DEPARTAMENTO DE PSICOLOGÍA EVOLUTIVA Y EDUCACIÓN

INTEGRATION INFORMATION PROCESSES FROM MULTIPLE DOCUMENTS

RAQUEL CERDÁN OTERO

UNIVERSITAT DE VALENCIA
Servei de Publicacions
2006

Aquesta Tesi Doctoral va ser presentada a Valencia el día 24 de Novembre de 2005 davant un tribunal format per:

- D. Manuel de Vega Rodríguez
- D. Ramiro Gilabert Pérez
- D. Jean-François Rouet
- D. Wolfang Schnotz
- D. Vicente Sanjosé López

Va ser dirigida per:
D. Eduardo Vidal-Abarca Gámez

Depòsit legal:
I.S.B.N.:84-370-6492-9
Edita: Universitat de València
Servei de Publicacions
C/ Artes Gráficas, 13 bajo
46010 València
Spain
Telèfon: 963864115

Departament de Psicologia Evolutiva
i de l'Educació

PROCESOS DE INTEGRACIÓN DE INFORMACIÓN A PARTIR DE DOCUMENTOS MÚLTIPLES

INTEGRATION INFORMATION PROCESSES FROM MULTIPLE DOCUMENTS

TESIS DOCTORAL EUROPEA PRESENTADA POR: **RAQUEL CERDÁN OTERO**

DIRECTOR: **DR. EDUARDO VIDAL-ABARCA GÁMEZ**

INDEX

1. THEORETICAL INTRODUCTION

1.1 LEARNING THROUGH QUESTIONS FROM A TEXT

Learning in school settings implies handling with expository texts, one or several on the same topic, from which learners have to acquire knowledge. The text has to be read, understood and its contents assimilated to the extent that learners are not only capable of repeting them afterwards, but can make inferences and apply this knowledge to new situations. All of this implies that students are able to read and understand the texts building an interconnected mental representation that includes textual information, inferences and other background knowledge to fill in the coherence gaps that may be present in the texts.

Thus, comprehending a text implies a construction of a mental representation in which the actions or mental processes students undertake when reading the text are crucial for the final learning outcome. These processes have been widely presented in Kintsch and van Dijk's comprehension model (Kintsch, 1998; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). This model assumes that comprehending a text involves a series of processing cycles. During each cycle, the reader acquires a small amount of information, roughly corresponding to a sentence. This involves constructing semantic propositions that underlie the meaning of the sentence, connecting the propositions through various types of links (e.g., coreference, causal and temporal relationships), and maintaining a small subset of propositions in working memory, in order to connect them to the next processing cycle. This process goes on during each subsequent cycle, allowing the reader to construct progressively an interconnected network of semantic propositions, or textbase according to Kintsch and van Dijk's (1978) terminology.

Text comprehension also involves the retrieval of knowledge from the reader's long term memory. Retrieval from long term memory is cued by the concepts and propositions encountered in the current cycle. Knowledge retrieved from LTM is integrated with text information and becomes part from the reader's representation in long term memory, or *situation model* (van Dijk & Kintsch, 1983).

Consequently, there are at least two different comprehension levels that can be reached after reading a text, depending basically on the kind of mental processes displayed when reading. If the reader sticks to the elaboration of ideas based on the text and makes few or no inferential connections among them, the mental representation that he or she would acquire would be situated at the level of the text-base. As this type of mental representation lacks of connections among ideas and few relationships between these ideas and reader's background knowledge, it would not allow the reader to reason about the new-learned contents or apply them to new situations, only literal remembering of the contents would be possible. Therefore, the level of learning derived from it could be defined as shallow.

Conversely, a reader who actively connects textual information through different types of links such as inferences and, additionally, makes connections to previous knowledge is said to have constructed a situation model from the text that enables him or her to use this knowledge for any kind of purpose, whether simple remembering, recognition or the capability to solve inferential situations based upon the contents present in the text. This level of learning from text could be defined as deep comprehension.

It would be desirable to help students construct rich interconnected situation models of the texts they normally read for learning, by promoting the nuclear processes to comprehend texts at a deep level. Some students have learned the basic strategies on how to read a text to acquire knowledge and display the main processes involved in comprehension (e.g., inference connection and background knowledge activation) almost spontaneously. Nevertheless, there are a great number of students that present difficulties in comprehending the ideas present in the texts and how these ideas are connected to each other. Most of the times the only mental representation they build after reading a text is a shallow network of textual ideas lacking of connections among themselves.

It is a responsibility of the educational system to teach and foster student´s abilities and strategies to understand texts at a deep level that allow the possibility to apply knowledge to many kinds of situations. Traditionally, different kinds of tasks have been used for this purpose, answering questions while or after reading a text is one of them. Though, questions can be very different and can promote different learning levels depending on the question-answering processes they may induce.

Generally speaking, the design of questions with instructional purposes has been aimed at fostering text comprehension and learning. In fact, there is ample evidence that adjunct questions deeply influence the processing of instructional materials (see Andre, 1979; Hamilton, 1985; Hartley & Davies, 1976; Rickards, 1979 for reviews), especially if they promote the production of inferences and/or the integration of text elements. Inferences and integration are basic processes when constructing a mental representation from text, according to Kintsch and van Dijk's comprehension model (Kintsch, 1998; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983).

Two types of instructional questions, high versus low level questions, have been distinguished attending to two dimensions, i. e., the mental processes they induced, and the amount of information they required to be answered. The combination of the two resulted in better or poorer comprehension. Thus several studies have demonstrated that working with high-level questions promoted better comprehension and learning than working with low-level questions (i.e., Vidal-Abarca, Mengual, Sanjose & Rouet 1996; Vidal-Abarca, Gilabert & Rouet, 1998). High-level questions were defined as those in which the learner had to comprehend, manipulate and connect several units of information via complex inferences. These mental activities needed for answering high -level questions directly led to better comprehension, as the learner was being encouraged to engage in the main activities involved in the construction of a richly connected mental model from text.

For instance, if a learner was asked to give some causal explanation of a fact presented in a text and the information needed to construct this explanation would need to be integrated through inferences across several locations in the text, then the learner would be answering a high-level question. This kind of question would lead to a better comprehension, as it would direct the learner to the most appropriate mental processes to construct a highly coherent mental model from the text. In contrast, if this same learner was asked to answer a very specific question in which he or she would only need to look for some specific qualities of a situation or an object located in a specific paragraph or sentence of the text, then he or she would be answering a low-level question. The learning outcome after answering low-level questions would be lower, as the only processes these kinds of questions would induce would be searching and locating information, processes which have not been proved to be very helpful in understanding and learning from a text better. Low-level questions would thus be defined as those directed at specific units of information and in which few or no inferences need be drawn. Therefore, they would promote poorer comprehension and learning, in comparison to high-level questions.

The distinction between high- vs. low-level questions is not unique. In fact, there is ample literature on the use of different types of questions in text comprehension and learning (Hartley & Davies, 1976; Andre, 1979; Rickards, 1979; Wixson, 1983; Hamilton, 1985; Langer, 1985; Goldman & Durán, 1988; Trabasso, van den Broek, & Lui, 1988; Graesser, & Franklin, 1990; Graesser; Lang, & Roberts, 1991). Specifically, Goldman and Durán (1988) identified five types of questions depending on the relationship between the question and the text and the demands made on the knowledge base. These questions varied in terms of their relation to the text and the types of processing required to answer them. In general terms, type 1, 2 and 3 questions shared their verbatim relationship to the text, but varied in the kind of text processing activities needed for answering; type 4 questions required integration across segments, and finally type 5 questions required reasoning beyond the text.

Should we try to integrate the dual dychotomy high vs. low level questions to Goldman and Durán's typology of questions, we could say that high-level questions would be equivalent to Goldman and Durán's type 4 questions. On the other hand, low-level questions would be those in which the answer can be located in specific segments of the text and can be extracted either by copying or by making minimal inferences across close sentences. In this type of question, therefore, there is always going to be a verbatim relationship between the question and the text, as in Goldman and Durán's type 1, 2 and 3 questions. In this way, the principal distinction between high- and low-level questions is the location of the answer (concentrated vs. dispersed) and the need or not of integration across segments (e.g., by summarizing, comparing and contrasting).

Following this high vs. low level questions classification, the beneficial effects derived from answering high-level questions compared to the poor results obtained after answering low-level questions have also been confirmed in recent research. Cerdán, Vidal-Abarca, Martínez, Gilabert and Gil (submitted) conducted three experiments to test the efectiveness of these two kinds of questions answered after o while reading a text dealing about Atomic Models and presented using a special software called *Read&Answer* (Martínez, 2003) which allowed us track the question-answering-behavior on-line. Students answered high or low-level questions from a text on Atomic models. Whereas low-level questions were 19, high-level questions were only 5.

However, the amount of information that needed be revised to correctly answer the questions was the same. What differed was the location of the answer (i.e, close or dispersed in the text) and the mental processes each type of question induced (i.e, connection via inferences or location processes). Results indicated that high-level questions facilitate deep comprehension but not surface recall of text, independently of having the questions after or while reading the text and that high- and low level questions promoted specific text inspection patterns and answering strategies, which varied depending on the level of success in the task, and were indicative of the different mental processes each kind of question induce.

Thus, the authors found across three experiments that high-level questions questions promote better comprehension and learning, as evidenced by long-term learning measures. The reason is that they make the student engage in additional text processing, mainly connecting ideas and establishing relationships between them, which are primarily involved in learning from text at a deep level. Additionally, the students answering high-level questions spent less time in the experimental session than those answering low-level questions. This result confirms prior research on the long-term benefits on comprehension and learning from answering high-level questions after reading a text (Vidal-Abarca, Mengual, Sanjose & Rouet 1996; Vidal-Abarca, Gilabert & Rouet, 1998).

On the contrary, high-level questions did not significantly differ from low-level questions in free recall measures. In fact, differences between the high and low-level group only appeared when the authors measured inferential comprehension and reasoning beyond the text, which assesses the construction of a coherent situation model from the text (Kintsch, 1998; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983) that goes beyond explicit textual information. Nevertheless, when using text-base measures such as free recall differences among groups were neutralized. This was an interesting result as it provided support to the main claim of the authors that high-level questions were especially effective in helping students learn from text at a deep level.

In the same vein, the authors found no significant differences in the global scores obtained as product from answering questions. That is, performance on the questions, as opposed to surface or deep post-learning measures, did not differ between the high and the low level question group but a measurement on the construction of a deep representation from text (i.e., comprehension questions) did create the expected differences. Therefore, it was not the inmediate effects of answering the questions what created significant differences between high and low level questions, nor surface learning measurements. We did find the beneficial effect of high level questions when we considered long-term learning measures, which reflected the construction of a coherent representation of text, which could only have been built if specific processing activities (i.e, inference making) would have been displayed while reading the texts to answer the questions. These processes were thus the only responsible for increases in learning from text in those answering high-level questions. And despite the efforts needed to look for the information in the text and integrate it, high-level questions proved more beneficial in the long run.

This result directly leads us to the need for making a distinction in the concepts of performance and learning and emphasizing on learning activities which might be more resource demanding when performing them but more beneficial for learning in the long term. Schmidt and Bjork (1992) published a highly interesting paper on some counterintuitive phenomena associated to the notions of training, performance and learning. Globally, they present a set of evidence showing that experimental manipulations that maximize performance during training can be detrimental in the long term and, conversely, manipulations that degrade the speed of acquisition or increase difficulty can nevertheless support the long-term goals of training.

According to these authors, learning should not be indexed by the improvements in skills across practice but be measured as the level of posttraining performance. If the level of performance in a specific domain has increased in the long term after training and there is a capability to transfer this increase to related tasks and altered contexts, then the training would have fulfilled its goals, that is, creating long-term learning. Nevertheless, learning is many times confused with performance or the inmediate success when doing the training or being under an experimental treatment aiming at increasing a specific ability.

Schmidt and Bjork (1992) find it essential to establish a distinction between the momentary strength or accesibility of a response and the underlying habit strenght of that response. In fact, many theorists have traditionally recognized that experimental variables applied during training can have two distinct kinds of effects (e.g., Estes, 1955; Guthrie, 1952; Hull, 1943; Skinner, 1938; Tolman, 1932). The first of them are the relatively permanent effects that should be called learning effects. The second group of effects is those which might be temporary effects derived of the experimental manipulations, which exaggerate or diminish performance differences while the treatment variables are operating. These performance differences normally vanish as soon as the subjects are allowed to rest or when the manipulation is removed.

Due to a misuse of these concepts, learning is erroneously assumed to refer to that set of processes occurring during the actual practice on the tasks of interest, as assessed by performance measures. To clarify these concepts and shed some light on this confusion, Schmidt and Bjork (1992) provide examples of experiments conducted both in the motor and verbal domains which clearly demonstrate that increases in performance do not necessarily correlate with increases in post-performance learning and the opposite.

Another interesting case they illustrate is that of experimental situations in which introducing difficulties for the learning during performance can enhance long-term learning. This effect can be observed in the context of text comprehension experiments. Mannes and Kintsch (1987) asked subjects to study a passage of a text, preceded by an outline that was in either the same or a different organization as the text materials. When the subjects were asked to recall the original text materials, the same organization outline was more effective.

But when the subjects were asked to do creative problem-solving tasks that required a deeper understanding of the text materials, the different organization outline was more effective.

Similarly, McNamara, Kintsch, Songer and Kintsch (1996) conducted a pair of experiments to investigate the role of text coherence in the comprehension of science texts and how it interacts with variables such as global and local text coherence, reader's background knowledge and different levels of understanding. They found that readers knowing little on the domain of the text benefit from a coherent text, whereas high-knowledge readers benefit from a minimally coherent text. The authors argue that the poorly written text forces the knowledgeable readers to engage in compensatory processing to infer unstated relations in the text. On the other hand, rewards obtained from this active processing in text were apparent at the level of deep comprehension or the situation model rather than at the superficial level of text-base understanding.

Therefore, these two examples clearly illustrate how active processing during performance and apparent increase in difficulty can promote effective long-term learning, and that this learning is mainly visible when deep learning measures are used. They also contribute to clarify the notions of on-line performance on the task and posttrainig long-term learning effects, providing evidence to the claim raised by Schmidt and Bjork (1992) that these concepts refer to different kinds of outcomes after experimental manipulation and that learning should only be considered as the final increase in an specific ability or group of abilities, that remains permanent after the removal of the experimental conditions and allows the generalisation of these abilities to other related contexts.

In accordance with these results, the beneficial effects obtained after answering high-level questions and the absence of differences with low level questions both in performance and in free recall measures (Cerdán, Vidal-Abarca, Martínez, Gilabert and Gil, submitted) can be explained through this paradigm. Therefore, students participating in these set of three experiments and answering high level questions did not experienced an apparent increase during performance. Indeed, the task seemed to be more resource demanding as it implied reading several units of information and connecting them. Despite that, it proved more effective in the long term, when deep comprehension measures were obtained. Thus, learning from text was effectively promoted by answering high level questions that made the students actively engage in the main processes involved in learning from text at a deep level, in continuity with similar findings obtained in previous research (Vidal-Abarca, Mengual, Sanjose & Rouet 1996; Vidal-Abarca, Gilabert & Rouet, 1998).

1.2 LEARNING AND INTEGRATING INFORMATION FROM MULTIPLE SOURCES

If it was desirable to promote to the maximum student's learning from text and we have concluded that specific tasks can help in increasing the level of comprehension and learning from a text, it is even more pressing the need to train and increase students' abilities and strategies to obtain information from multiple sources and integrate it, taking into account that school settings and also informal learning situations are rapidly moving towards this new way of learning through the reading and combination of different sources of information.

However, some nuclear issues regarding reading from multiple sources and integration of information should be clarified, in order to be able to establish a set of hypothesis on how to promote to the maximum learning and integration from this new perspective. First of all, we should report on how research has approached the situations in which learners extract and integrate information from several different sources and if there is a general model that accounts for the mental processes involved when combining different texts.

A growing interest on how students are capable of combining different sources of information and the strategies they apply for this purpose emerged in the context of history learning. Wineburg (1991) alerted that history teachers were commonly confronted with the need to teach by using different original or primary sources of historical evidence, yet little research had been conducted in order to guide these teachers. Because learning historical contents implies more than reaching a solution, but reconstructing an event or set of events by using different sources, Wineburg (1991) conducted an experiment using the think-aloud methodology in which he raised the question of how people construct an understanding of historical events from a group of fragmented and contradictory documents.

