). How do students integrate multiple texts? An investigation of top-down processing. *European Journal of Psychology of Education*, *36*, 599–626. https://doi.org/10.1007/s10212-020-00497-y

List, A., Grossnickle, E. M., & Alexander, P. A. (2016). Profiling students' multiple source use by question type. *Reading Psychology*, *37*, 753–797. https://doi.org/10.1080/02702711.2015.1111962

Loewenthal, K. M. (2001). *An introduction to psychological tests and scales*. Psychology Press.

Magnusson, C. G., Roe, A., & Blikstad-Balas, M. (2019). To what extent and how are reading comprehension strategies part of language arts instruction? A study of lower secondary classrooms. *Reading Research Quarterly*, *54*, 187–212. https://doi.org/10.1002/rrq.231

Mason, L. H., Dunn Davison, M., Hammer, C. S., Miller, C. A., & Glutting, J. (2013). Knowledge, writing, and language outcomes for a reading comprehension and writing intervention. *Reading and Writing*, *26*, 1133–1158. https://doi.org/10.1007/s11145-012-9409-0

McKeown, M. G., Beck, I. L., & Blake, R. G. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, *44*, 218–253. https://doi.org/10.1598/RRQ.44.3.1

McMaster, K. L., Van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., Bohn-Gettler, C. M., & Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, *22*, 100–111. https://doi.org/10.1016/j.lindif.2011.11.017

Miras, M. (2019). Leer para aprender integrando varias fuentes [Reading to learn by integrating several sources]. In I. Solé (Coord.), *Leer, comprender y aprender. Propuestas para favorecer el aprendizaje a partir de textos [Reading, comprehending and learning. Proposals to promote learning from texts]* (pp. 123–146). Horsori.

Mosenthal, P. B. (1996). Understanding the strategies of document literacy and their conditions of use. *Journal of Educational Psychology*, *88*, 314–332. https://doi.org/10.1037/0022-0663.88.2.314

Muijselaar, M. M., Swart, N. M., Steenbeek-Planting, E. G., Droop, M., Verhoeven, L., & de Jong, P. F. (2017). The dimensions of reading comprehension in Dutch children: Is differentiation by text and question type necessary? *Journal of Educational Psychology*, *109*, 70–83. https://doi.org/10.1037/edu0000120

Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology*, *101*, 740–764. https://doi.org/10.1037/0022-0663.88.2.314

Nadal, E. (2021). *La síntesi escrita com a eina d'aprenentatge a l'educació secundària: una proposta d'intervenció [Written synthesis as a learning tool in secondary education: an intervention proposal]*. (Unpublished doctoral dissertation). University of Barcelona.

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.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. The National Institute of Child Health and Human Development.

Ness, M. (2011). Explicit reading comprehension instruction in elementary classrooms: Teachers' use of reading comprehension strategies. *Journal of Research in Childhood Education*, *25*, 98–117. https://doi.org/10.1080/02568543.2010.531076

OECD. (2019). *PISA 2018 Reading framework*. Retrieved from https://www.oecd-ilibrary.org/sites/5c07e4f1-en/index.html?itemId=/content/component/5c07e4f1-en

Occelli, M., & Valeiras, N. (2013). Los libros de texto de ciencias como objeto de investigación: Una revisión bibliográfica [Science textbooks as an object of investigation: A bibliographic review]. *Enseñanza de las Ciencias*, *31*(2), 133–152. https://doi.org/10.5565/rev/ec/v31n2.761

O'Flahavan, J. F., Hartman, D. K., & Pearson, D. P. (1989). *Teacher questioning and feedback practices after the cognitive revolution: Replication and extension of Guszak's (1967) study* (Technical Report No.

461). Retrieved from Illinois Digital Environment for Access to Learning and Scholarship website: http://hdl.handle.net/2142/17957

Olson, D. O. (2015). Literacy and the technologies of knowing. In R. J. Spiro, M. DeSchryver, M. Schira Hagerman, P. M. Morsink, & P. Thompson (Eds.), *Reading at a crossroads? Disjunctures and continuities in current conceptions and practices* (pp. 21–25). Routledge.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*, 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175. https://doi.org/10.1207/s1532690xci0102_1

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, *76*, 1239–1252. https://doi.org/10.1037/0022-0663.76.6.1239

Pearson, P. D., & Gallagher, M. C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology*, *8*, 317–344. https://doi.org/10.1016/0361-476X(83)90019-X

Pearson, D. P., & Johnson, D. D. (1978). *Teaching reading comprehension*. Holt, Rinhart and Winston.