One of the most interesting results derived from this research was a classification, obtained from the analysis of the participants' think-alouds, of the main heuristics or strategies used by students when reading from multiple texts to learn historical contents. The first of them was the *Corroboration* heuristic, defined as the act of comparing documents and its contents with one another. The author formulated this heuristic as follows: *"whenever possible, check important details against each other before accepting them as plausible or likely"* (Wineburg, 1991). The second one was the *Sourcing* heuristic, applied to identifying the source of the document before reading the body of the text. Finally, the third heuristic was *Contextualization* or the act of situating a document in a concrete temporal and spatial context. The use of these heuristics differentiated expert historians from students and, according to Wineburg (1991), they represent a kind of *"syntactic knowledge"* that goes beyond content domains but allows the combination and use of several and different sources. In short, these heuristics have full sense in history learning, as this domain mainly involves working with different sources and the ability to correctly identify key concepts and trustworthy sources to reconstruct a set of events in time.

From this starting point, further research was conducted on students' processes and strategies when reading multiple historical sources. Thus, Rouet, Britt, Mason and Perfetti (1996) investigated if students' ability to reason with and about documentary evidence was influenced by the composition of the document set given to the students, whether primary or secondary sources. As most remarkable, they found that college students with little previous experience in history, contrary to Wineburg's findings, can learn from and reason with multiple documents. Students were able to gather different types of information on the historical controversy they were presented with and integrate it in a coherent essay. They were able to keep track of the sources of information and to refer to these sources in an appropriate way. Furthermore, studying primary documents influenced students' evaluation of documents and promted the use of references in the essays. The authors conclude that using multiple and varied documents can be an enriching activity in the history curriculum and, generally, it seems to improve students' quality of essays and help them combine information.

Rouet, Britt, Mason and Perfetti (1996) establish a clear distinction that is completely relevant for us. First of all, they insist that *reasoning about documents* is not equivalent to *reasoning with documents. Reasoning about documents* implies that, when learning from multiple sources, there may be a need to evaluate each piece of information on the basis of the type of document it is. In fact, proficiency in reasoning about documents seems to be an important element of expertise in many disciplines, especially history (Rouet, Britt, Mason & Perfetti, 1996; Wineburg, 1991), but it may be less important when working with other kinds of texts, such as expository texts, in which *who says what* is not so relevant as when dealing with documentary evidence and historical facts.

On the other hand, *reasoning with documents* would be the ability to use document information when solving a problem (Rouet, Britt, Mason & Perfetti, 1996). This activity is especially relevant when students have to write an essay based on multiple sources. When writing an essay students must refer to source information and then organize and relate this information in their essays (McGinley, 1992; Nash, Schumacher & Carlson, 1993; Spivey & King, 1989). Reasoning with documents or the exposure to multiple documents may in fact increase students' ability to engage in complex reasoning and there is indeed evidence that exposing students to multiple documents changes how they reason.

For instance, Perfetti, Britt and Georgi (1995) studied a group of college students over an 8-week period as they sequentially read a set of multiple texts describing U.S. negotiations to build the Panama Canal. Interestingly, they found that as students read the different sources over time and acquired more events and details, they engaged in more complex reasoning. They began to give more supportive reasons for their claims, more qualifiers and used longer causal chains. Therefore, it may be possible that this increase in the students' quality of reasoning was due to an exposure to multiple texts as well as due to an increase in domain-specific knowledge.

Other experiments have been conducted in order to clarify nuclear questions in working with multiple sources. For instance, wider research on the effects of discipline expertise in using documents in history (Rouet, Favart, Britt & Perfetti, 1997), or studies to analyse the effects of instructing how to identify and use source information, with successful and interesting results (Britt & Aglinskas, 2002). From this last study, where are particularly interested in the final result of the training. Thus, students trained in identifying source information and working with multiple documents using a computer-based environment were those who wrote essays on the topic of the texts that were more integrated.

In reaching this point, we are prepared to present a proposal for a general model on how students integrate historical information from multiple sources and represent it in memory. This model (Rouet, Britt, Mason & Perfetti, 1996; Perfetti, Rouet & Britt, 1999) points out that understanding multiple sources would involve the same levels of representation as understanding single passages. For each document, the reader has to understand the literal meaning of the text and build a situation model (Kintsch, 1998; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). However, the situation models of different documents must interact in several ways: (a) they may overlap, (b) a document may be part of the situation model of another document and, (c) different documents may yield incompatible situation models. Rouet, Britt, Mason and Perfetti, (1996) and Perfetti, Rouet and Britt, (1999) suggest that, when learning from multiple documents, readers build and additional level of representation, where both sources and contents of the document set are represented (Britt, Rouet, Georgi & Perfetti, 1994). This *argument model* or *documents model* accounts for argumentative relations between documents. It also allows the students to maintain contradictory information in a coherent representation. Finally, it may serve as a retrieval structure to perform specific learning tasks after reading multiple documents.

It is interesting to note that the reading of multiple texts produces additional representations that include relations between the texts. In many cases the relations are only implicit and may be unrecognized by the reader. In other cases, one text has information that builds on information learned through previous texts, as updating a situation model. It may also be the case that texts explicitly contradict each other, forcing the reader to recognize the connections between the texts. In any case, documents or texts are connected in a situation of reading from multiple sources. The reader, displaying connection processes which may be similar to those applied in connecting information in one text, should be capable of constructing a higher order representation, the *argument* or *documents model*, which contains both intertextual links among documents (i.e., *Intertext Model*) and an integration of each of the situations described in individual texts (i.e., *Situations Model*).

Thus, both the *Intertext model* and the *Situations model* are part of a broader *Argument* or *Documents model*. The Intertext model includes a node for each document and labeled links between documents and the situations they describe. According to the authors, every node has available slots for source, rethorical goals and content. Including slots for source or rethorical information may be more relevant in history documents but may be less important when dealing with expository texts. On the other hand, this Intertext model will be connected to those situations described in the texts. The connections then provide a full *Documents model*, one with texts and situations. In multiple documents, the ideal situation would be to reach an accurate and integrated representation of all situations described in all texts, which therefore constitutes a higher level of representation in the Documents model.

The degree to which documents are related to each other and situations connected and creating an integrated representation in memory determines the quality of the mental representation derived from reading from multiple sources. On the other hand, if readers fail to connect texts and construct a common higher order documents situation, the level of success after reading multiple sources would have been really low. Perfetti, Rouet and Britt, (1999) suggest that several factors such as different kinds of tasks or learner's goals may mediate the extent to which documents are connected and a higher order integrated situation built from the text.

In fact, when reading multiple sources, the goal of reading includes more than just learning the propositional content of the documents. Readers generally use document information in order to perform a specific task. Perfetti, Rouet and Britt, (1999) hypothesize that the kind of task should matter for multiple-document learning, just as it matters for single-text learning. Thus, the task should have an influence on how readers evaluate, memorize and use information from multiple sources. There is indeed research that confirms this influence.

Wiley and Voss (1999) conducted two experiments in which they provided students with information in a web site with multiple sources of information on a historical event and instructed them to write arguments, narratives, summaries or explanations. They found that performing argument tasks after reading the multiple sources produced increased conceptual understanding of the main topic presented in the texts as well as the most integrated and transformed essays on the topic students read about. Wiley and Voss (1999) argue that not all kinds of tasks were expected to produce learning from multiple sources and the reason why the argument task had the best result can be interpreted in light of the kind of mental processes it induces when working with texts.

In fact, students may benefit conceptually from tasks that promote the construction of a situation model integrated from all documents, just as it happens in the case of single-text learning (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted). Wiley and Voss (1999) remark that there are tasks, such as writing arguments, that promote integration from multiple sources and learning from the different documents. On the other hand, tasks that can be performed with a more superficial representation of the texts and establishing few connections among texts, would not lead to better understanding. It seems that writing arguments or essays from multiple sources promotes more transformation from the original sources and active students' engagement in the task. Indeed, transformation is regarded as a more active and constructive process in which the writer relates the contents of sources in new ways by making novel connections within source material as well as connections to the reader´s knowledge.

Using the terminology proposed by Scardamalia and Bereiter (1987) writing arguments from multiple sources produces the most beneficial results because it makes the student engage in *knowledge-transforming* processes from the diverse sources he or she is confronted with. Additionally, despite writing arguments or global essays from diverse sources may be a more demanding task in comparison to others, it turns out to be the most profitable in learning from multiple sources.

This result has continuity with past research demonstrating that conditions that make reading more effortful are finally the most beneficial for learning (McNamara, Kintsch, Songer & Kintsch, 1996; Mannes & Kintsch, 1987, Schmidt & Bjork, 1992). Hence, there seems to be tasks such as argumentative essays that actually promote and increase the quality of the integrated mental construction from different sources of information on historical events, the explanation being that that they make the student actively extract information from the sources and connect it via transformation processes. The final result of performing such tasks after reading multiple sources would be the actual construction of a Documents model in which all texts are connected and the situations described in each text integrated in a higher order situations model (Rouet, Britt, Mason & Perfetti, 1996; Perfetti, Rouet & Britt, 1999).

Thus, we have reviewed the existing literature on how students learn from multiple texts and we have also presented a proposal for a general model aiming at explaining which mental processes are responsible for integrating different sources of information into a single mental representation. Nevertheless, there is a set of questions and uncertainties that raise after reviewing the existing literature on multiple texts research (Wineburg, 1991; Perfetti, Britt & Georgi, 1995; Rouet, Britt, Mason & Perfetti, 1996; Rouet, Favart, Britt & Perfetti, 1997; Perfetti, Rouet & Britt, 1999; Wiley & Voss, 1999; Britt & Aglinskas, 2002).

The first of them is related to the concept of integration, which constantly appears in the different studies we have reviewed but is seldom clarified. It seems that integration is one of the nuclear processes in reading multiple documents, as it should help in constructing a coherent mental representation from different sources. To create coherence into the reader´s mind some mental activity has to be undertaken. And this is precisely the integration process, which would allow the connection among different units of information coming from the different texts. Integration would thus be a mental process that connects different units of information into the reader's mind. In the context of learning from a single text this connecting process was called inference making (Kintsch, 1998) and its activation resulted in deep learning. What about when learning from multiple texts? What kind of connecting processes could fall into the category of integration? Would they have behavioral correlates, similarly to the single-text situation? Displaying these processes would result in better learning from multiple texts and in the construction of a richly interconnected situations model from the several sources?

An approximation to the process of integration from multiple sources would be one of the classical heuristics Wineburg (1991) established in order to explain what students do when dealing with different historical documents. Thus, the corroboration heuristic was the general skill of checking facts or interpretations from a particular document against other, independent sources. It mainly involved comparing the information from the various sources to identify which statements or units of information were unique, which were contradicted and which ones were incomplete in one document and needed the combination with other units of information from other sources (Britt & Aglinskas, 2002). Corroboration would also be similar to the concept of reasoning with documents, as the ability to use document information when solving a problem (Rouet, Britt, Mason & Perfetti, 1996).

In any case, both corroboration and the concept of reasoning with documents leave many questions opened, basically how integration processes actually take place and why some kind of tasks are more beneficial than others in learning from multiple documents. Despite the limitations of many definitions of integration as a cognitive process, one successful and interesting approach to the concept of integration explained in terms of processing activity and that gives the first answers to the questions we are arising is the *reinstatement-and-integration strategy* proposal (Mannes, 1994; Mannes & Hoyes, 1996).

According to Mannes, readers who try to integrate knowledge derived from separate sources may use a *reinstatement-and-integration strategy*. During comprehension, readers who recognize information located in other sources will reinstate this information in short-term memory together with its original surrounding context. Because the current text and context occupy short-term memory at the same time, the reinstatement may afford the opportunity to actively construct new links between information from prior and current context.

The explanation of integrating information in terms of cognitive processes is clarified by Mannes in detail (Mannes, 1994; Mannes & Hoyes, 1996. Thus, a reader constructs a representation of the text material on the basis of its coherence relations (van Dijk & Kintsch, 1983). As propositions are encountered, they enter the reader's short-term memory buffer, and when possible, interpropositional relationships are derived. When the buffer is full, propositions currently in the buffer are removed and are often copied to long-term memory to make room for new propositions, and the process of deriving relationships between propositions in the buffer repeats itself. As this is a cyclic process, sets of text propositions are constantly encountered in proximity and occupy the same buffer. Consequently, in these cases there is a high probability of the reader becoming aware of the relationships between them (Kintsch & van Dijk, 1978).

In the situations of multiple text reading, different sets of propositions would be activated by a current reading that had been previously read in other texts. In these cases, if reinstatement of the previously read propositions and context occurs in short term memory, there is a chance for reinstated and new information to be integrated or connected by the drawing of inferences or the construction of elaborations. Nevertheless, this process does not normally occur automatically and seems to depend on the characteristics of the task that would promote different levels of connections among sources and the degree to which the match or mismatch between the sources favors or limits an active processing of the texts.

Thus, Mannes (1994) conducted a set of two experiments in which two groups of students studied an outline of an expository text and then read the original text. Whereas one of the outlines was similar to the text students would read afterwards, the other of the outlines differed from the text in the sense that it was written from another perspective. After the reading session, the students were assessed on several learning measures such as a networking ideas task consisting on drawing relations among ideas, a cued response task and a summary.

Interestingly, Mannes (1994) found that readers being given a different outline before reading the target text were able to build a richer domain representation. After the reading, they included outline material in their text summaries, thus suggesting that the two sources of information had become integrated and that participants considered outline material to be among the important things they had learned. In addition, the participants inferred more relationships between outline and text material as evidenced by the final learning measures.

They interpreted these results in terms of previous research that present studies on tasks that were apparently more difficult and resource demanding to perform but produced the highest learning effects (Mannes & kinstch, 1987; McDonald, 1987). Thus, reading two sources of information that apparently do not match each other seems to be a more difficult task as when the two sources present the same type and structure of information, as the reader in the first case has to actively find and produce the connections among the sources, mainly through the reinstatement-and-integration strategy.

This active processing to connect units of information explains why the different outline task produced better learning effects or the construction of a deep mental representation of the sources that allows further inference making and problem-solving.

Additionally, Mannes and Hoyes (1996) found behavioral correlates in terms of reading times for the integration process that takes place when two texts show diverse perspectives on the same issue and the reinstatement-and-integration process is activated. They conducted three experiments in which subjects read a target text after studying an outline that either did or did not conform to the perspective of the target text. Those readers for whom the text and outline presented multiple perspectives read more slowly than did readers for whom the experimental materials presented a single perspective. According to Mannes and Hoyes (1996) this probably occurred because they needed extra time to incorporate the current and original learning contexts for the target domain. But although this need for integrating may have been time consuming, it did not produce a deficit in comprehension.

Moreover, there was a general result that new appearing sentences were read more slowly than old sentences. And, especially, those new appearing sentences that occurred in a different context were read the most slowly. Again, the explanation to these results is given from the reinstatement and integration strategy perspective. The reader of such sentences first fails at a memory search and then attempts to integrate the new information with the currently active contents of memory.

The macroproposition that remains active from the previous reinstatement produces expectations that are different from those based on the current context. Thus, the integration process takes place, being demanding and time consuming.

Complementary, Mannes and Hoyes (1996) also conducted an informal think-aloud study in which, overall, the main finding was that readers who were hypothesized to have engaged in more integrative processing produced longer protocols that contained more elaborations. In any case, the aim of this final think-aloud study was to complement the big amount of on-line evidence that points out the existence of a mental process that allows integrating multiple perspectives in one coherent representation. The detailed definition of the reinstatement and integration strategy and the on-line processing evidence that authors provide allows us to have a clear perspective from which we would be able to study integration processes that occur under some specific circumstances. But some further points would still need to be clarified regarding research conducted so far on integration of information from multiple texts.

One of these points needing further clarification would be the role of the type of text in learning from multiple documents. So far, research on multiple documents has been conducted with historical documents. These kinds of texts have obvious peculiarities not shared with other kinds of texts the main of which is that they present controversies and discrepancies. Historical knowledge, in comparison to other kind of knowledge, is fraught with different interpretations which readers have to integrate to reach the organisation of a series of events in time and space. This is precisely one of the reasons why studying learning from multiple texts was especially appropriate in the history knowledge domain and why the ability to correctly indentify the source information from each document remarked in several studies (e.g., Britt & Aglinskas, 2002).

Nevertheless, the need to integrate and combine information from different sources of information is also present in many other fields using other kinds of texts to present information and which have remain unexplored in the light of learning from multiple documents. It is the case of expository texts. These are the most common kind of texts used in schools to promote learning of different contents. They are also the vehicle for transmitting scientific knowlege and neutral facts. They present neutral and valid knowledge normally exempt from controversies. Consequently, argumentative relations would not be as important in these kinds of texts as causal relations. Additionally, It seems that identificating source information would also be secondary when reading multiple expository texts, differently to historical documents, being more relevant what kind of information is presented in each document and how these units of information combine one to each other among texts, primarily by means of causal connections among units of information.

Despite the growing need to conduct specific research on how integration of multiple sources takes place when dealing with expository texts, llitle has been done on this issue. We should consider, though, an interesting article by Strømso, Braten and Samuelstuen (2003) in which they present a study of students' strategic use of multiple sources during expository text reading. They conducted a longitudinal think-aloud study to examine how students' strategic processing and their linking of multiple sources work together during the reading of expostory text.

Analysis of university students' think-aloud protocols obtained during several study sessions along the school year revealed that students primarly used a group of four different strategies when they read multiple texts. The first of them is *Memorization*, which occurred when the reader provided evidence of selection and rehearsal of text information, without trying to transform or move beyond the content given in the text. They called the second strategy *Elaboration* and was coded if the reader elaborated on or tried to make the text more meaningful by building connections between ideas located in the text, or by connecting ideas located in the text with ideas located somewhere else. The third of the strategies observed in students when reading multiple documents was *Organization* and appeared when the reader tried to relate, group, or order information and ideas given in the text. Finally, students showed also signs of *Monitoring* by providing evidence of assessing or regulating text comprehension.

Two main results should be highlighted in this study. First of all, the different uses of the above mentioned strategies. Thus, whereas memorization and organization were primarily used to process information located in one single text, the students used monitoring and, especially, elaboration, to construct more linkages external to the text they were currently reading.

Moreover, those students who focused their elaboration strategies on external sources were the ones who performed the highest at the year-end examination. The second main result was that, globally, students tended over the school year to focus their study strategies more on external than on internal sources. That is, they concentrated more and more on establishing relations from the text they were reading to other related sources.

Strømso, Braten and Samuelstuen (2003) argue that in these two main results underlies the same psychological explanation. Thus, when applying the elaboration and monitoring strategies to relate different sources of information, students were making an effort to build an integrated situation model, which also explained why they performed the highest in the final examination. On the other hand, the shift to external sources over the school year can be explained by the proximity of the final examination. The students' strategic processing seemed to become less directed toward the construction of a textbase, only from a single text, but more directed at the formation of a situation model over time.

In conclusion, Strømso, Braten and Samuelstuen (2003) have provided us with an interesting initial framework of how students deal with and learn from multiple expository texts and have concluded that the process playing a most determinant role in integrating information is elaboration, which seems to make students build an integrated situations model from all documents. However, despite these initial results and this attempt to clarify how integration from multiple expository texts actually takes place, some relevant questions still remain unanswered.

1.3 HOW LEARNING AND INTEGRATION PROCESSES FROM MULTIPLE EXPOSITORY TEXTS CAN BE ENHANCED BY USING SPECIFIC TASKS: OBJECTIVES AND HYPOTHESIS FOR THIS STUDY

Once we have reached this point, we are prepared to present the main objectives and hypothesis of this study, which might be easily suspected from the above paragraphs. Briefly, our main concern is to find new ways by which students profit the most from learning from multiple expository texts and contribute with new perspectives to the concept of integration from multiple documents. Additionally, we expect to shed some light on the mental processes displayed when reading multiple expository texts.

In order to be able to fulfil these goals, we raise the following set of questions and hypothesis. First of all, similarly to the single-text case, there should be tasks that are more effective than others in promoting integration of information and long-term learning from multiple expository texts. Those tasks would be especially effective in making students extract different units of information from the several sources, contrasting or corroborating one against the other (Britt & Aglinskas, 2002) and connecting them via integration processes (Mannes, 1994; Mannes & Hoyes, 1996).

The final outcome after performing these tasks would be increase in deep learning or the construction of a higher-order representation in which the different units of information from each document would be integrated (Perfetti, Rouet & Britt, 1999). It may happen that those tasks, in comparison to others, were more resource demanding, as they would make the student be more active in extracting and connecting relevant units of information.

Nevertheless, despite this performance effort, they would be the most effective in creating long-term learning (Mannes & Kintsch,1987; McNamara, Kintsch, Songer & Kintsch,1996), though differences with other kind of tasks would not be apparent when measuring performance or surface recall of the texts (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil,submitted).

To find such a task and a context in which it can be understood, we have designed a multiple expository texts learning situation. Our main concern was to choose a topic which could be partially tackled across three texts, that none of them fully presented it, but instead a reader would need to extract, contrast and combine specific units of information presented in each document, to be able to reach a clear view on the topic all texts dealt with. Moreover, we were especially interested in studying expository texts, which clearly presented information exempt from controversies. To meet these criteria, we selected three texts about the *biological mechanisms that make bacteria become resistant to antibiotics*.

Each of the texts partially presented the main issues related to these biological mechanisms that change the efectiveness of some antibiotics under specific circumstances. They also included information which was not relevant for understanding the main points of the problem. Thus, because none of the texts explained all the antecedents, causes and implications of this phenomenon in detail, in order for a reader to construct an integrated representation of how bacteria become resistant to antibiotics and why, integration across relevant units of information coming from each document should take place.

Only the student that read all three texts in detail, extracted the relevant units of information in each text and connected them via integration would be capable of constructing a coherent representation of the topic. In contrast, readers failing at extracting the relevant pieces of information in each text and connecting them, or those sticking only to one document and discarding the others or, finally, readers fixating only in some specific segments instead of all of them, would not be able to construct a veridical and coherent representation of the main issue presented across the documents.

We hypothesize that the task that would prove more effective in fostering the selection of relevant pieces of information across documents and their connection via integration would be making the students answer *global or very broad questions* which would focus students' attention on the main topic of the text. Thus, directly asking for the causes, reasons and biological mechanisms that make bacteria resistant to antibiotics would make the student read all three texts and connect the relevant pieces of information to reach a clear view of what he or she is being asked for. Similarly to the effectiveness of argumentative tasks in multiple historical text research (Wiley & Boss,1999) or the case of high-level questions in single-text learning (Vidal-Abarca, Mengual, Sanjose & Rouet 1996; Vidal-Abarca, Gilabert & Rouet, 1998; Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted), we expected that students answering global questions from the three texts would be able to acquire a deep and integrated representation of all the relevant pieces of information distributed across the three documents.

We also hypothesized that this global question would be especially effective in constructing this coherent representation given that it would directly promote integration across relevant units of information.

In contrast, we designed a complementary task in order to contrast the global question task to other types of questioning from multiple sources. For this purpose, we elaborated *brief questions* that focus students' attention only on a single text. Thus, they only ask for pieces of information present in one of the three texts, which are normally present in one sentence or in some cases require the production of inferences across close sentences in the same text. Therefore, the main feature of these brief questions is that, in contrast to global questions, they make the students concentrate on reading and extracting information from a single text, thus reducing the possibility to integrate information across documents. If we established a parallelism with other types of tasks used in previous research (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted), they would be equivalent to low-level questions for two reasons: a) because the location of the answer is closer in the text, b) because few inferences need be drawn to answer the question. Therefore, brief questions, in comparison to a global task and in continuity with past research (Cerdán et al., submitted) should not induce integration and deep learning from multiple sources as much as performing a global task.

In conclusion, we expect to find increased quality of learning and the display of information integration processes after answering a global question from three complementary sources about *bacteria resistance to antibiotics*.

Comparatively, brief questions should not allow so much the construction of an integrated mental representation of the texts, as they would direct the students' attention only to one of the sources, thus reducing the possibility of integrating information from the three texts.

Nevertheless, brief questions should produce benefits when considering other learning levels, that is, performance on the task and surface recognition of the ideas in the texts. Therefore, in reaching this point it is essential to mention the three learning levels we will consider as measures of the different quality levels in learning from multiple documents. The shallowest measure would be taking into account performance on the task. If results follow previous patterns similar to those obtained in single-text learning (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted), we should find no differences among tasks at this level. Or it might be the case that brief questions would increase performance as they would direct the student to the individual units of information in the texts. Given that brief questions focus students's attention not on integrating units of information, but on considering them individually, they would be especially helpful to extract these units of information for the task, thus increasing performance. However, they would limit the possibilities to integrate them to construct a higher-order mental model from all documents. Conversely, answering a global question may limit students' possibilities of extracting all the units of information for the task (i.e., performance), as it seems to be harder to choose what to select and integrate from a single instruction, but benefit integration of ideas instead.

The second learning level we include for measuring the different learning outcomes from the task would be surface construction of a mental model from the documents. That is to say, when students are capable of identifying and understanding the relevant units of information that would make up a mental representation that represents the problem of bacteria resistance to antibiotics and its biological mechanisms, based upon the three documents. Because students have correctly identified and understood the main points presented in each of the documents, they would be able to correctly identify these ideas in any kind of final test which would measure recognition and individual understanding of these units of information. Recognition of individual idea items would only measure surface learning from the three relevant sources, but it would leave apart how these ideas connect to each other to construct an integrated mental representation from the three sources. In other words, we would be measuring superficial understanding and recall of individual ideas but not integration across units of information, which is the focus of our interest and really determines the quality of learning from multiple sources.

We do not expect to find significant differences between tasks at this learning level. On the one hand, answering brief questions may foster the construction of this surface construction based on the identification and understanding of the main ideas on the topic. Let us remember that brief questions have been designed to make students be concentrated on the isolated ideas that make up the problem raised in the three documents. They have been designed to promote concentration on the main issues presented in each document individually and, because most of the times they foster understanding of these individual items via inference making across paragraphs in one text, they would be especially effective in promoting isolated comprehension from the individual points raised in each of the documents to understand why bacteria become resistant to antibiotics.

On the other hand, the global question task, being the most effective from the two tasks, would promote a versatile final learning. As this task would foster integration of all the relevant units of information to construct a higher-order representation of all the documents, this construction would also imply a lower-level construction, that is, the individual understanding of the ideas making up the integrated mental model from the three sources. Therefore, this lower-level construction needed for reaching the higher-level stage of learning would allow the student be versalite with the acquired knowledge, successfully performing all kinds of final learning tasks and always yielding high results. Consequently, when measuring superficial acquisition of the isolated units of information presented in the documents, we would not expect to find great differences between tasks.

Finally, the global question task should be the most effective when measuring the construction of an integrated mental representation of the relevant units of information present in each of the documents. As we have already hypothesized, the global question task would foster integration across documents in a greater degree than brief questions, thus allowing the construction of an integrated higher-order representation from the three sources.

Hence, when measuring the quality of this integration and the capability of students to apply this integrated knowledge to new but related situations, the global task should produce the best results.

In relation to this point, we raise the question of integration as a process and our interest to find some on-line evidence demonstrating that global questions actually foster on-line integration processes. In other words, if global questions yield indeed the best results when measuring the construction of an integrated mental representation, this should be because they make the student actively connect information from the three sources when performing the task in a greater extent than brief questions. To be able to make this inference, we would need on-line evidence showing that integration processes are more present when performing the global question task but less present when performing other kinds of questioning tasks.

This takes us back to some points of interest that we remarked earlier regarding integration as a process. First of all, we wondered what kind of connecting processes could fall into the category of integration and if they would have behavioral correlates, similarly to the single-text situation. We also pointed out if displaying these processes would result in better learning from multiple texts and in the construction of a richly interconnected mental model from the several sources. These main questions are thus completely valid at this point of our discussion.

We have established to consider integration in terms of the cognitive processes explained by Mannes (1994) and Mannes and Hoyes (1996), who provided strong empirical support demonstrating that integration of information implies connection of complementary units of information, partially related but not completely similar, which are simultaneoulsy activated in short-term memory. According to their findings, when students integrate different units of information during reading multiple sources, interesting differences appear in their reading speed rates.

In brief, if readers are exposed to a multiple texts reading situation in which they have to contrast and integrate information, their reading speed would be more slowly in comparison to another group not exposed to a multiple text situation. This decrease in reading speed would be explained by the fact that integration of units of information would occur during reading and this need of additional processing would need extra time. Another interesting result they found was that new information was read more slowly and this especially ocurred when this information needed to be combined with information previously read.

It should be remarked that the empirical support provided by these authors came from on-line time-based measures. Indeed, they supported their claim that integration occurred in a greater extent in multiple perspectives texts situations by tracking how students behaved while reading. This is in fact an interesting approach to untangling the mental processes involved in comprehension and when performing learning tasks. Indeed, it has successfully been used in text comprehension studies to examine processing from text more deeply and accurately.

For instance, Coté, Goldman & Saul (1998) used a software called *Select-the-Text* (Goldman & Saul, 1990a) to present texts sentence by sentence on the computer screen and they examined children's strategies for processing informational text to understand and remember new information. *Select-the-Text* collected information on how long each sentence was read and the order in which sentences were accessed. These outputs were very valuable on-line data which enriched conclusions drawn in the study. *Select-the-Text* served as basis for elaborating *Read&Answer* (Martínez, 2003), a software that also presents texts and questions electronically and enables the researcher to obtain useful reading and answering processing records. By using *Read&Answer* we were able to determine which were the main strategies associated to answering high and low-level questions and it helped us in obtaining on-line evidence supporting our claim that high-level questions promoted connections among units of information and deeper learning in a greater extent than low-level questions (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted). Similarly, interesting on-line evidence demonstrating that integration occurs while answering a global task and reading multiple sources could be obtained by using *Read&Answer* to present texts and tasks electronically. This will be in fact one of the main tools we will use to deepen into integration processes. However, a more detailed description on *Read&Answer*, the main outputs and how the experimental design was created will be presented in further sections of this study.

Therefore, in continuity with previous findings (Mannes, 1994; Mannes & Hoyes, 1996) and primarily based on on-line tracking of reading and question-answering behavior, we set the following hypothesis on how the global task would show integration processing evidence on-line.

Thus, given that answering a global question would make the student locate the relevant points on bacteria resistance to antibiotics in each of the documents and combine those, answering a global question would induce integration processes. These integration processes should be apparent in the way students read information across documents. Mainly, reading speed should decrease when reading the relevant units of information making up the general representation of the problem and, conversely, reading speed should become faster when reading irrelevant pieces of information, which are also present in the three documents. If this ocurred, it would be strong empirical evidence that integration is actually taking place when reading to answer a global task. On the other hand, as brief questions focus students' attention only on one of the texts, integration processes should not occur to such an extent, neither the above reading pattern suggested for the global task.

Moreover, if integration is promoted by the performance of a global task, this integrating process should also be apparent in how students read the three texts across time. Therefore, considering the whole experimental sequence, students with the global task should read more slowly at early stages of their sequences and read faster when finalizing their task. This would indicate that when they encounter new information to be combined for the first time, integration of information needing additional processing time would occur. And, conversely, when the students would re-read the information at later stages of their reading sequences, they would only be reviewing information which would have already been integrated.

In terms of Mannes (1994) and Mannes and Hoyes (1996) it would be old information. Thus, it would not need the display of integration processes in short-term memory.

On the other hand, we expect not to find such a clear pattern for brief questions. Brief questions focus students' attention on a single text and on the isolated ideas making up the general model on how bacteria become resistant to antibiotics. Thus, if there is little opportunity for integration to occur, the reading sequence will have a different pattern than the global questions' reading sequence. Globally, we believe that the reading speed will not experience a progressive decrease in the course of the experimental task, but on the contrary it may happen that students will read faster even at later stages of their performance sequence. The explanation would be that each of the brief questions focuses on isolated pieces of information, which could be slowly read to answer the questions even at later stages of the experimental sequence.

On the whole, we have developed so far our main hypothesis and expectations for this study. Briefly, we are interested in finding tasks that promote deep learning and integration from multiple expository texts and we believe that a global task, in the terms described above, would probably create this expected outcome in a greater extent than brief questions. Moreover, we expect to find on-line evidence to support our claim that integration actually takes place when performing such tasks.

If we are able to come to this point, we will hopefully shed some light on what integrating information from multiple sources consists of, which is our broader aim for this study. To examine all the set of questions raised so far, we designed an experiment in which university students read the three texts on bacteria resistance to antibiotics and, simultaneously, they performed a global task or answered brief questions. Both reading and answering being performed using a computer software called *Read&Answer* (Martínez, 2003), which permits the electronical presentation of texts and the possibility for students to examine the three texts while answering.

Complementary, to deepen into strategic reading and answering behavior in comprehending multiple sources of information and hence complement on-line evidence, we used in experiment 1 the think-aloud method in half of the sample, the other half performing under silent conditions. This way, we would be able to obtain congervent data to support our claim that some tasks are more effective than others when working with multiple sources. Additionally, because we compare silent performance to performance and thinking-aloud, it would be interesting to test the degree to which making students think-aloud interferes or not with performance and learning from multiple documents. Nevertheless, before detailing our hypothesis on the effects of thinking-aloud in working with multiple sources, we believe that a brief revision on the use and effects of this methodology should be presented.