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

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology*, *102*, 299–312. https://doi.org/10.1037/a0017763

Raphael, T. E., & Au, K. H. (2005). QAR: Enhancing comprehension and test taking across grades and content areas. *The Reading Teacher*, *59*, 206–221. https://doi.org/10.1598/RT.59.3.1

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*, 181–221. https://doi.org/10.3102/00346543066002181

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

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

Rouet, J. F., & Britt, M. A. (2014). Multimedia learning from multiple documents. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 813–841). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.039

Rouet, J. F., & Potocki, A. (2018). From reading comprehension to document literacy: Learning to search for, evaluate and integrate information across texts. *Infancia y Aprendizaje*, *41*, 415–446. https://doi.org/10.1080/02103702.2018.1480313

Sáiz, J. (2011). Actividades de libros de texto de Historia, competencias básicas y destrezas cognitivas, una difícil relación: Análisis de manuales de 1º y 2º de ESO [History textbook activities, basic competencies and cognitive skills, a difficult relationship: Analysis of 7th and 8th grade manuals]. *Didáctica de las ciencias experimentales y sociales*, *25*, 37–64.

Sánchez, E. (2010). *La lectura en el aula: Qué se hace, qué se debe hacer y qué se puede hacer [Reading in the classroom: What is done, what should be done, and what can be done]*. Graó.

Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, *45*, 669–700. https://doi.org/10.3102/0002831208316955

Schön, D. A. (1987). *Educating the reflective practitioner*. Jossey-Bass.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*, 40–59. https://doi.org/10.17763/haer.78.1.v62444321p602101

Spencer, M., Gilmour, A. F., Miller, A. C., Emerson, A. M., Saha, N. M., & Cutting, L. E. (2019). Understanding the influence of text complexity and question type on reading outcomes. *Reading and Writing*, *32*, 603–637. https://doi.org/10.1007/s11145-018-9883-0

Stadtler, M., Bromme, R., & Rouet, J. F. (2018). Learning from multiple documents: How can we foster multiple document literacy skills in a sustainable way? In E. Manalo, Y. Uesaka, & C. A. Chinn (Eds.), *Routledge research in achievement and gifted education. Promoting spontaneous use of learning and reasoning strategies: Theory, research, and practice for effective transfer* (pp. 46–61). Routledge/Taylor & Francis Group.

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*, 63–90. https://doi.org/10.1177/0741088312466532

Taboada, A., & Guthrie, J. T. (2006). Contributions of student questioning and prior knowledge to construction of knowledge from reading information text. *Journal of Literacy Research*, *38*, 1–35. https://doi.org/10.1207/s15548430jlr3801_1

Vidal-Abarca, E., Gilabert, R., Martínez, T., Sellés, P., Abad, N., & Ferrer, C. (2007). *Test de Estrategias de Comprensión [Test of comprehension strategies]*. Instituto Calasanz de Ciencias de la Educación.

Vidal-Abarca, E., Mengual, V., Sanjose, V., & Rouet, J.F. (1996, September). *Levels of comprehension of scientific prose: The role of text and task variables*. Paper presented at the International Seminar on Using Complex Information Systems, Poitiers, France.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge University Press.

Wei, R.C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the US and abroad*. National Staff Development Council and the School Redesign Network at Stanford University.

Wiley, J., Steffens, B., Britt, M. A., & Griffin, T. D. (2014). Writing to learn from multiple-source inquiry activities in history. In G. Rijlaarsdam (Series Ed.), P. Klein, P. Boscolo, C. Gelati, & L. Kilpatrick (Volume Eds.), *Studies in writing, writing as a learning activity* (pp. 120–148). Emerald Group Publishing Limited.

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, *80*, 351–361.

NÚRIA CASTELLS (corresponding author) is a lecturer in the Department of Cognition, Developmental and Educational Psychology, at the University of Barcelona (Spain); email nuria.castells@ub.edu. Her work focuses on examining students' learning from multiple texts and the impact of professional development programs on the promotion of reading, writing and knowledge acquisition in Secondary and Higher Education.

MARTA MINGUELA is a lecturer in the Department of Cognition, Development and Educational Psychology at the University of Barcelona, Spain; email martaminguela@ub.edu. Her research focuses on understanding reading and writing processes, the role of metacognition in reading

comprehension, the teaching of self-regulation of the reading process, and the role of educational psychologists to improve teaching practices.

ISABEL SOLÉ is an honorary professor in the Department of Cognition, Development and Educational Psychology at the University of Barcelona (Spain); email isoleg@ub.edu. Her research focuses on reading comprehension, the processes of reading and writing and their relationship to learning, from elementary school to university.

MARIANA MIRAS is an Emeritus professor in Development and Educational Psychology at the University of Barcelona (Spain); email mariana.miras@ub.edu. Her research focuses on the teaching and learning processes of writing throughout

schooling, the use of reading and writing as learning tools, academic writing, and the cognitive and emotional aspects involved in the learning and teaching processes.

ESTHER NADAL is a lecturer in Educational Psychology at the Open University of Catalonia (UOC), Spain; email enadale@uoc.edu. Her research focuses on the use of reading and writing and learning from multiple texts at secondary education and university.

GERT RIJLAARSDAM is a professor in Innovative Language Education at the Universities of Amsterdam (NL) and Trondheim (Norway); email g.c.w.rijlaarsdam@uva.nl. His research focuses on process and intervention studies in languages, literature and arts.