The think-aloud method has been demonstrated to reflect what is available in working memory, accessible to consciousness and codable in language (Ericsson & Simon, 1980; 1993) and would therefore be indicative of which mental processes and contents are responsible for how students perform and learn from an specific task. Consequently, by making students verbalize what they are thinking simultaneously to the task (i.e., *concurrent verbal reports*, Ericsson & Simon, 1980; 1993) or after performing the task (i.e., retrospective *verbal reports*, Ericsson & Simon, 1980; 1993), we can analyse in greater detail on-line processing of text, strategies and problem solving.

In fact, there is a wide range of empirical studies that have successfully used this methodology in different fields of interest in Psychology. To our interest is the growing use of this methodology to study general comprehension processes and strategies that students apply when reading texts, validating that the content of think-aloud protocols is linked to comprehension (Chi et al., 1989; Yuill, Oakhill & Parkin, 1989; Coté & Goldman, 1999; Goldman & Durán, 1988; Magliano et al., 1999; Pressley & Afflerbach, 1995; Trabasso & Magliano, 1996b; Whitney et al., 1991; Zwaan & Brown, 1996). Moreover, not only has this methodology been used to study on-line processing of one text, but it has also been used in classical and recent studies of multiple text comprehension and integration processes from multiple perspectives, as we reported earlier, yielding highly interesting results. (Wineburg, 1991; Mannes a & Hoyes, 1996; Strømso, Braten & Samuelstuen, 2003).

However, two critical issues are long-standing in past research as regards to the validity of think-aloud data and the degree of interference or not with performance and learning.

Critical voices have claimed that readers, when thinking aloud, do not provide a veridical report of their underlying mental processes (e.g., Nisbett & Wilson, 1979). Rather, they would construct a text representation and then use it to “tell a story” about their understanding (Long & Bourg, 1996). Readers thus seem to construct a verbal report that conforms to the pragmatics of the situation. Some of the information in their reports would reflect processes that ocurred when they comprehended the text, but other information would reflect processes that occurred when they constructed their “story” (Long & Bourg, 1996).

Despite this, using a think-aloud procedure would provide very valuable data on processes and strategies that occur during comprehension, and the validity of these data would be increased if convergence with other kind of data were provided (Long & Bourg, 1996; Whitney & Budd, 1996). Indeed, successful attempts have been done to provide converging evidence for the conclusions drawn from protocol data (Trabasso & Suh, 1993; Magliano & Millis, 2003; Suh & Trabasso, 1993; Trabasso & Magliano, 1996). Additionally, using the think-aloud methodology as only a part of a converging evidence strategy has been presented in Magliano and Graesser’s (1991) three-pronged approach to comprehension research.

Nevertheless, another controversy arises regarding the degree to which using a think-aloud procedure interferes with performance and learning or not, when comparing to silent reading conditions.

Ericsson and Simon (1980; 1994) argue that subjects can generate verbalizations, subordinated to task-driven cognitive processes, without changing the sequence of their thoughts, and slowing down only moderately due to the additional verbalization. They insist that if subjects are only asked to verbalize thoughts per se (type 1 and 2 verbalizations), no interference effects would be apparent, providing wide evidence based on research. Nevertheles, some experiments have requested subjects to provide reasons, justifications and elaborations (type 3 verbalizations). In such cases, they find studies in which verbalization specifically affected performance (Ericsson & Simon, 1993), but the directions in which it affected performance were diverse, in some cases increasing performance and learning, in other cases hindering both of them.

The above discrepancy is also present in text comprehension think-aloud studies. There is evidence that having students verbalize their thoughts while performing a task creates no difference compared to a silent reading condition. Thus, two studies with adult readers found no differences in comprehension between readers who stopped to give think-aloud comments and those who read silently (Crain-Thoreson, Lippman & McClendon-Magnuson, 1997; Fletcher, 1986). But, contrarily to these results, Loxterman et al. (1994) found that sixth graders who thought aloud while reading outperformed silent readers on recall and comprehension measures.

Based on this result, Coté, Goldman & Saul (1998) hypothesized that protocols may actually encourage the construction of a better representation of the text by inducing more active engagement with the text, leading to better encoding of material, more connections of prior knowledge and hence a better mental representation. They conducted an experiment using think-aloud data in one condition compared to silent reading to examine children's strategies for processing informational text to understand and remember new information. When they compared recall performance between the think-aloud and silent reading condition they found a surprising result. Whereas the think-aloud procedure facilitated the 6th grader's performance, it hindered that of the 4th-grade students.

Contrary to Coté, Goldman & Saul (1998), we have found evidence that thinking-aloud may hinder comprehension and learning from text in college and university students. For instance, Wade and Trathen (1989) studied the role of noting of ideas in a text and its effects on recall. Interestingly, they found that asking students to describe their strategies while studying reduced the amount of information they noted and had negative effects on recall. They conclude that verbal reporting during reading affects both the process and product of studying.

However, we would like to highlight results obtained by Magliano, Trabasso and Graesser (1999). They conducted an experiment to study how strategic processind during comprehension is affected by properties of the text and how different strategies can affect text retention. Additionally, think-aloud protocols were collected to predict sentence reading times for other participants reading silently. As regards the effects of think-aloud, they found that a strategy to explain led to an increase in memory relative to a strategy to understand when reading was done silently.

Thus, reading strategies did not have an impact when thinking-aloud. The authors argue that it is possible that thinking-aloud wiped out the general benefits of reading to explain. Moreover, they point out that possibly the process of describing the thoughts that occurred at a sentence may had strengthened the memory representations for the explicit text. Indeed, story recall was better when thinking-aloud than when reading silently. In short, thinking aloud may improve memory for text because it requires a more conscious processing of text (Magliano, Trabasso & Graesser, 1999).

After reviewing the main issues concerning the use of the think-aloud methodology and its implications on performance and learning, we are able to make some predictions of the kind of data that the analysis of students' think-alouds will provide us and the possibility that thinking-aloud will interfere or not with performance and learning on the task in experiment 1. Overall, our interest in using the think-aloud methodology when students are performing their task is to deepen into the process of integrating information from another perspective already explored by other authors in multiple texts learning situations (Wineburg, 1991; Mannes & Hoyes, 1996; Strømso, Braten & Samuelstuen, 2003). Additionally, we expect to find convergent evidence four our main claim that integration of information taskes place in a greater extent when performing a global task.

However we especially want to analyse to which degree making students think-aloud while answering questions from multiple sources interferes or not on performance and learning. Given that we have evidence on both sides, we expect to find both possibilities in our study.

On the one hand, it might happen that thinking-aloud, as hypothesized by Ericsson and Simon (Ericsson & Simon, 1980; 1993) will not interfere with learning from the task, and maybe only creating an slightly increase in experimental time due to the need of thinking-aloud and performing the task simultaneously. On the other hand, what if our studies follow results obtained by other authors who studied comprehension and learning processes in university students (Wade and Trathen, 1989; Magliano, Trabasso & Graesser, 1999) and found that thinking-aloud hindered learning from text? It might be the case that thinking-aloud implies indeed an additional processing effort to construct a story to tell the experimenter (Long & Bourg, 1996). Because processing resources in short-term memory are limited (e.g. Miller, 1956) and thinking-aloud would increase the cognitive load in short-term memory (Paas, Renkl & Sweller, 2003; Sweller et al., 1998), these factors would result in decreased learning due to a short-term memory overload.

In terms of the *Cognitive Load Theory* distinction (i.e., Sweller, 1994), thinking-aloud would increase *extraneous load* in short-term memory, not intrinsic to the task itself, but due to the characteristics of the instructional design, in our case making students think-aloud simultaneously to the task. Moreover, it might happen that this extraneous load placed on students' short-term memory would be added to an additional cognitive load due to the nature of the task itself. Indeed, *intrinsic load* or the load placed on working memory due to the nature of the materials to be learned (Sweller, 1994) might have been increased beacuse of the fact that reading multiple sources and answering different kinds of questions was more resource demanding than answering questions only from one text. Therefore, the interaction of an increase of the difficulty of the task or intrinsic load and the extraneous demands due to the need of thinking-aloud would be responsible of yielding worst performance and final learning, in comparison to students who did not thought aloud. In any case, this is only one of the two possibilities we expect to find.

In sum, we have developed a broad theoretical framework and clearly raised the main points we are interested in, to come to this point in which we exactly know what we are looking for and how this could be approached from an experimental point of view. It is thus high time to explore the experimental data and how these data can answer our questions and hypothesis. We will start with the development of our first experiment (i.e., *Experiment 1*) in which, as we discussed earlier, we analyse two main aspects. The first of them, are gobal tasks more effective than brief questions in promoting integration and learning from multiple expository texts? And, if so, would they have behavioral on-line correlates demonstrating that integration actually takes place? The second question assesses the degree to which making students think-aloud simultaneously to the task interferes or not with performance and learning from multiple sources.

2. EXPERIMENT 1: PERFORMING A GLOBAL TASK vs. ANSWERING BRIEF QUESTIONS WITH MULTIPLE EXPOSITORY TEXTS: THE ROLE OF INTEGRATION PROCESSES AND THINKING-ALOUD.

Experiment 1 thus examines how effective is performing a global task, which makes students concentrate on integration from multiple sources, in comparison to answering brief questions, which focus students' attention on single documents. We therefore present a multiple expository texts learning situation in which university students read three texts on bacteria resistance to antibiotics and simultaneously answered one global or four brief questions on the information contained in the three documents. All students performed their task using an electronic presentation thanks to the software *Read&Answer* (Martínez, 2003), but only half of the sample was instructed to think-aloud.

2.1 Method

Participants

Fifty university students contributed to the development of experiment 1. They were all enrolled in a Psychology undergraduate programme at the University of Valencia and participated for course credit. They had a mean age of 20 years. They were selected from this specific sample of Psychology undergraduates to meet two criteria for our study: that they were university students with deeply-rooted habits of handling with and learning from expository text. Because we were specifically testing how learning strategies applied to one text would be apparent in multiple texts situations, we were interested in selecting a sample used to learning from expository text.

They were randomly distributed into each of the conditions in our experiment. This way, whereas 11 students were assigned to the *global task&think-aloud* group, 12 students went into the *brief questions&think-aloud* group. Hence, the think-aloud students totalling 23. On the other hand, 14 students were assigned to the *global task&no-think-aloud* group and the other 13 were assigned to the *brief questions&no-think-aloud* group. Thus, the no-think-aloud students totalling 27. The reason why there were slightly fewer students in the think-aloud condition is because we had to discard some of them because of failures in the registration of their verbal reports. In any case, we ended up having a similar number of students performing a global task (i.e., 25) and answering brief questions (i.e., 25).

Materials

Control measures. To verify that experimental conditions did not differed in a set of measures that could contaminate subsequent results, we used three specific tests that would further guarantee that the initially conducted randomization was indeed effective. The most important of all three is the *previous background knowledge test*. As students were expected to learn a group of contents on bacteria resistance to antibiotics from three different texts, we should certify that students did not master these issues previously. If this would be the case, the student should be descarded, because we specifically wanted novices on knowing how bacteria become resistant to antibiotics and the underlying biological mechanisms.

On the other hand, we did not want experimental conditions to significantly vary on previos background knowledge, given the big impact that this variable may have on performance and learning from multiple texts.

The *previous background knowledge test* presents students with a 15 true/false item questionnaire that directly assesses a basic knowledge on the biological mechanisms responsible for producing resistant bacteria to antibiotics. The maximum score students can reach is 15, thus being each of the items scored 1. All items clearly cover the main relevant contents that students will read about in the multiple texts learning session. We made an extra effort in order for this match to be possible, as we considered that students should only be assessed on the basic contents they will be expected to learn. An example of an item of the previous background knowledge text and aiming to be representative of the whole test says as follows: *"Bacteria are multi-cell organisms"*. Clearly, this is a false item that attempts to examine the degree to which students have or not a basic notion of what bacteria are. Another reprensentative item of the previous background knowledge was the following: *"Antibiotics are medicines that destroy or kill bacteria"*. This time, we were testing if students knew what antibiotics were used for, as this would be an elementary issue that would be treated in greater detail in the multiple texts situation. In general, the *previous background knowledge test* was created to have a control measure for what students knew in anticipation of the main topic they would encounter in the multiple texts session and the rest of the 14 items were produced in a similar way.

The second control measure we used was a test on *lexical access*. Given that the ability to activate the meaning from words from a written representation is a prerrequisite for comprehension at a higher level, we should measure the extent to which students have a fluent lexical access, according to their age. This would serve us especially to compare our experimental conditions and see if they are equivalent in this low-level stage of language. If this is not the case, data adjustements should be carried out, so that conclusions drawn from performance with multiple texts are not confusing.

To assess lexical access, we used a comercial test that precisely covers this dimension (PROLEC-SE, Cuetos, Rodríguez & Ruano, 2000). The test consists of a list of 40 words and a complementary list of 40 pseudo-words. Students are required to read both of them aloud and time and number of errors committed are registered. This way, we have an accurate measure of how automatized has become reading in this adult university students sample. Indeed, as adult readers, they would be expected to read faster the words they know, without committing mistakes, and slightly slow down when reading unkown words, increasing the number of mistakes. We should find no significant differences among experimental conditions in this dimension either.

Finally, the last of the three control measures we used was a test on *writing speed on the keyboard*. Given that all the experimental session, both reading and answering questions, would be conducted using the computer, we were interested to test if there would be any student having special difficulties in writing using a keyboard, in such case to be discarded, or if experimental conditions significantly differ or not in this ability, which we expected they would not. To assess *writing speed on the keyboard,* participants were given 2 minutes to copy a 146 word text onto a computer. Afterwards, we considered the number of words they were able to copy.

Software Read&Answer. The multiple texts were presented on a computer screen using the application *Read&Answer* (*Martínez, 2003)*. Similarly to *Select-the-text* (Goldman & Saul, 1990), it presents readers with a full screen of text. All text except the segment (i.e., a sentence or a paragraph) currently selected by the reader is masked. Readers unmask a segment by clicking on it and when they unmask another segment, the first segment is remasked. Thus, only one segment is visible at a time, but the graphic features of the text (e.g., paragraph indentation, length of the paragraphs, position of the segment in the text) are visible to the reader. Readers can reread the segments in any order they choose (*see figure 1.1*)

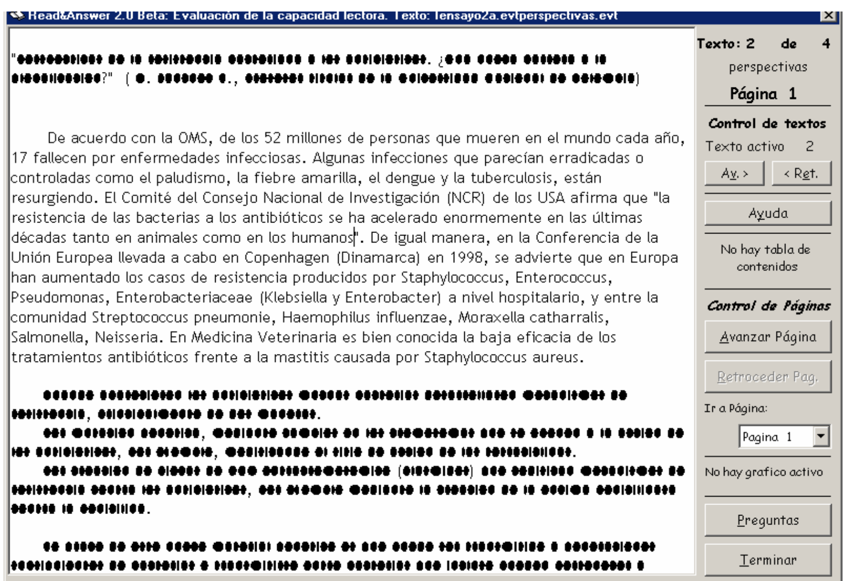


Figure 1.1 Read&Answer Text screen

Read&Answer includes other possibilities that *Select-the-text* does not, which are especially useful for recording the reader's behavior when she or he is involved in question-answering tasks based on a long text or multiple texts. Read&Answer presents the text on different screens corresponding to the different pages. A simple interface allows the user to navigate among them. In addition, visual information (e.g., diagrams, figures, pictures, etc) can be inserted into the text as a segment. On the other hand,, if we wish to create multiple texts learning sessions, as it is the case, *Read&Answer* includes this possibility. The reader only has to click a button on the right side of the screen, in the navigation toolbar, to change from one text to another.

Read&Answer also presents the reader with a question screen, which is divided into two parts, the upper part for the question and the lower part for the answer. This question screen can be accessed through a button on the navigation toolbar from any text that the student may be reading at a given moment, thus allowing flexibility in the question-answering process. In the question screen, the user clicks on each part to either read the question or write in the answer box. A simple interface allows the reader to move from one question to another and from the question screen to the text screen, and viceversa (*see figure 2*).

Read&Answer automatically generates three outputs. The first is a list of all the segments active at any given moment and sequentially ordered following the students' performance in the experimental session. It also provides the length of time each segment was active. A piece of the text (e.g., a paragraph), a specific question, and the answer to every question are all segments. Thus, every action the reader undertakes, whether it be reading a text segment, reading a question, rereading a text segment, or writing an answer, is recorded and included in the sequential list. The second output is a summary of the reader's behavior when he or she reads the text and answers the questions. *Read&Answer* provides a different summary of the two types of study behavior, reading the texts and answering the questions. The summary includes: (1) the number of words in the text segments, (2) the total amount of time the segment was exposed, and (3) the reading rate per word. The third output is the record of the reader's answers to each question.

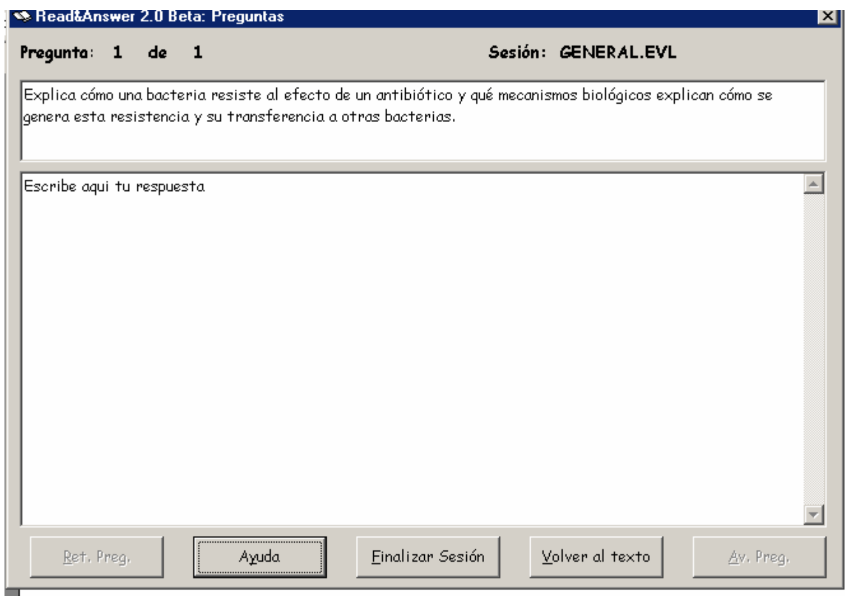


Figure 1.2 Reading and question-answering process in Read&Answer

Consequently, these three outputs allows us track in detail the on-line reading of texts and question-answering process. Indeed, as we reported earlier, we have successfully used them in past research to track students' question-answering behavior in one text (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted). Now we expect that they will provide us with valuable data of how students read multiple texts, answer different kinds of questions and integrate information from complementary sources of information.

Multiple texts, complementary contents and associated tasks. To emphasize which task we expected to produce the most integration from multiple sources we briefly described how we had designed a multiple texts learning situation. Thus, we had selected three expository texts that explained which biological mechanisms are involved in producing bacteria resistance to antibiotics. We wanted the texts to be exempt from controversies and easy to understand, not deepening into complex explanations of the underlying biological resistance mechanisms, but understandable by any non-expert audience. For this purpose, we carried out a selection of texts coming from popular science web sites that clearly presented the main topic we were interested in. The fact that they were from popular science web sites did not imply lower quality of contents, because they were written by prestigious scientists and institutions who adjusted their discourse to a novice audience, which was precisely what we were interested in.

After conducting a precise analysis of contents, we finally selected the three texts present in our study, following length, comprehensibility and content-relevance criteria. Generally, our main interest was to create a multiple texts learning situation based on complementary relationships across sources. This implies that, in order for a reader to wholly understand the bilological mechanisms that make bacteria resistant to antibiotics, he or she would need to extract relevant pieces of information coming from each of the three texts and combine them, so that they make sense. In other words, not all sources give the whole perspective of how bacteria become resistant to antibiotics, but instead it should be the reader who extracts relevant units of information across sources and integrates them in a higher-order representation that goes beyond the individual texts. In other words, the reader has to construct a puzzle using specific pieces that are present in each of the texts. This implies the ability to select which pieces are relevant in each text and find out how they would combine (i.e., integrate) to construct the puzzle which should explain why bacteria, under some specific circumstances, become resistant to antibiotics.

Therefore, we selected three texts that complied with our criteria, that is, presenting understandable biological explanations of how bacteria become resistant to antibiotics, combining relevant and non-relevant information, not excessively long and complementary to each other. The first of them is called *Perspectives on bacterial resistance to antibiotics. A new barrier to globalisation?* and the author is *G. Sánchez*, *full professor at the National University of Colombia*.

It is a 658 word text distributed across two *Read&Answer* pages. Because *Read&Answer* organises reading through masking and unmasking of paragraphs, this text was divided into 6 segments, 3 of which presented in the first page and the other 3 presented in the second page. From this 6 segments, only one was relevant for constructing an integrated representation of the biological resistance mechanisms, the rest of the information was completely irrelevant. This relevant segment was presented in page 1, and was paragraph 3. We coded it as Perspectives 1, 3 (i.e., P 1-3) to clearly indicate which relevant segment we are referring to in further analysis.

The second of the texts we included to make up our multiple texts learning situation is *Genetics of bacterial resistance*, with autorship falling in the *National Spanish centre for the research on infectious Diseases*. It is a 390 word text, shorter than the previous one, distributed across two *Read&Answer* pages. It was also divided into paragraphs, 3 present in the first page, and 2 present in the second page. Thus totalling 5 paragraphs, from which only two were relevant to understand why bacteria become resistant to antibiotics. The first of the relevant segments is the third one located in the first page (i.e., G 1-3) and the second relevant paragraph is the first one located in the second page (i.e., G 2-1). Thus, the proportion of relevant information in this text is higher than in the first text. And one of the main reasons is that it directly presents a genetic explanation of the underlying biological mechanisms in resistance. However, the principal explanation is that it overlaps contents with other texts, thus part of the relevant information is a repetition of what a reader would find in one of the other two sources.

In any case, the only reading of this text is not enough for understanding the biological mechanisms and their implications in detail, but need the combination of other units of information in the other two texts.

Finally, the third of the texts we used for this study is *Anti-Bacteria Resistance,* steming from the *National Institute of Allergy and Infectious Diseases, National Institute of Health of USA*. It is the longest of all three, 684 words, and the one with the highest proportion of irrelevant information, as it mainly deals with research implications of bacteria resistance to antibiotics. Despite this, one of the paragraphs contains key information for the construction of a mental model to understand this phenomenon. Hence, this text is distributed across three *Read&Answer* pages containing overall 8 paragraphs to unmask. In the first page, there are three paragraphs, none of them relevant. In the second page, we again find three different paragraphs. This time, the first paragraph we encounter is relevant (R 2-1). Finally, the last page presents two irrelevant paragraphs.

Therefore, so far we have presented how we selected a group of three complementary texts to promote students integration across sources. Nevertheless, we have still not developed in detail which contents we considered to be the only necessary and relevant to construct a mental model of the biological mechanisms responsible to produce resistance and how these contents were distributed across sources. This is precisely our focus now.

We conducted a detailed content analysis of the three texts to clearly specify which units of information were students expected to extract and connect. With units of information we mean individual sentences containing at least a subject and a verb. These units should be equally distributed across the three texts, so that a reader confronted to the three sources would need to read the three texts in detail, extract the relevant units of information from each source and connect them via integration, in order to fully understand why bacteria become resistant to antibiotics. Furthermore, these units of information would not appear isolated, but grouped into paragraphs in each of the texts. As we have specified above, the texts normally presented one or two relevant paragraphs were all units of information were contained.

As it will be apparent in the following *table of contents across texts*, content analysis of the texts help us establish 22 main units of information that would make up a mental representation of how resistance is produced. These ideas were grouped into four main areas and presented in the three different sources.

TABLE OF DISTRIBUTION OF CONTENTS ACROSS TEXTS			
AREA 1	***DESCRIPTION OF BACTERIA***	**TEXT**	**SEGMENT**
1	*Bacteria are single-cell organisms with a reduced amount of genes.*	R	2-1
2	*They multiplicate themselves fastly.*	R	2-1
3	*They are highly capable of adapting to any kind of environment.*	R	2-1
4	*This is a key factor for the development of resistance to antibiotics.*	R	2-1
AREA 2	***BIOLOGICAL RESISTANCE MECHANISMS***		
5	*When antibiotics appeared, many bacteria developed resistance mechanisms.*	P	1-3
6	*There are two resistance mechanism.*	P	1-3
7	*Genetics mutation.*	*P,G*	*P 1-3; G 2-1*
8	*The transfer of resistance genes from one bacterium to other bacteria (Plasmids)*	*P,G*	*P 1-3; G 2-1*
9	*Genetics mutation implies changes in the genetics information of bacteria.*	*P,G*	*P 1-3; G 2-1*
10	*These changes prevent them from the antibiotics effects.*	P	1-3
11	*Plasmids are pieces of extra-cromosomic DNA.*	P	1-3
12	*Plasmids produce resistante mechanisms to antibiotics.*	P	1-3
13	*Plasmids may produce resistance to more than one antibiotic. This is called multirresistance.*	G	2-1
AREA 3	***RESISTANCE'S IMPACT AND TRANSFER***		
14	*Even one random mutation can have a big impact, due to the high bacteria's multiplication rate.*	R	2-1
15	*Resistance can be transmitted to other generations of the same bacteria.*	*P, R*	*P 1-3, R 2-1*
16	*Resistance can also be transmitted to other generations of bacteria not related to the original ones.*	*P, R*	*P 1-3, R 2-1*
AREA 4	***RESISTANCE RESPONSES***		
17	*Bacteria respond in three different ways to the letal effects of antibiotics.*	G	1-3
18	*They can modify the antibiotic's chemistry.*	G	1-3
19	*They can degrade the antibiotic.*	G	1-3
20	*They can modify the target of the antibiotic in the bacteria.*	G	1-3
21	*They can stop antibiotics from penetrating.*	G	1-3
22	*They can expel the antibiotic.*	G	1-3

Thus, the 22 main units of information were distributed across 4 content areas. As it can be observed in the above presented table, the first of the areas is called *Description of bacteria* and contains 4 units of information present only in one of the sources. Thus, they are developed in the text *Anti-bacteria resistance* (i.e., R) and appear in paragraph 1, in the second page (i.e., R 2-1). Knowing what bacteria are and how they reproduce is a basic content for understanding how resistance appears, being this the reason why we include this area as essential for building a mental model on bacteria resistance to antibiotics.

The second of the areas deals with the nuclear issues of the mental model, that is, the genetic explanations of why bacteria can produce resistance to some antibiotics. This second area, *biological resistance mechanisms*, consists of 9 units of information which mainly appear in the text *Perspectives on bacterial resistance to antibiotics*, in page 1, paragraph 3 (i.e., P 1-3). Five out of 9 units of information exclusively appear in P 1-3, but 3 units (i.e, ideas 7, 8 and 9) overlap with the *Genetics* text. Indeed, the text Genetics also covers this three idea units (i.e, G 2-1) repeting contents from the text Perspectives. Finally, one idea unit (i.e., number 13) is exclusively covered by the text Genetics (G 2-1). In any case, as most of the contents are primarily developed in the text *Perspectives*, we assume this text to be the main source for this area.

Area 3 is a smaller group of 3 units of information dealing with the impact and transfer of bacteria resistance. Ideas 14, 15 and 16 are primarily present in the text *Anti-bacteria resistance*, similar to area 1 (R 2-1).

Nevertheless, some contents (i.e., ideas 15 and 16) overlap with the text Perspectives (P 1-3). Thus, ideas 15 and 16 are equally presented both in the *Anti-bacteria resistance* and *Perspectives* text. It should be noted that the fact of having some contents overlapping does not affect the integration process, as it only occurs in 5 units of information out of 22. Additionally, it is only contents that are equally presented in two sources, so operatively they work as belonging only to one source, because a reader extracting them from one of the sources would not need to inspect the other source to obtain the same information.

Finally, area 4 is completely covered by the text Genetics (G 2-1). It mainly deals with the different responses bacteria may use to prevent antibiotics from working on them. It is thus a complementary content to understand how bacteria respond once they have become resistant. In a mental model of how and why bacteria create resitance to antibiotics, the current contents would occupy a less relevant position.

To sum up, we have selected 22 units of information which need to be extracted and integrated from the different sources to make up a mental model of how bacteria produce resistance mechanisms. These 22 idea units are distributed into content areas and differentially present in the three complementary sources. Thus, area 1 is present in the text *Anti-bacteria Resistance*; area 2 has as a primary source the text *Perspectives*; area 3 is covered again with the text *Anti-bacteria Resistance* and, finally, area 4 is mainly present in the text *Genetics*.

Therefore, any student wishing to construct an integrated mental model would need to inspect all sources, extract the relevant units of information and, finally, integrate them. To conclude, it should be noted that the order of presentation of the sources was randomized in the experimental session, so that the potential effect of the position of the text would not interfere with the integration process.

This detailed table of basic contents was the basis for the design of the different tasks we used in Experiment 1. We had suggested the comparison of two different kind of tasks. On the one hand, making students perform a global task that would make the student inspect the three sources, extract the relevant units of information and integrate them, to construct a mental model of how bacteria produce resistance. This global task, thus, would cover the 22 units of information presented previously and, moreover, it would not promote an isolated identification of these ideas, but their connection through integration processes.

In the present experiment, our interest is to compare this global task to 4 brief questions which would directly match the 4 areas in the table of contents. Therefore, each brief question would call for the isolated contents contained in each area. Because each of the areas is primarily focused on one text and the extraction and understanding of isolated areas would not guarantee integration among them, we do not expect deep learning effects from this task. A detailed description of the two kinds of questions and their corresponding contents is presented below in the *hierarchichal table of questions across contents*.

HIERARCHICAL TABLE OF QUESTIONS ACROSS CONTENTS IN EXPERIMENT 1			
QUESTIONS: CORRESPONDING IDEAS AND TEXT SEGMENTS		**IDEA**	**TEXT**
GLOBAL	***Explain how bacteria resist the effects of antibiotics and which biological mechanisms explain this phenomenon and its transmission to other bacteria.***	**22 IDEAS**	**P,G,R**
BRIEF 1	*Which characteristics of bacteria have an influence on the development of bacteria resistance to antibiotics?*	1,2,3,4	R 2-1
BRIEF 2	*Which biological mechanisms permit bacteria become resistant to antibiotics?*	5-13	P 1-3 **OR** G 2-1
BRIEF 3	*Can resistance be transmitted to other bacteria?, under which circumstances?*	14-16	P 1-3 **OR** R 2-1
BRIEF 4	*How can bacteria resist to antibiotics?*	17-22	G 1-3

Briefly, as it is apparent in the hierarchichal table of questions and corresponding contents, the global task requires for the students to cover and integrate the 4 content areas through the inspection of the 3 texts, extraction of relevant segments and integration of them. It is the only kind of task with more possibilities to allow the reader the construction of an integrated mental model on the topic of our interest. In this first experiment, we will contrast the global task to 4 brief questions which focus students' attention on the 4 content areas, though it promotes the isolated comprehension of the areas of contents. Because of this, they will not produce deep learning and integration of contents.

Similarly as we discussed when detailing the distribution of contents across texts, it might be argued that our claim that brief questions focus students' attention only on one of the texts is not true. In fact, from the *hierarchical table of questions* it might seem that two brief questions (brief 2 and 3) could be answered from two different sources, thus promoting some kind of integration. As we already clarified, there are two cases in which contents overlap between texts. But this overlapping relationship implies repetition of contents and not information needing to be combined. Therefore, it would not matter for a reader to select the contents from one or another source. Thus, operatively, contents are present in only one source and students' attention when answering is directed to an isolated source of information. As we have signaled in the hierarchical table of contents, the relationship between overlapping contents is exclusive (i.e. OR).

Once detailed the main contents which make up the components of a mental model explaining why and how bacteria become resistant to antibiotics, and once justified how we elaborated the questions and the reasons why we expected that a global task would be the one having more possibilities to produce integration, we will clarify one important issue regarding the questions students answered in the experimental phase. And this issue is precisely how questions were scored in order to obtain a performance measure for each of the participants. Obtaining a clear *performance measure* was indeed part of our set of hypothesis regarding the benefits of different kind of tasks in different learning levels. We start with performance, which accounts for how well students comply with the demands of the task.