APPENDIX A

Sample of prior knowledge text questions

3. The founder of Islam was:

   a) Allah

   b) **Mohammed**

   c) Abdallah

   d) The archangel Gabriel

15. During the reign of the Catholic Monarchs, the Emirate of Granada, where most of the Muslims on the peninsular lived:

   a) **Became part of the Crown of Castile**

   b) Became part of the Crown of Aragon

   c) Became part of the Kingdom of Portugal

   d) Became part of the Kingdom of Navarre

19. The symptoms of the Black plague include:

   a) The appearance of black scabs that cover the skin

   b) Intense stinging all over the body

   c) A strong smell on your breath

   d) **Shivers all over the body**

APPENDIX B

Texts used in the posttest and examples of the different types of questions

Posttest TA. Social groups and ways of life in Al-Andalus society

In 711, a Muslim army formed by approximately 60,000 Arabs and Berbers from North Africa crossed the Straits of Gibraltar and began to conquer the Visigoth kingdom of the Iberian Peninsula. Most of the Visigoths who inhabited the Peninsula at that time were Christians or Jews. As a result of the conquest, the structure of society changed through the arrival of these new groups, the Arabs and the Berbers.

The conquering army comprised far fewer Arabs than Berbers. However, following the conquest of the Peninsula, it was the Arabs who governed the territories and cities of Al-Andalus. They tended to live in the medina, the main walled quarter of a Muslim city and the

location of its markets, the central mosque, and a few residential neighborhoods. Although they communicated mainly in their mother tongue, Arabic, they also learnt Latin, which was the language most widely used in the Iberian Peninsula until that point.

By contrast, most of the Berbers, who had arrived on the Peninsula as soldiers in the Muslim army, became farmers or shepherds and lived in the countryside. Nonetheless, small numbers of them were merchants and lived in the *ravals*, the neighborhoods located outside the medina. Initially, they continued to communicate in Berber, their own language, but over time they learnt and began to use Arabic.

Despite the arrival of the Arabs and Berbers, the majority of people living in Al-Andalus (4-5 million) were still Visigoth and followed different religions, mainly Christianity and Judaism. Following the conquest, the Muslim rulers decreed that these groups could continue to practice their religions in exchange for the payment of higher taxes than were levied on the Muslim population. These taxes served to promote the development of agriculture and industry, and also contributed to the general running of the state.

Part of the Christian population decided to convert to Islam, adopting the Arabic language and customs, and in this way they avoided paying the special taxes. These Christians became known as Muladies, from an Arab word meaning "adopted." Over time, this community grew until it became the largest group within Al-Andalus society.

The rest of the Christian population, which did not convert to Islam, continued to pay extra taxes and to practice their own religion. Thus, they maintained some of their own traditions (Christmas, the Feast of St John, etc.) and also certain customs that distinguished them from the Arabs, such as drinking wine; they also used their own language, Latin, to communicate with one another. However, they also adopted certain Islamic customs related to ways of dressing, cooking, or house building, and they learnt Arabic to communicate with Muslims and to read their texts. These Christians thus became known as Mozarabs, meaning "Arabized." Despite their embracing certain aspects of Muslim culture, the fact that these Mozarabs sought to preserve the right to choose their own representatives did not sit well with the Muslim rulers. Consequently, they were persecuted and over time they either ended up converting to Islam or fleeing to Christian kingdoms to the north of the Iberian Peninsula.

As for the Jews, they were another minority that retained its own religion and customs in exchange for the payment of higher taxes. They lived in their own neighborhoods, known as the *juderia*, and maintained their own social organization under the authority of the Rabbi. They continued to use Hebrew for religious ceremonies, but spoke Arabic for everyday affairs.

The coexistence between these different social groups gave rise to a diverse and bilingual society. The Arabs learnt the language of the conquered land, while the other groups adopted the language of their conquerors, with the latter becoming the language of everyday communication in Al-Andalus.

Posttest TB. Economic activities in Al-Andalus

During the Visigoth era, the economy of the Iberian Peninsula, as well as that of the rest of Western Europe, was based on subsistence agriculture. With the arrival of Islamic civilization, agriculture continued to be the main economic activity, but it took on a different character. Production increased considerably and produce was now sold in addition to being used for own consumption. At the same time, cities came to play a fundamental role as centers of commerce and artisan production.

With regard to agriculture, after the conquest the Arabs redistributed the most productive and fertile land of the Visigoths. They were now the landowners, whereas the Berbers worked the land in exchange for part of the harvest, just as most of the Muladies and Mozarabs did, since the majority of them lived in the countryside. Cereals, vines, and olives continued to be the most important crops, although the Muslims also introduced new kinds of crops that are still cultivated to this day, for example, rice, eggplants, artichokes, citrus fruits, peaches, and apricots for human consumption, as well as plants such as cotton and indigo for the textile industry.

New irrigation techniques were also developed. Ditches (canals) and *norias* (water wheels worked by animals in order to draw underground water) were built and enabled water to be transported and distributed to areas where it was scarce, thus improving the productivity of the land.