To measure performance, we again had the table of units of information across texts as reference. In fact, the 22 main idea units making up an hypothetical to be constructed mental model, were the ideas students were expected to include in their answers, whether it be in a single answer (i.e, global task) or in four answers (i.e, brief questions). Therefore, to score students's responses and obtain for each of them a global performance measure, equally for all conditions, we counted the number of different idea units from the table of contents that they had included in their responses. Thus, if a student perfectly performed the global task, including the 22 ideas in their answer, this student would obtain for performance 22 points. Performance scores were transformed into percentages, thus the previous case would have obtained a 100% of success in performance.

The scoring procedure was conducted by two experimenters. To guarantee equality in criteria, at least two question sets were scored and compared among scorers. A common agreement we reached is the degree to which idea units of the table of contents would match how students included these ideas in their answers. Generally, we considered that there was match when ideas were expressed: a) with very similar words, nearly the same wording or, b) paraphrasing the idea unit, though expressing the same content. Discrepancies would be discussed and solved. This training scoring phase ended when discrepancies were below 5% of the total amount of corrections and answering sets were scored separately.

True/False recognition&understanding of ideas test. After explaining in detail the texts, contents and associated questions we used to make students learn from a multiple expository text working session, it is time to specify how we measured the degree to which students had profited from the task. First of all, we constructed a final true/false item questionnnaire to measure if students were able to identify the idea units present in the sources and understood what they exactly expressed. In general, we are aiming at measuring an isolated level of learning, a kind of text-base measure of the components that would make up a mental model of bacteria resistance to antibiotics. Given that the isolated understanding of these components would not guarantee integration of contents, we considered that this true/false final learning text would only measure a surface construction of a mental model from the documents, a kind of structure with no integration among ideas.

The True/False recognition&understanding of ideas test consisted thus of 18 items that directly matched the main contents that students were supposed to have gone through in the experimental session. This matching was not similar in wording, but it sometimes included a paraphrasis relationship or even some kind of transformation from the original idea to assess understanding. In general, each of the items directly matched by means of the above relationships with one of the 22 ideas present in the table of contents, though some items included more than one unit of information from the original table of contents.

As examples of some items in this true/false questionnaire, we have the followings: *“Plasmids are DNA pieces”* or *“Plasmids produce resistance mechanisms against antibiotics”.* The first of them is false and the second one is true, and both of them directly correspond to ideas 11 and 12 in the table of contents. Had the students identified and understood these two ideas in the experimental phase, they should have no problems in correctly answer them for the true/false questionnaire.

Application&transfer test. To conclude, the last of the measures we used to assess the quality of the construction of an integrated model from the three sources was an *aplication&transfer of knowledge practical case*. By using this kind of test, we were trying to have a deep quality of learning measure. Because this test made students apply an integrated knowledge of how bacteria become resistant to antibiotics to a new but related situation, we assumed that this would be the only kind of measure that would indicate integration across contents and sources. Therefore, the kind of task yielding higher scores on this test would be the one effectively creating integrated and long-term learning from multiple sources.

To measure application&transfer of knowledge, students were presented with the following practical situation: “*Imagine you have a sore-throat and you go to the doctor to have a treatment prescribed. The doctor informs you that the origin of your illness is a bacteria infection. You should take antibiotics for a whole week. You start your treatment, following the doctor’s advice, but after three days you feel perfectly well and decide to stop taking antibiotics”.*

Based on this practical situation, students were required to answer the following questions: "*1- Do you think the antibiotic has been successful in destroying all the bacteria responsible for your illness, why?", "2- Which kind of explanation would you give?", "3- What kind of further implications would there be because of you stopping to have the antibiotics*?

Clearly, students should relate their new acquired knowledge of how and why bacteria may resist to antibiotics and under which circumstances to a new situation. Those students succeding in this task would be the ones who had integrated the isolated contents on bacteria resistance present across three sources, because they had been able to apply it to new contents, showing deep understanding. This *Application&transfer test* was socored up to a maximum of 9 points, and a percentage of success was obtained for each student. The scoring system also implied two experimenters and a training-phase in which discrepancies were discussed and solved.

Think-aloud registration devices. Finally, to obtain the think-aloud data of half of the sample in this Experiment 1, we used a microphone connected to the computer and the recording software Microsoft provides in Windows XP ®. Audio outputs for each student were transformed to MP3 files for practical reasons.

Procedure

Students taking part in Experiment 1 were assessed over two sessions carried out in two different days. Broadly speaking, session 1 was used to obtain the control measures scores and to train students to use the software *Read&Answer*, whereas session 2 included the whole experimental phase, both the process of reading multiple texts and answering questions and the final learning assessments. However, some considerations should be pointed out regarding the think-aloud and no-think-aloud participants' procedure.

Think-aloud participants were assessed individually, because of the requirements of the think-aloud procedure. In session one, they were first randomly assigned to one of the two experimental conditions, that is, performing a global task or answering 4 brief questions and then they performed the previous background knowledge test. After that, they were assessed on writing speed on the keyboard, having 2 minutes time to copy in a Word® document all they could from another text which was also presented in the same Word® document. We also obtained the lexical access measure, by making students read aloud a list of 40 words and a complementary list of 40 pseudo-words.

What was specific for think-aloud students was the need to verbalize their thoughts while performing the task. A specific training should be conducted to teach students say whatever came to their minds during performance. Because externalizing short-term memory contents associated to performing a task is not something students are used to doing continuously, we carefully designed a training and modelling phase for students to clearly know how to think-aloud while reading the texts and answering the questions. All this training and modelling was conducted in the first session.

The training and modelling phase had two main objectives. First, to familiarize students with the software *Read&Answer* and the performance sequence they were expected to follow.

Secondly and specific for the think-aloud group, to learn how to verbalize anything that would came to their minds during performance. In order that think-aloud participants would get used to the software and exactly knew how they were expected to answer the questions and read the multiple texts, we conducted an initial training in which, by using a *Read&Answer* example of multiple texts reading and question-answering, we instructed students how to read the multiple texts, how to change texts, how to move from one page to another in the same document and, finally, how to access the question-answering screen and go back to the text screen in order to inspect the sources. We explicitly indicated the participants that a brief summary of how to to perform the task and the instructions they were expected to follow would be present in a single-page document, called Instructions, which would preceed the three texts in the software.

Our instructions included a specific requirement, which was that students should start by reading their task before reading and inspecting the different sources. The reasons for this were two. First, to make performance homogeneous among participants, so that everybody would start their sequence by reading the requirements of the task they would be confronted with. Secondly and most important, to comply with our experimental hypothesis, which focus on multiple text processing guided by a task.

Once the think-aloud participants had completely understood the requirements of the task and the instructions they would have to follow, we concentrated on teaching them think-aloud. For this purpose, a twofold instruction was designed, which included both explicit training and modelling.

The explicit training consisted of a detailed explanation of what verbalizing during performance may imply, that it would be only saying aloud whatever thoughts crossed their mind while reading the texts and answering the questions and that they were not expected to produce an extra effort to verbalize, as the only thoughts that they should say aloud would be those that would naturally appear while performing the task. Therefore, we were especially interested in trasmitting the idea that verbalizing the thoughts would not imply extra effort, thus no extra processing, and that students should behave naturally, as if they were on their own saying out loud what they are thinking, which is something any student or person has experienced at some point in their lives.

Additionally, the explicit training included specific considerations on how and when think-aloud in the experimental phase. Because the experimental design included reading of texts, movements in the software and writing processes, we should clearly indicate students how to verbalize. For this purpose, we gave students the following indications.

First, that they were expected to verbalize after reading each of the text paragraphs of the multiple sources. Because the texts were presented masked in the screen and students read the text paragraph by paragraph, we indicated the students that the moment to verbalize would be after reading each paragraph aloud. Reading of paragraphs would also be done aloud, in order to facilitate the tracking of students' verbalization when analysing the audio files. The second of the indications to think-aloud was that students should also verbalize while moving through the software (i.e, changing texts, going to the question-answering screen), only when any thought crossed their minds simultaneously to actions in the software. Finally, to obtain data on the writing process, we indicated students to verbalize before, while or after reading their questions, which would be also read aloud, and writing their answers. And only in case any thought crossed their minds in the question-answering process. This way, we would obtain representative think-alouds of the whole performance sequence and all relevant components involved in our design.

After the explicit training phase, we included a modelling session in which the experimenter, by using a reduced version of a multiple texts situation with questions to answer, modelled for five minutes how students would be expected to think-aloud. In order to make a homogeneous modelling, we anticipatory designed this modelling so that it would afterwards be performed exactly the same for each student being modelled. Moreover, we tried that a variation of types of verbalizations would appear, so that no student would be biased in how to think-aloud. For instance, verbalizations implying comprehension of information or repetition of contents should appear in the same proportion so as not to influence any student in any direction.

Therefore, session 1 was used for obtaining the control measures data and for training students both in think-aloud and in using *Read&Answer*. We needed an average of 1 hour to perform all our tasks. Session 2, on the other hand, took place no later than two days after session 1. Similarly to session 1, all the procedure was individually conducted due to the think-aloud methodology. Students were invited to seat in front of the computer and were placed a headphone with a microphone to record their verbalizations. This way, they started working on their own. No time limit was established, the only criteria for finishing being having performed the task they had been assigned to. Globally, students employed an hour and a half to perform the whole experiment. While participants were performing the task and thinking-aloud, we particularly tried that they felt as if they were on their own. For this purpose, the experimenter placed herself out of the student's sight and would only appear in case there was any eventuality. Only in those cases that students stopped without apparent reasons to think aloud, would the experimenter indicate the student: "*please go on thinking-aloud!"*. Finally, when students finished their multiple-texts task and once away from the computer, they performed the true/false recognition&understanding questionnaire and, finally, they answered to the questions from the application&transfer case. Thus, overall, this second session implied an average of two hours performance.

We earlier pointed out that the procedure slightly varied for those having or not to think-aloud. Those in the no-think-aloud condition, whether in the global or brief questions task, also went through two sessions with similar characteristics to the above described, except that there was no training in thinking-aloud. Thus, session 1 was used to obtain the control measures and to train students in the use of the *Read&Answer* software, whereas in session 2 students performed their experimental phase and solved the true/false questionnaire and the application&transfer practical case. However, differently to the think-aloud-condition, sessions were not conducted individually, but by grouping participants into blocks of three or four students simultaneously working under silent conditions in individual computers in the same room.

Dependent measures

In order to fully understand the whole experimental procedure and the data analysis that we have conducted to test our hypothesis, we should first clarify which measures we considered and how they relate to the main hypothesis in this study. We want to justify what we have analyzed so that it appears perfectly clear how these measures relate to broader objectives in this study. Therefore, this is the aim of the current section.

(a) Control measures. As we presented earlier, we obtained three independent measures to test the degree to which experimental conditions were homogeneous or not in a group of variables that may influence further analysis. Thus, we obtained a global score for each student in *previous background knowledge*, being the maximum that students could obtain in this test 15 points. We also obtained several measures linked to *lexical access*.

This way, in reading words we counted the *number of mistakes* students made when reading aloud and the *time in seconds* they needed to read the list of 40 words. Similarly, in reading pseudo-words we also obtained the number of mistakes and time needed to read the whole list aloud. Finally, the last of the control measures was *writing speed on the keyboard*. From this measure, we counted the *number of words* students were able to copy in two minutes time.

(b) Assessment of type of task effects on performance and learning from multiple expository texts. To test the main hypothesis in this study, that is, the effects of different kind of tasks on different learning levels, we used three kinds of measures which aim to correspond to increasing quality levels in the construction of an integrated mental model from the multiple texts. As we have earlier specified, not all kind of tasks would prove similarly effective in each of the learning levels.

The first of the learning levels was *performance* on the task and, according to some authors (i.e., Schmidt & Bjork, 1992), it should not be considered a learning level by itself, but a transitory effect due to an experimental manipulation. In any case, performance would indicate how well students had succeeded in their task and, therefore, would have some interest if related to other learning measures. We obtained a *performance score* for each student by counting how many relevant units of information from the 22 units spread across the three texts the student would have included in his or her answers. The global performance score would be finally transformed into a percentage score, to facilitate interpretation.

The second of the learning measures aimed at assessing a further step in the construction of a mental model of how bacteria produce resistance to antibiotics. As we explain earlier in detail, making student recognise and understand specific content items directly related to those units of information making up a mental model of bacteria resistance would indicate how well students had acquired and understood the isolated units of information. Though, it would not inform us if students had integrated this knowledge.

Thus, to measure this shallow or text-base dimension of learning, we obtained a global score from the students' answers to the *true/false recognition&understanding of ideas questionnaire.* The maximum score that could be obtained in this test was 18 points, corresponding to the number of items it contained, and this score was finally transformed into a percentage score.

Finally, we used a specific measure to assess the degree to which students had not only acquired and understood the isolated units of information making up a mental model on bacteria resistance, but they had also been able to integrate this knowledge into a coherent representation. Students suceeding at this stage would be the only ones having integrated the different units of relevant information presented in each text. Thus, to assess this deep level of learning we had the *Application&transfer practical case*. The maximum score in this test was 9 points which would be also transformed into a percentage to facilitate further interpretation.

(c) Measures of how students distribute their time in the experimental session. We earlier argued that, in case a global task would prove more beneficial for promoting integration from muliple sources, there should be some kind of on-line evidence demonstrating that students were actually integrating information. Or, globally, on-line indications that performing a global task is indeed a more effective way of working with multiple expository sources. This is precisely the aim of this set of measures.

For this purpose, we analyzed how students distributed their time across tasks in the experimental session. This way, we would have an overview of students' reading and question-answering behavior across time. Because *Read&Answer* provides us with an output appearing all actions that students perform with their corresponding time in seconds, this allowed us consider separately how long each student had employed to: a) perform the whole experiment (i.e., *Time in Experiment*); b) read the single-text page of instructions (i.e., *Time in Instructions*); c) read the three multiple texts (i.e., *Time in Text*); d) read the question/questions (i.e., *Time in Question*) and finally, answer the question/questions (i.e., *Time in Answer*). Consequently, these measures would globally indicate students' question-answering behavior in multiple texts guided by a task, whether a global task or 4 brief questions.

(d) Measures of how students read and integrate information from multiple sources. Similarly to the previous set of measures, the current ones aimed at assessing the degree to which a global task promoted integration of information from multiple sources in a greater extent than brief questions. These measures, thus, would provide strong evidence that combination of relevant units of information into a higher-order representation is actually taking place.

In assessing reading and integration from multiple sources, we exclusively concentrated on *reading speed rate*, believing that it directly indicates processing of information. Reading speed rate was obtained dividing the number of seconds a student had employed in a given textual unit by the number of words this unit would contain. Thus adjusting the amount of time devoted to a paragraph according to its lenght.

The first of the measures we considered was *the question reading rate*, which would indicate how fast or slow questions had been processed. This measure would not direct so much to integration but would inform how students process the demands of their task. Either students would have had one global task or 4 brief question, or either a student would have read the questions fewer or more times than other, we averaged the reading speed rate in questions, so as to have a global measure of how questions were read, independently of the number of visits or number of questions.

Secondly, we created a repeated-measures dependent variable. This was *Reading speed rate*, which was subdivided into *relevant reading speed rate* and *non-relevant reading speed rate*. As all participants would read both relevant and non-relevant information distributed across the three texts, this allowed us create a repeated-measures dependent variable, which would surely enrich further analysis.

The *relevant reading speed rate* was obtained by averaging the reading speed rates in all visits students had undertaken in the experiment to read relevant paragraphs. Let us remember that we had 4 relevant paragraphs, distributed two in one text and the other ones in the two different texts. Conversely, the *non-relevant speed rate* was obtained by averaging the reading speed rates in all visits students undertook to read irrelevant paragraphs. We strongly believed that this measure would shed light on the integration process.

Finally, to assess the degree to which students regulated their reading and integration process across time, we designed another repeated-measures dependent variable called *Reading rate across experiment*. It was subdivided into three components: *reading rate in time 1*, *reading rate in time 2* and *reading rate in time 3*. To obtain this measure, we primarily used the output sequence from *Read&Answer*, which provides a sequentially ordered account of all the actions the student has undertaken in the experiment and their corresponding time. From this sequence, we discarded all which was not reading of textual paragraphs, though keeping the sequential order. Then we divided the sequence into three thirds and calculated the reading rate for each of the segments making up each third. It should be noted that we discarded for each of the participants reading rates which were below one standard deviation from the mean reading rate each participant had. This way, we discarded all which was not reading of text according to the participants' reading speed. The reason for this was to obtain a clear pattern of the reading speed adjustment across the experiment.