In the cities, artisan products of high quality began to be produced, most notably with regard to paper, ceramics, leather goods, gold and silverware, weaponry, textile dyes, and woven silks. The Muslims also created gardens that served to conduct botanical experiments. Using new techniques, they developed perfumes and balsams, as well as insecticides to deal with crop pests and syrups and narcotic substances that physicians could use to relieve the pain suffered by patients during operations. The Arabs controlled a large part of this production, with most of the actual work being done by Mozarabs and Jews.

The city was also the place of trade, with both the Muladies and Jews working as merchants. In the souk (the city market), one could find an enormous variety of products—both agricultural and artisan—from across the Islamic world. Trade between members of the Muslim community in the Iberian Peninsula was mainly conducted in two currencies: the gold dinar and the silver dirham. As artisan production was scarce in the rest of Western Europe, the Muslims also sold their goods to other Christian kingdoms in exchange for gold or slaves.

Read the texts about Al-Andalus society and answer the following questions:

Sample literal question: Who was living in the Iberian Peninsula when the Muslims conquered it?

Sample intratextual inferential question: Why did Arabic end up being the most widely used language in Al-Andalus?

Sample intertextual inferential question: What similarities and differences existed between the Muladies and the Mozarabs?

Reliability analyses for the different type of questions per moment

The reliability for each type of question per test moment using Cronbach's alpha corrected for test length using the Spearman–Brown extension and Pearson correlations is shown in Tables C1 and C2.

TABLE C1
Cronbach's Alphas for Types of Comprehension Question on Pretest (T1), Posttest (T2), and Follow-up Test (T3)

	T1	T2	T3
Literal questions	.65	.60	.71
Intratextual questions	.75	.60	.70
Intertextual questions	.81	.82	.70

TABLE C2
Pearson Correlations between Comprehension Question Results on Pretest (T1), Posttest (T2), and Follow-up Test (T3)

	T1		T2		T3	
	Intra. questions	Inter. questions	Intra. questions	Inter. questions	Intra. questions	Inter. Questions
Literal questions	.193**	.239**	.347**	.217**	.274**	.309**
Intratextual questions		.518**		.373**		.251**

Note. ** $p < .01$

Sample of learning test questions

4. The term *hegira* means:

a) The start of the Islamic calendar

b) **The fleeing of the founder of Islam due to the persecution to which he was subjected**

c) The pilgrimage to Mecca once in your life

d) The fast during Ramadan

7. Religion played a relevant role in the rapid expansion of Islam. The causes of this expansion include:

a) The fact that, when they found out about it, believers found a religion that was easier to obey

b) The fact that it became popular among all of those who did not have a religion

c) **The fact that in conquered regions, Muslims paid less taxes**

d) The fact that the prayers were more simple

Professional development sessions delivered to teachers in the intervention group

The first seminar started with a presentation and discussion of the purpose of the project, namely, to diversify the kind of questions that teachers ask their students about texts and to improve the way they help students to solve them. Then teachers were provided with the first pair of texts for the teaching unit (Texts 1 and 2, see Table 3) and were asked to individually plan questions they might ask their students. The pool of teachers' initial questions were collected for its use in the following seminar and as a control measure. The first dossier (reading strategies) was introduced to them and given to them as an assignment for the next seminar.

In the second seminar, teachers were introduced to the kinds of questions that students can be asked to foster their comprehension: literal, intratextual inferential, and intertextual inferential. After this presentation, teachers were asked to analyze, in small groups, the questions they had proposed in their initial pool of questions for texts 1 and 2. Once they had identified the kinds of questions they had produced in the previous session according to the classification we had presented to them, we also discussed issues related to how questions can be produced, always bearing in mind the main objectives of the lesson. The teachers noted the importance of reading the texts to find out the most relevant ideas, identifying the kinds of inferences they would like students to make, and thinking about possible connections that might be of relevance in order to better understand the texts. We showed them the questions we had designed for the pair of texts and compared these questions with theirs. At the same time, they identified and discussed the difficulties in answering the questions and possible ways to deal with these difficulties in the classroom. For instance, they noted the importance of teaching students to understand the wording of the questions fully before looking for the answer in the texts; teaching them to reread and organize the information from the two texts in order to identify possible related ideas; and ensuring that their answers were complete and well organized. Then, we discussed and agreed with the teachers on using specific ways of presenting the texts and enhancing students' comprehension of them, to help students give better and more complete answers to both literal and inferential questions. For instance, when asking their students oral questions, they became aware of certain questioning techniques such as giving students sufficient time to respond; paraphrasing incomplete responses; following up with additional questions; and asking open questions (with "why" and "how") to find out how the students had arrived at a particular response or to discover a misunderstanding. In addition, we modeled the reading strategies introduced during the session. At the end of this seminar, teachers had a general idea of the most relevant strategies to use to enable students to answer questions of varying complexity. Teachers also had an initial general plan to work with the first pair of texts and the questions to ask their students once they started the teaching unit (see Table E1). This general plan was reviewed in session #4. They were also assigned to reread Dossier 2 (types of questions that can be formulated to promote students' comprehension).

In the third seminar, the topic of the Dossier 3 (general and specific characteristics of the process of reading from multiple sources) was presented and discussed with the teachers in light of the different types of questions that had been introduced in the previous seminar. In the second half of the session, teachers were shown the second pair of texts (T3 and T4 in Table 3) and were asked to develop different types of questions on these texts in small groups. After sharing the questions they had created and

through guided practice, the seminar ended with agreement on the questions (literal, intratextual, and intertextual inferential) to be used in the classroom. We had already designed specific questions of different complexity for each pair of texts. However, since a professional development program usually aims at helping teachers to develop new competences, it was important to encourage them to develop their own literal and inferential questions (both intra- and intertextual) for the texts that were going to be used in the intervention in the classroom context. Thus, the final questions that were agreed with the entire group to work with texts 3 and 4 in their classrooms, were questions produced by the teachers that coincided, or were remarkably similar to our pool of questions. For the next seminar, teachers were asked to reread Dossier 3 and plan a teaching unit using what had been discussed during the seminars.

In the fourth seminar, the difficulties encountered in planning the sessions for the teaching unit were discussed with the teachers. They made changes to their plans after they had reflected on the content addressed during the seminars. In this context, we discussed and demonstrated again the following strategies that might be useful:

- Having students begin with careful, silent reading of the texts.
- Clarifying concepts and responding to student queries about the meaning(s) of the texts.
- Encouraging students to analyze and verify their understanding of each question before answering—attending to the wording of the questions.
- Identifying the ideas that needed to be reflected in an answer, which are more complex for inferential than for literal questions.
- Making predictions based on the text(s).
- Making inferences based on the information contained in one or both texts in a pair. This included selecting information based on the wording of the question; connecting information with the most appropriate connectors (between paragraphs in one text, or between paragraphs from various texts) and organizing the information (selected from more than one paragraph or more than one text).
- Motivating students to write clear, full, and well-structured answers to each question.
- Paraphrasing the information contained in the text(s).
- Modeling all the previous strategies to make the thinking of the teacher explicit when different types of questions are solved.

This revision allowed the teachers to plan a detailed script to be followed during the implementation of each pair of texts (see Table E1).

This plan included starting the sessions by stating the purposes of the activity, followed by activating prior knowledge about the texts that were going to be read and making some predictions. Then students would read silently and identify unknown words. This reading would be followed by reading aloud in the group class and solving doubts, and teachers asking some questions orally. After the information had been clarified, students would start reading the second text of a pair following the same procedure. In the next session, the teacher would start by asking what had been done in the previous session and then rereading the text aloud to clarify the meanings. At the end of this activity, the teacher would present all the questions and, for the first pair of texts, model how students should proceed to answer some of them. This would allow students to solve the other questions in pairs. In the same session, they would probably review some of the answers and the teacher would help the students select, connect, and organize the information from one or more texts. This sequence would be similar with the other pairs of texts, which would allow more independent work by the students as they made progress in answering the questions by themselves.

After discussing the general script, teachers prepared a new plan for working with the pair of texts 5 and 6 (see Table 3) in their respective classrooms, and they also designed the questions they would ask. The seminar ended with an agreement on how the teachers would work with these texts in the classroom and the questions they would ask their students, following the same procedure as that of the previous seminars. In this session, we also introduced the checklist they were asked to use to register what they do in the sessions of the teaching unit, and showed the teachers how it had to be filled each time they worked with the content of the texts.

After the fourth seminar, teachers started to implement the teaching unit, starting with texts 1 and 2. Following the instructions and decisions in the agreed script, the fifth and sixth seminars were mostly devoted to reflecting on and discussing the difficulties encountered, if any, and finding ways to solve them.

During the second half of the sixth seminar, teachers were asked to individually plan questions they could ask their students considering the final pair of texts. Since the final pair of texts was going to be used for the posttest, we had already planned the questions for the students. Since most of the questions proposed by the teachers were similar or close to ours, we could easily come to an agreement with them to use ours for the posttest.

A summary of the professional development seminars is provided in Table E2.