We strongly believed that by maintaining only reading rates which would indicate processing to some degree we could have more interpretable patterns in terms of integration. We then averaged the reading rates in the first third (i.e., *Time 1*), in the second third (i.e., *Time 2*) and in the last third (i.e., *Time 3*). As we stated in the hypothesis section, it might happen that the student integrating information will read slower in Times 1 and 2, thus having higher reading rates. This would indicate that integration and processing of segments is taking place. Contrarily, it should happen that reading rate dramatically decreases in Time 3, thus indicating that the student is reading much faster, because he or she is only reviewing already processed and integrated information.

In conclusion, dependent measures in *section d* are mainly based on reading rate assessment and directly focus on the processing and integration of information from multiple sources. Therefore, in conjuction with dependent measures from section b and c, they should help us in giving a coherent explanation of how and why integration of information occurs.

(e) Think-aloud measures. Finally, the last measures we obtained were a group of *think-aloud measures* derived from a categorial analysis from the verbal protocols students produced. Our main interest in obtaining these kinds of measures was to deepen into strategic processing in multiple text reading and to obtain convergent data that would strengthen our hypothesis that a global task is more effective in learning from multiple texts.

For this purpose, we conducted a detailed categorisation and analysis of participants' think-aloud protocols. Obtaining and analysing verbal protocol data required that we carried out a very systematic analysis procedure, which followed the next sequence of steps. First, the transcription of audio recordings belonging to each student to a analysable format. Then, a decision-making process in order to establish which would be our main categories in analyzing the verbal protocols. Once established the main categories, the allocation of each verbalization in one of the analytical categories. Finally, a quantitative treatment of the data which would allow us interpret the significance of results.

In order to transcribe each participant's verbal protocol, we considered as segmentable units the individual ideas each student would produce in their thinking-alouds. By ideas we meant sentences, grammatically independent and understandable by themselves, having at least one subject and one verb. This way, each student's verbal protocol would be made up of as many segmentable units as ideas he or she would have verbalized when thinking-aloud. As an example, one student would verbalize the following units when trying to understand a paragraph from one text: "*Here they are mentioning that bacteria become resistant to antibiotics. Thus medicines stop being useful, when enough time has gone by and bacteria have become resistant. Then, farmacological industries have to be continuously innovating, because old medicines are not useful anymore*". Based on our segmentation criteria, we obtained 6 individual and codeable units from this verbalization.

Each of the students' think-alouds was segmented this way and by two experimenters, who initially segmented together, solving discrepancies, and then segmented separately. For this purpose, we used an Excel® file for each student in which we inserted each student's ideas in different entries. Finally, we would have separated Excel® sheets for each student, all prepared and ready to be coded.

Our main objective in coding was to establish a coding system which would reflect all the reading and question-answering process that took place in the experimental session. That is to say, our categories should reflect the following performance levels: a) The use of the software and performance of the task in an electronical environment, b) Text comprehension and the use of different reading strategies and, finally, c) The writing process involved in answering the question/s. It should be noted that this coding process was again conducted by two experimenters, which agreed on what and how to code after a training session in which discrepancies would be solved. After that, coding was conducted individually. The creation and application of the coding variables required an iterative process in which different proposals were raised and tested in a pilot verbal protocol. After at least more than three modifications of the initial coding proposals, we obtained the definitive coding system, including the following categories: Task verbalizations; search verbalizations; Relevant and non-relevant comprehension verbalizations; Superficial Text processing verbalizations; writing and superficial writing verbalizations.

- *Task Verbalizations (T)*: We considered Task verbalizations those units associated to the process of performing the assigned task and also any verbalization related to the use of the software, as we considered that verbalizing on the software was linked to performing a task in an electronical environment. For instance, a participant would start the experiment by saying aloud: *"these are the instructions, aren't they?, I read them aloud. Ok, I will start by reading the questions".* Obviously, the student was referring to the initial requirements associated to the task. In many cases, task verbalizations reflected an excessive concern about the use of the software and the superficial actions (i.e., reading text, answering questions) carried out in the course of the experiment, instead of focusing on finding and understanding the relevant pieces of information. For example, a student answering brief questions mainly verbalized on the task, producing verbalizations as the following: *"ok, going back to text to complete the answer, reding question..., I answer, now I go back to text one to look for information for question 3, I return to question...".* In general, we expected that this would be the trend present in subjects who received brief questions and thougt aloud. As we initially hypothesized that brief questions would promote a more superficial processing of text and no focusing on integration of information across sources, this would be apparent in their pattern of verbalizations. The contrary was expected for participants having a global task. Because they would understand the need for finding and understanding the relevant pieces of information, their think-alouds should reflect this concern.

- *Search Verbalizations (S):* Verbalizing on the process of searching for information would imply referring to searching and locating both relevant and irrelevant pieces of information. Searching and finding was indeed a key process for those who received a global task or brief questions. Due to the inherent characteristics of the experimental design, both groups were encouraged to search for information in the texts. Therefore, both groups would be equally expected to verbalize on the search process. However, it might be the case that brief question participants would verbalize more on searching for specific units of information, as the kind of task they were assigned to promote the location of units of information instead of integration. Search verbalizations were, for instance, when partipants continuously pointed out the need for searching for a specific piece of information *“ok...,I think this was located in text three, I will reread text three to find the answer”,* when they actually found it *“ohh, the three responses of bacteria against antibiotics, there they are” or* when they did not find it “*it is not here” there is nothing here for the question”.*

- *Relevant (Rc) and non-Relevant comprehension (nRC):* In order to reflect the degree to which participants would verbalize to understand and integrate information we established this twofold category. *Relevant Comprehension verbalizations* would be those produced in the course of reading texts and aimed at better understanding the relevant pieces of information and integrating them to build a higher-order representation of the problem why bacteria produce resistance to antibiotics. They would reflect an active processing of text and efforts to understand the information, for example by drawing inferences, elaborating summaries, giving explanations, etc. For instance, one participant answering the global task would verbalize the following way to better understand the mechanisms by which bacteria produce resistance to antibiotics: “*here they more or less give a different point of view of how resistance is produced. Given that* *they are mono-cells and with a tiny amount of genes, by mutation of these genes, they can create resistance, which is transmitted to their descendants".* Verbalizations to understand information could thus be made on relevant information, that is to say, on the group of relevant 22 units of information for the task. In this case, we called them *Relevant Comprehension verbalizations (RC)*. However, they could also be made on other non-relevant information, in which case we called them *non-Relevant Comprehension verbalizations (nRC)*. Participants answering a global task should be the ones verbalizing more to understand relevant information, because of the task promoting more the comprehension and connection of relevant units of information. On the other hand, non-Relevant Comprehension verbalizations should equally be distributed between global and brief questions participants. When considering neutral information, not relevant for the task, we expected that the global task group would not significantly outperform the brief questions group.

- *Superficial Text Processing (STP):* Complementary to the above category, we also considered verbalizations produced at such a superficial level that they were mere repetitions of the textual information read aloud, without contributing to a deep processing of the main contents relevant for the task. Thus, after reading the following textual segment "*Bacteria resistance is due to changes in the genetics information of the bacteria (mutation) or* *to the acquisition of new genetics information through plasmids. Plasmids are non-essential elements in the bacteria, which nevertheless provide them with great advantages to fight against antibiotics"* a student would say aloud *"it says that plasmids are non-essential elements. They provide bacteria with advantages to fight against antibiotics".* Obviously, the student was almost repeating verbatim what he or she had read aloud. This kind of verbalization would not contribute at all to understanding textual information at a deep level and was thus classified as superficially-related to comprehension. Superficial verbalizations were even completely unrelated associations to the textual information read aloud. As an example, another student would verbalize this way after reading a segment which presented bacteria that had produced resistance to diverse antibiotics *"ok....and now I access my ...what you call...background knowledge. Aha, I studied the staphilococcus when I was in high school..."* Thus, the student had verbalized by making completely irrelevant associations which did not contribute at all to understanding why bacteria become resistant to antibiotics. In general, we expected that STP verbalizations were more present in brief questions participants. Because brief questions fostered a more superficial revision of the sources, they would promote more STP verbalizations than in global task participants, who were mainly focused on reading, understanding and integrating relevant units of information.

- *Writing verbalizations (W):* In order to reflect the writing process when the students answered their questions, we included an additional category which would account for how students managed, controlled and regulated both the process of selecting the key information for a specific answer and the writing process in itself. Writing verbalizations would thus reflect self-regulation of the answering process "*I go to question 3, because I don't know what to answer in this one", "I think this one is correct this way", "ok...I think question 2 is not correct..".* On the other hand, verbalizations reflecting control of the writing process in itself *"Then. I start writing that bacteria may modify the antibiotic or degrade it. Comma, then...bacteria may also prevent antibiotics from working. Full stop".* We expected that both global and brief questions participants would equally verbalize on the writing process, as writing an answer or several answers was a basic process in both groups.
- *Superficial writing verbalizations:* Complementary to the above category, we included a final type of verbalization which would reflect a superficial management of the writing process and, in some cases, a wrong selection of the answer. By superficial management of the writing process we mean an excessive concentration on minimalist dimensions when writing such as

spelling “does it have an accent or not?”. By wrong selection of the answer we mean such cases in which students verbalized on selecting and using a wrong piece of information for an answer or even produced some irrelevant associations while writing the answer which did not contribute to its correctness at all. This category would thus be parallel to the Superficial Text Processing category applied when thinking-aloud on textual information. Therefore, we may expect similar results as in that previous category, in which we expected brief questions participants to verbalize more superficially, because of the type of task that would promote this kind of processing to a greater extent, in contrast to global task participants. Similarly, we may expect brief questions participants to produce superficial verbalizations when answering to a greater extent.

We have described so far which categories we used to analyse each participants’ verbal protocols and be able to obtain interpretable data from them, useful in terms of providing convergent data to other measures we use in this study to prove that there are tasks that are more effective than others in promoting integration and learning from multiple sources. Therefore, verbal protocols were coded by two experimenters following the above criteria. This way, we obtained for each student an accumulative score in each of the categories based on the counting of the number of times each unit in the protocol was coded using an specific category.

(f) Visual analysis of the combination of information process. To conclude, we decided to present a very specific complementary case analysis of how students actually combine the different units of relevant information in the multiple texts situations. The idea to conduct this case analysis came from the limitations of previous measures. In fact, none of them clearly indicated how students manage to combine the different relevant units of information across the reading of multiple sources and how this combination varies depending on the task.

Therefore, we concentrate on the combination of different, not the same, relevant units of information, which is indeed the key element in integration. Previous measures, such as relevant speed rate in comparison to non-relevant speed rate, are based on the averages of reading rate per word in a group of relevant paragraphs students visited. It might have been the case that one student visited the four different relevant paragraphs in the three texts and the mean reading rate in these relevant paragraphs actually represented the combination of information, or, contrarily, it could have happened that this student only visited the same one or two relevant paragraphs, reading and re-reading them all over the reading sequence. When obtaining the mean reading rate per word, it would not indicate this fixation and lack of combination.

Hence, because integration mainly involves the combination and connection among different, not the same, units of information, we decided to present a very specific informal case analysis which would illustrate our idea of integration as combination of different units of information and would try to differentiate the reading process for the global task and the brief questions, respectively. For this purpose, we selected one participant from each of the experimental conditions (i.e, Global-think-aloud; Global-no-think-aloud; Brief-think-aloud; Brief-no-think-aloud) and we analysed in detail the reading sequence across the experiment from the multiple sources. The criteria for selecting this single participant in each of the conditions were that they were *high-learning participants.* By high-learning we mean that their score in the final learning measures, both true-false and application case, was a high score, in comparison to the other participants (i.e, more than 60 %). With this kind of selection we tried to reflect cases in which combination of information, guided by task, would probably appear more clearly.

In each of the selected cases we concentrated on one of the outputs *Read&Answer* provides. Because we were interested in the temporal ordered reading of text paragraphs across the experiment, we used the output sequence to track the reading order of textual segments across the experimental time. As we presented earlier, the output sequence provides a sequentially ordered list of all segments active at a given moment, including readings of text segments and visits to questions and answers. As we were only interested in tracking the combination of information, we discarded from the sequence all which was not reading of text. Once we had the sequentially ordered list of textual paragraphs students visited in the experiment with their corresponding time in seconds, we calculated the reading rate per word for each segment by using the number of words of each paragraph. Similarly as in the reading rate across experiment analysis, we discarded for each of the participants reading rates which were below one standard deviation from the mean reading rate in each of the participants, with the purpose to obtain a clearer view of the reading process, eliminating quick visits to the textual segments. When we had the reading rate per word sequentially ordered and paired to the textual paragraphs students had visited, we obtained a *linear one-bar graphic* based on the ordered reading rates per word which would represent both the order and speed at which students had read the different textual paragraphs. As we were mainly interested in observing how the combination of information would take place, we should make visible the reading of different units of information in the *linear one-bar graphic*. For this purpose, we colored the textual paragraphs represented as reading rate in the graphic according to the type of paragraph it was.

All non-relevant textual paragraphs in the three texts were painted in grey. Let us remember that we had four relevant textual paragraphs distributed across three texts. Paragraph *P 1-3* (i.e., Perspectives, segment 3, page 1) was colored in red. Paragraph *G 1-3* (i.e., Genetics, segment 3, page 1) was colored in dark blue and paragraph *G 2-1(*i.e., Genetics, segment 1, page 2) was colored in light blue. Both *G 1-3* and *G 2-1* sharing the main color to indicate that they both belonged to the same text. Finally, paragraph *R 2-1* (i.e, Resistance, segment 1, page 2) was colored in green. The aim of this different colorings was not trivial at all, but would help in visually appreciate the existence of combination of information or not.

Therefore, by analysing very specific cases using an informal methodology we wanted to provide one more complementary source of data to understand what integration of information consists of. As we believe integration mainly involves the combination of different units of information and as we had hypothesized that this process would occur in a greater extent in the global task, this should be apparent in the present case analysis. We would understand the combination as the presence of re-readings of the different relevant paragraphs across the reading sequence and the progressive lack of non-relevant readings in the last stages of the reading sequence. Contrarily, because the brief question-answering process would promote an isolated concentration on units of information to answer the independent questions, combination of information as defined above should not occur. Indeed, we should expect to find a linear process of readings of the relevant paragraphs and no combined re-readings of the same segments across time.

Additionally, we expect to find a higher number of readings of non-relevant paragraphs even at later stages of the reading sequence. The reason for this expectation might be due to the question-answering process in brief questions. Because answering four independent questions may promote a more *inspection&search for isolated units of information pattern* (Cerdán et al., submitted), the probability to engage in readings of any paragraph, relevant or non-relevant in the experimental sequence would be higher than in the global task. On the other hand, differences depending on the presence of thinking-aloud or not should also be expected to occur.

In conclusion, this final analysis would only be aimed at illustrating how the integration process may take place in each of the tasks in terms of combination of information. It would be an informal analysis only to complement the systematic analysis obtained with the other on-line measures.

Now, after having explained in detail all the measures we used in this study to test our hypothesis and after having deeply justified the sense and scope of each kind of measure, we are prepared to report the statistical analysis that we conducted, all the results we obtained and how they relate to our main purpose in this Experiment 1 and to broader aims in the present dissertation.

2.2 Experiment 1: Results

(a) Control measures

To analyse the impact of a set of control measures that could interfere in subsequent results we conducted 2x2 Anovas, with independent variables *Think-aloud (think-aloud vs. No-think-aloud)* and *task (global task vs. Brief questions)*. The dependent variables were the control measures we described in earlier sections and from which we expected that they would not yield significant differences between groups: *Background knowledge, Time reading words, Time reading Pseudo-words, mistakes in words, mistakes in Pseudo-words and, finally, writing speed on the keyboard*. As expected, we found no significant differences in any of these variables *(see Table 1.1)*. Therefore, we had strong guarantees that experimental conditions did not differed in measures which could have probably interfered in subsequent analysis. Conclusions drawn on learning and integrating information from multiple sources will not be due to variations in background knowledge, lexical access or writing speed in the samples, but exclusively to our experimental manipulation.

	TABLE 1.1 CONTROL MEASURES IN EXPERIMENT 1											
	B. Knowledge		Time Words		Time Pseudoword		Mistakes Words		Mistakes Pseudoword		Writing Speed	
TA	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
G	9,55	2,98	31,27	6,62	44,45	7,63	0,00	0,00	1,36	1,86	67,27	16,54
B	10,25	2,01	28,58	4,21	43,92	7,33	0,08	0,29	1,00	1,28	64,00	13,02
nTA												
G	9,29	2,92	29.00	7,83	41,86	8,10	0,00	0,00	1,50	1,61	61,85	21,05
B	9,08	2,22	30,00	4,98	41,62	6,71	0,15	0,38	2,00	1,83	62,53	14,44

Table 1.1 Control measures in Experiment 1.