TABLE E1
Suggested Lesson Plan for the First Pair of Texts

First session	
Steps	**Examples of what to say and ask**
1) Present each of the two texts, reading the titles and proposing a general aim that justifies reading them and helps students to activate their previous knowledge.	*We will learn more about Islam and about ourselves, because in our culture there are many traces of Muslim heritage. We will be able to review what we know and perhaps even the stereotypical views we have of Islam through what we are learning.* *A thorough understanding of the subject will help us get to know other peoples better and will favor the development of mutual respect.* *What do you know about Islam? In what period were the Muslims in the peninsula? What is their influence in our cultural inheritance?*
2) Encourage students to read each of the texts silent. Before reading, provide students with some specific clues regarding the text they are going to read. Tell the students that as they read they should try to identify things they do not understand and then discuss them.	*Now that we will be working on the subject of Islam, these first two texts will help us understand its origins, where it appeared, its basic precepts, and what was happening in the world when it appeared. Islam is a religion; the texts will help us to see what other religions were influential at the time it first emerged, and it is worth exploring whether there are any relationships between these religions.* *Which other religions do you think were present at the time? What would the precepts of Islam have been at that time?*
3) After the first silent reading of the texts, read the first text aloud and explain any difficulties. Then, follow the same procedure with the second text. Try to get students to think about the connections between the two texts.	*Do the texts talk about the same thing? Are there any ideas that appear in both texts? Which ones? Do they deal with similar themes?* *Does what is said in one text help to understand or complement what is said in the other? How?*
Second session	
Steps	**Examples of what to say and ask**
1) Start the session by outlining again the purpose of the readings and asking the students what they remember about the texts read in the previous session. This should be a group activity involving everyone.	*What were we reading yesterday? What were the texts about?*
2) Provide students with the questions they will have to answer in writing. The questions should combine both inferential (intra- and intertextual) and literal questions, so as not to divert students' attention to the search for details, and direct them toward the goals of the lesson. To help students with their answers, show them how to answer a question of each type (literal, intra-, and intertextual inferential), and discuss with them how they would do it. Read the first question aloud (e.g., What was the coexistence like between Muslims and the inhabitants of the Iberian Peninsula?), and ask students what they think they should look for in the texts and where they think the answer might be found. For the intertextual inferential questions, provide students with some specific clues that help students to see the importance of performing a careful analysis of the texts. Take notes of the students' responses and write their answers on the digital screen; together, discuss how complete the answers are.	*What is this question asking?* *Where do you think we can find the answer? In which one of the texts? Is there relevant information in both texts?* *What would be an appropriate answer?* *How would you start the sentence? Would it be useful to repeat the wording of the question?* *Should we add anything else? What else? Tell me.* *What other information or ideas from the texts have no relation to the question, or only a slight relation?* *What other words could we use instead of…* *If the answer is not literally presented in the texts, how can we express it?* *Let's read the answer. Is it complete enough? Does it make sense?* *I remember that there was information on this topic in both texts; do you remember where this information was?* *I will reread it to see if I can find it; here it says that … and here it says that …* *What can I do so as not to forget what is said in the texts? I'll write down the information because this will help me to answer the question.*
3) After showing how to answer a question of each type, ask students to answer the other questions in pairs. Supervise how the pairs answer the questions, identify possible difficulties they experience, and show how to solve them.	*Now that we have done the three different types of questions together, it is your turn to answer the questions. You will do this task in pairs and you can come back to me if you have any problems.*
4) Go over the answers all together. Put up some examples on the digital screen to help students reflect on the selection, connection, and organization of the information.	*Is it essential to include this information? Could we find something else to say? What would be the most relevant idea? Are these ideas connected? How? What would you say first? How would you connect the two ideas?*

TABLE E2
Professional Development Seminars

Professional development seminars	Content of the seminar	Materials used	Assignment for the next seminar
Seminar 1	Presentation of the professional development program and its main objective, that is, improving the kinds of questions that teachers ask their students about texts. Getting to know each other: Teachers introduce themselves to the group. Pool of teachers' initial questions: Teachers are given time to plan the kind of questions they might ask their students regarding the first pair of texts in the teaching unit. Brief introduction to basic reading strategies.	Dossier 1: Reading strategies. Texts 1 and 2	Review Dossier 1 and identify potential difficulties with understanding its content and using it in the classroom, for discussion during the next session.
Seminar 2	Identifying the different kinds of questions that students can be asked so as to foster their comprehension: literal, intratextual inferential, and intertextual inferential questions. Analysis of the questions that teachers had proposed in their initial pool of questions and their potential to promote better comprehension; agreement on the questions to be used in the classroom. Modeling of different questioning techniques.	Dossier 2: Types of questions that can be formulated to promote student comprehension	Review Dossier 2 and identify potential difficulties with understanding its content and using it in the classroom, for discussion during the next session.
Seminar 3	Discussion of the importance of helping students to deal with multiple texts. Designing different types of questions for a pair of texts that present complementary information (guided practice, ending with agreement on the questions to be used in the classroom).	Dossier 3: General and specific characteristics of the process of reading from multiple sources Texts 3 and 4	Review Dossier 3 and identify potential difficulties with understanding its content and using it in the classroom, for discussion during the next session. Apply the content of the previous sessions, designing different types of questions for a new pair of texts. General planning of the teaching unit using what they have learned in the training program.
Seminar 4	Reflection/discussion on the difficulties encountered when teachers completed their assignment from the previous seminar and when designing different types of questions for a new pair of texts. Agreement on the general planning and script of their teaching unit. Discussion of teachers' proposals for working with texts 5 and 6 in the classroom, and of a proposal by the researcher that tries to incorporate some of the teachers' ideas. Agreement on how the teachers will work with these texts in the classroom and the questions they will ask their students.	Dossiers 1, 2, and 3 Texts 5 and 6	Begin the teaching unit with students (starting with texts 3 and 4). Fill in the checklist after each class.

(continued)

TABLE E2
Professional Development Seminars (*continued*)

Professional development seminars	Content of the seminar	Materials used	Assignment for the next seminar
Seminar 5	Reflection/discussion on the difficulties encountered, if any, when implementing the content of the training program in their classrooms, and how to solve them. Discussion of other ways of working with two texts in the classroom and proposing different kinds of questions to their students.	Dossiers 1, 2, and 3	Instruct the teaching unit as agreed during the training. Fill in the checklist after each class.
Seminar 6	Reflection/discussion on the difficulties encountered, if any, when implementing the content of the professional development program in their classrooms, and how to solve them. Pool of teachers' final questions. Open discussion of final teacher queries or remarks. Evaluation of the training program.	Texts: Posttests TA and TB (students' posttest texts)	Instruct the teaching unit. Fill in the checklist after each class.

APPENDIX F

Analysis and results of teachers' pool of questions

The pool of teachers' initial questions was subjected to content analysis to determine whether they proposed literal or inferential (intra- or intertextual) questions. Once the classification criteria had been established, two researchers independently carried out the analysis, shared their conclusions, and resolved any disagreements by consensus. The independent-scoring-based kappa index for interrater agreement ranged from .88 to 1. The vast majority of the questions offered by teachers in both groups were literal (63.5%), followed by intratextual inferential questions (31.4%). Only 5.1% were intertextual.

A Mann–Whitney U test found no statistically significant differences between the IG and CG teachers regarding the type and number of questions prepared for the first pair of texts before the professional development program (see Table F1).

The question sets used on the final pair of texts by the teachers in the IG and the CG revealed significant differences for the intertextual questions ($t = 2.739$; $p = .021$; IG $M = 1$; CG $M = 0$). This showed that the CG teachers did not use this type of questions with their students. We also calculated the mean overall time that teachers in both groups dedicated to working on each pair of texts. We found no significant difference between the two groups (around 3 hours for each pair of texts; $U = 7.5$; $p = .55$).

TABLE F1
Means, Standard Deviations, and Mann—Whitney U test Results for the Questions Prepared by the Teachers in the Two Groups

Type of question	Control group M (SD)	Intervention group M (SD)	Comparison of means
Literal	13.66 (11.52)	9.50 (4.97)	$U = 16.50$, $p = .81$
Intratextual inferential	3.83 (2.71)	5.33 (3.20)	$U = 12.00$, $p = .39$
Intertextual inferential	1 (.89)	1 (1.26)	$U = 17.00$, $p = .93$
Total	19.33 (12.87)	16.83 (8.40)	$U = 15.50$, $p = .69$

Teacher log and checklist of activities and IG teachers' adherence to the intervention

The teacher log and checklist included:

In relation to the Teaching Unit of Islam, we ask you:

a) to fill in the "registration" sheet for each session you conduct in the Teaching Unit, indicating the activities that you carried out, the sequence in which you carried them out, and the time spent on each activity.

b) to add, for each of the sessions, three examples (photocopied) of the tasks solved by the students (if possible, already assessed by you).

REGISTRATION SHEET

SESSION 1:

DATE OF THE SESSION: …………………….. DURATION OF THE SESSION (IN MINUTES): ……………..

TEXT / TEXTS READ: …………………………………………………………………………………….

Were they digitized or not?……………………………………………………………………………..

OTHER MATERIALS USED: ……………………………………………………………………………

1. INDICATE WHETHER YOU HAVE DONE ANY OF THE FOLLOWING ACTIVITIES, THE APPROXIMATE TIME SPENT ON EACH ONE, THE TEXT USED, AND THE ORDER IN WHICH YOU CARRIED OUT THEM OUT:

Activities	Have you done the activity? Yes/No	Time spent	Text used	Order
Oral and collective reading of the text				
Silent reading of the text				
Oral presentation by the teacher				
Resolution of doubts about specific words of the text				
Oral questions about the content/meaning of the text[1]				
Resolution of written questions about the text[2]				
Underlining of main ideas of the text				
Writing a summary of the main ideas of the text				
Discussion or debate of aspects of the content (indicate which one: …………………………. …………………………………………..)				
Watching a video, film, or documentary				
Chronological frieze reading				
Creation of chronological friezes				
Reading historical maps				
Creation of historical maps				
Search for data on the web (indicate what type of data you have been asked to search for: …………………………………………… …………………………..)				
Search for texts on the web (indicate which topic they have been asked to search for: ………………………………………. ………………………………………..)				
Carrying out short written essays on the topic				
Others (explain)				

In order to assess the degree of fidelity of the IG teachers to the planned scripts, the information in the checklist was organized for each session. The sequence of activities followed by one of the teachers in the IG and by another teacher in the CG is shown below.

IG Teacher 1: Mean time per session = 50 minutes

Title of the subtopic	Origin and principles of Islam: the historical context of its emergence		Spread of Islam in the Iberian Peninsula: Al-Andalus		Cultural contributions of Islam: the legacy of Al-Andalus		Characteristics of Al-Andalus society: economic activity and cultural exchange
Sessions	SESSION 1	SESSION2	SESSION 3	SESSION 4	SESSION 5	SESSION 6	SESSION 7
Texts used Activity order	T1 + T2		T3 + T4		T5 + T6		Posttest TA+Posttest TB
1	Silent reading of the text (T1+T2)	Oral and collective reading of the text (T1+T2)	Silent reading of the text (T3)	Oral and collective reading of the text (T3+T4)	Silent reading of the text (T5+T6)	Watching a video, film, or documentary	POSTTEST
2	Resolution of doubts about specific words of the text (T1+T2)	Resolution of written questions about the text (T1+T2)	Resolution of doubts about specific words of the text (T3)	Resolution of doubts about specific words of the text (T4)	Resolution of doubts about specific words of the text (T5+T6)	Oral and collective reading of the text (T5+T6)	
3	Oral questions about the content/ meaning of the text (T1+T2)	Reading historical maps	Silent reading of the text (T4)	Reading historical maps	Oral questions about the content/ meaning of the text (T5+T6)	Resolution of written questions about the text (T5+T6)	
4	Resolution of written questions about the text (T1+T2)		Chronological frieze reading	Resolution of written questions about the text (T3+T4)	Resolution of written questions about the text (T5+T6)		
5	*Homework: Questions (T1+T2)*		Resolution of written questions about the text (T3+T4)				

Title of the subtopic	Origin and principles of Islam: the historical context of its emergence		Spread of Islam in the Iberian Peninsula: Al-Andalus		Cultural contributions of Islam: the legacy of Al-Andalus		Characteristics of Al-Andalus society: economic activity and cultural exchange
Sessions	SESSION 1	SESSION2	SESSION 3	SESSION 4	SESSION 5	SESSION 6	SESSION 7
Texts used Activity order	T1 + T2		T3 + T4		T5 + T6		Posttest TA+Posttest TB
6			*Homework: Questions (T3+T4)*				

CG Teacher 1: Mean time per session = 50 minutes.

Title of the subtopic	Origin and principles of Islam: the historical context of its emergence		Spread of Islam in the Iberian Peninsula: Al-Andalus		Cultural contributions of Islam: the legacy of Al-Andalus		Characteristics of Al-Andalus society: economic activity and cultural exchange
Sessions	SESSION 1	SESSION 2	SESSION 3	SESSION 4	SESSION 5	SESSION 6	SESSION 7
Texts used Activity order	T1	T2	T3	T4	T5	T6	Posttest TA+Posttest TB
1	Silent reading of the text	Others: Thinking routine (3-2-1 Bridge)	Oral and collective reading of the text	Silent reading of the text	Oral and collective reading of the text	Silent reading of the text	POSTTEST
2	Underlining of main ideas of the text	Silent reading of the text	Oral presentation by the teacher	Underlining of main ideas of the text	Underlining of main ideas of the text	Resolution of doubts about specific words of the text	
3	Resolution of doubts about specific words of the text	Underlining of main ideas of the text	Underlining of main ideas of the text	Resolution of doubts about specific words of the text	Oral questions about the content/meaning of the text[1]	Underlining of main ideas of the text	
4	Resolution of written questions about the text	Resolution of doubts about specific words of the text	Resolution of doubts about specific words of the text	Oral presentation by the teacher	Others: Creation of conceptual map	Oral presentation by the teacher	
5	Reading historical maps	Discussion or debate of aspects of the content	Reading historical maps	Oral questions about the content/meaning of the text	*Homework: Finishing conceptual map*	Oral questions about the content/meaning of the text	
6	Others: brainstorming	Others: Creation of glossary	Resolution of written questions about the text	Reading historical maps		Others: Creation of collective scheme	
7		*Homework: textbook exercises*	Others: Creation of scheme	Others: kinesthetic activity: simulation game			

Title of the subtopic	Origin and principles of Islam: the historical context of its emergence		Spread of Islam in the Iberian Peninsula: Al-Andalus		Cultural contributions of Islam: the legacy of Al-Andalus		Characteristics of Al-Andalus society: economic activity and cultural exchange
Sessions	SESSION 1	SESSION 2	SESSION 3	SESSION 4	SESSION 5	SESSION 6	SESSION 7
Texts used Activity order	T1	T2	T3	T4	T5	T6	Posttest TA+Posttest TB
8			*Homework: textbook exercises*				

Once we had all the information from the checklist organized, we were able to estimate the percentage of sessions in which the teachers from both groups had been working on the "Resolution of written questions about the text" (*M* for the IG teachers = 70.14%; *M* for the CG teachers = 40.5%). In addition, we also identified the questions that formed the focus of the "Resolution of written questions about the text". IG teachers reported asking their students the questions which had been agreed during the professional development program (literal: 38.44%; intratextual inferential: 35.5%; intertextual inferential: 26.05%). The teachers in the CG posed the questions they would have asked their students normally, and the types of questions they asked were distributed as follows: literal, 70.42%; intratextual inferential, 26.63%; intertextual inferential, 2.94%. Thus, even when they asked their students to answer questions about the texts, the questions they asked were mainly literal.