(b) Assessment of type of task effects on performance and learning from multiple expository texts.

We conducted the following analysis to test the main hypothesis of this study, that is to say, is a global task the most effective way of fostering deep comprehension and integration of information from multiple sources, rather than answering brief questions? If so, would there be differences between tasks in three increased quality learning levels? Moreover, would thinking-aloud affect in some way learning through tasks from multiple sources? To answer these questions we conducted 2x2 Anovas, with independent variables *Think-aloud (think-aloud vs. No-think-aloud)* and *task (global task vs. Brief questions)*.

The dependent variables were the three learning levels we hypothesized would be related to different learning outcomes depending on the task.

Briefly, *Performance* on the task or a measurement of how well students had succeeded in including the main relevant units of information from the three sources. Performance would measure a specific learning level, not related to long-term learning, but to transitory effects of the experimental manipulation (Schmidt & Bjork, 1992). It would be the shallowest measurement we would obtain to assess benefits from answering questions from multiple texts.

Secondly, we measured surface construction of a mental model from the documents by using the scores from the true*/false recognition&understanding of ideas questionnaire*. This second learning level would guarantee understanding of the main ideas making up a mental model on bacteria resistance to antibiotics, but it would not indicate that students had integrated this knowledge into a coherent representation. For this last purpose, we used the scores from the *Application&Transfer practical case,* which mainly measured the extent to which students had been capable of integrating from the three documents the units of information of how bacteria become resistant to antibiotics and could apply this knowledge to a new but related situation. This measure would be the one reflecting integration of information from multiple sources.

Finally, to test the degree to which the above dependent variables were actually measuring the same dimension (i.e, learning) though at different levels, we conducted Pearson Correlations among the Performance, True/false questionnaire and Application&transfer scores. We also included background knowledge socores just in case this variable would mediate learning to some extent.

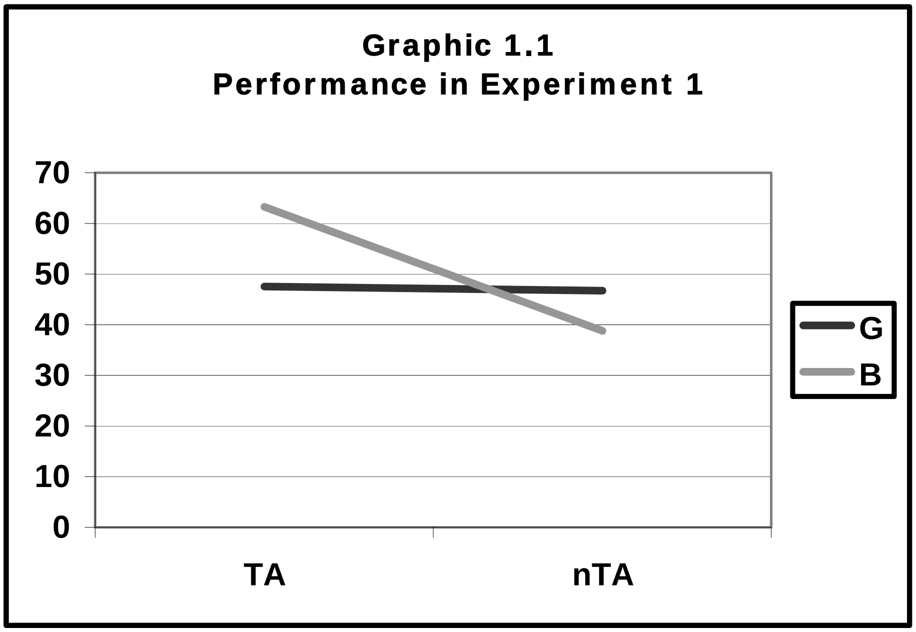
Performance results. The only result we obtained in *performance* was a marginal effect for the interaction, F (1, 46) = 3.12, p = 0.08. Thus, as can be seen in *table 1.2*, brief questions participants obtained higher performance scores, particularly when thinking-aloud. However, in the no-think-aloud condition, it was global task participants who slightly scored higher, though differences were not so marked as in the think-aloud condition.

TABLE 1.2
PERFORMANCE IN EXPERIMENT 1

		% Ideas in Answers	
		M	SD
TA	G	47,52	25,98
	B	63,25	25,45
nTA	G	46,75	22,58
	B	38,81	20,54

Table 1.2. Performance in Experiment 1

To have a clear view of this effect, let us have a look at *graphic 1.1.* Globally, the think-aloud condition seemed to have an increasing effect in performance, in comparison to the no-think-aloud condition. This increasing effect, however, was especially apparent for brief questions. Thus, brief questions participants who additionally thought-aloud obtained better scores in performance. This effect was neutralized when thinking-aloud disappeared. Moreover, there is a striking effect which is apparent after observing the following graphic: whereas global questions remained unaltered by the changes in thinking or no-thinking-aloud, brief questions were deeply affected.



Graphic 1.1 Performance in Experiment 1

We first interpret these results in terms of the type of task and, secondly, considering how the think-aloud method could have increased concentration on textual information. Thus, brief questions were designed to focus students´s attention on single documents and on specific areas of information. Students answering brief questions were globally expected to obtain better performance scores than students answering a global task, because the nature of brief questions directly lead students to the location of the main units of information on bacteria resistance to antibiotics. As they were directed to these units of information to a greater extent than global task participants, it was completely expected that they obtained better performance socores. Nevertheless, the dependance of brief questions on thinking-aloud was a surprising result for us.

It seems that thinking-aloud fostered concentration on the textual units of information in the three sources. This could be explained by the fact that students thinking-aloud were encouraged to read aloud this textual information and additionally verbalize whatever came to their minds in relation to it. We believe that this could have made students concentrate on textual information to a greater extent and thus include this information in their answers. However, it only happened in students answering brief questions.

Consequently, thinking aloud and answering brief questions which concentrated on the textual units of information had some kind of multiplicative effect which finally resulted in increased performance. On the contrary, a global task remained robust to this *textual-focusing effect* of thinking-aloud. If further results confirm our hypothesis that answering a global task is more beneficial for learning and integrating information from multiple sources, this would be an additional confirmation for the general benefits of global tasks in multiple texts situations.

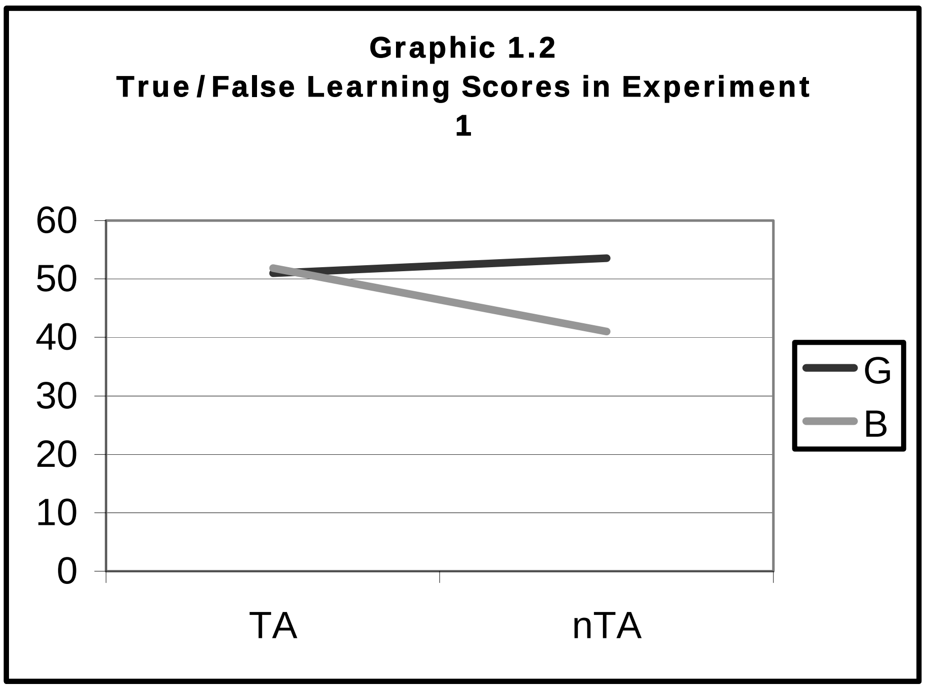
True/False questionnaire results. When measuring surface learning from the multiple sources, we found no significant results for any of the effects in the Anova. Though, there was a curious interaction trend: whereas true/false scores were very similar when Thinking-aloud for both the global task and brief questions participants, brief questions participants had lower scores in the abscence of thinking-aloud (see *Table 1.3*).

TABLE 1.3 FINAL LEARNING MEASURES IN EXPERIMENT 1				
	% True/False Questionnaire		% Application case	
TA	M	SD	M	SD
G	51.01	30.10	40.90	16.34
B	51.85	26.52	33.79	12.63
nTA				
G	53.57	22.70	53.96	15.47
B	41.02	23.62	40.59	11.19

Table 1.3 Final Learning measures in Experiment 1

This trend, though non-significant, reinforces the argument we raised in the previous section that brief questions in addition to thinking-aloud have a multiplicative effect. Thus, it seems that thinking-aloud forces the student to focus on a textual surface level based on the longer maintenance of the textual units in short-term memory, because of reading aloud and thinking aloud on the same units of information. Thinking-aloud, in addition to questions which promote concentration on surface textual information, reinforces the acquisition of these textual units at the same level as the global task but this effect seems to vanish as soon as thinking-aloud disappears. The hypothesized *multiplicative effect* may explain why brief questions participants scored equally to global task participants when thinking-aloud, but they had much lower results as soon as Thinking-aloud dissapeared.

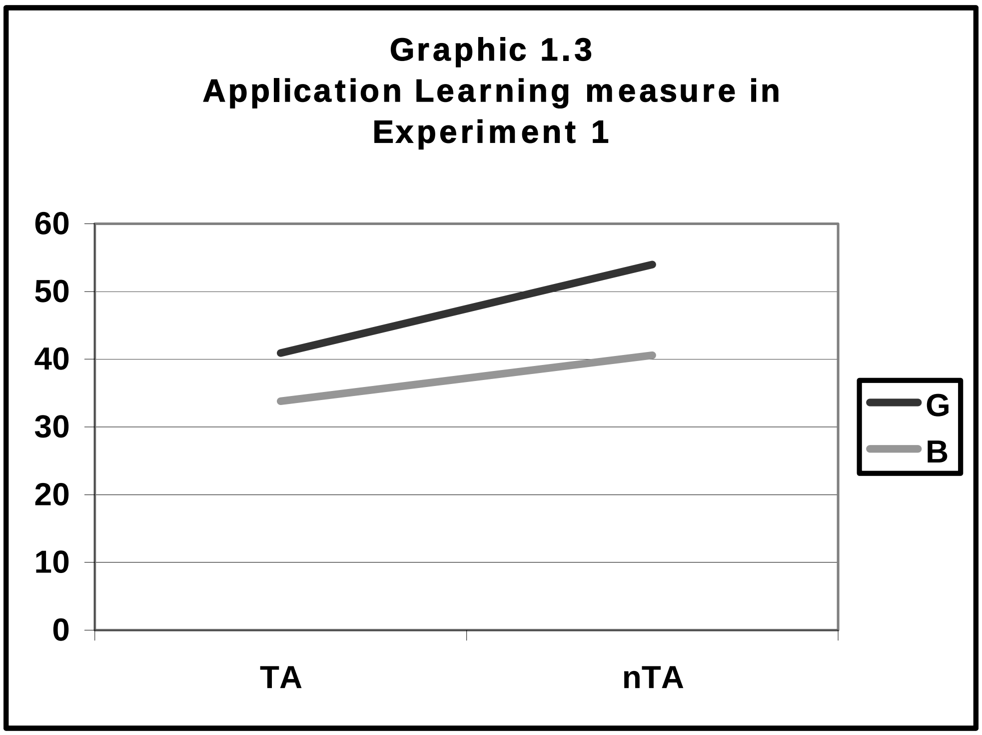
Therefore, when measuring a surface level of learning, based on the acquisition and understanding of the individual ideas making up a mental model of how bacteria become resistant to antibiotics, no significant differences were found between tasks. However, we did find a trend for the interaction which clearly indicated that brief questions were highly affected by thinking-aloud *(see graphic 1.2).*



Graphic 1.2 True/False Learning scores in Experiment 1

Application&Transfer practical case results. Finally, we present results obtained in the only dependent variable which would reflect deep comprehension and integration of information from the multiple documents. Both main effects were significant (*Table 1.3*).

Thus, the no-think-aloud group scored higher (*M* = 47.53, *SD* = 14.96) than the think-aloud-group (*M* = 37.19, *SD* = 14.96), *F* (1, 46) = 6.21, *p* < 0.05. On the other hand, participants answering a global task scored higher (*M* = 48.22, *SD* = 16.87) than participants answering brief questions (*M* = 37.33, *SD* = 12.16). These results can be graphically observed in the following *graphic 1.3*.



Graphic 1.3 Application Learning measure in Expriment 1.

Therefore, we have two strong, predicted and interesting results. First of all, we should comment on the learning effects of type of task. We had designed a multiple texts learning situation and used two kinds of tasks to promote integration of information and the construction of a coherent mental representation based on the three sources. We had hypothesized that the kind of task that would probably promote integration and learning at a deep level would be a global task, which would make students actively extract, combine and integrate different units of information into a coherent representation. Benefits derived from learning at a deep level would only be observed when using deep comprehension measures but not when other surface measures were used.

This is indeed the predicted result that we have obtained. Now we can state that, from the empirical evidence presented so far, it seems that a global task promotes integration and learning from multiple sources in a greater degree than brief questions, as it is apparent from the better scores in this application learning measure. On the contrary, in lower stages of learning (i.e., performance or surface recognition) this effect does not appear so clearly. Now we would need on-line evidence that effectively demonstrates that global tasks promote integration processes in the course of reading multiple texts. We present this evidence in subsequent analysis.

Secondly, it is highly interesting to have found that thinking-aloud hinders the process of learning from multiple texts at a deep level. However, it benefits the surface acquisition of textual units, mainly in brief questions, as apparent in performance and surface learning measures. Hence, there is some extraneous effect when students think aloud which disappeares when students do not verbalize, and which make the students concentrate on surface textual aspects, limiting the possibility of processing at a deep level and, consequently, hindering the process of learning and integrating information from the multiple sources.

We argued in the theoretical introduction that results concerning the effects of think-aloud were diverse, some reflecting neutral or beneficial effects in performance and learning when thinking-aloud (Crain-Thoreson, Lippman & McClendon-Magnuson, 1997; Fletcher, 1986; Loxterman et al. 1994; Coté, Goldman & Saul, 1998), and other studies showing the opposite trend (Wade & Trathen, 1989; Magliano, Trabasso & Graesser, 1999). In our case, we have found evidence for the hypothesis that thinking-aloud may in fact make students concentrate on surface information in text and on the other hand limit the effects of other higher-level processes (Magliano, Trabasso & Graesser, 1999).

Pearson Correlations among learning measures. To conclude this set of measures on the learning effects of different task when working with multiple documents, we included a complementary analysis to test the degree to which the three learning levels we hypothesized were related or not, and if relations among variables followed our theoretical predictions. For this purpose, we conduted *Pearson Correlations* among the following variables: *Performance scores, True/false questionnaire scores, application&transfer scores and, finally,previous background knowledge scores.* The reason for including previous background knowledge was to test if this variable somewhat influenced other learning outcomes. Results of this set of correlations are presented in the following table (*see Table 1.4*). Correlations marked with an asterisk were significant at 0.05.

TABLE 1.4
PEARSON CORRELATIONS AMONG OFF-LINE MEASURES
EXPERIMENT 1

	True-False	Application	Background Knowledge	Performance
True-False	1.00	.35*	.29*	.15
Application	.35*	1.00	.04	-.00
Background Knowledge	.29*	.04	1.00	.00
Performance	.15	-.00	.00	1.00

Table 1.4. Pearson Correlations among off-line measures in Experiment 1.

As we would expect, we found a significant correlation at 0.35 between the true-false questionnaire scores and the application&transfer measure. Because both variables measured learning, though at different levels, this result was completely expected, though it provided empirical confirmation to our claim that we were measuring the same dimension at different levels. We thus obtained a kind of validation to our learning measurements.

The second of the significant results was a surprising significant correlation between previous background knowledge and the true/false questionnaire at 0.29. Thus, previous background knowledge was related to scores in the true/false questionnaire, but not on deeper learning measures. This result may make sense if we consider that the true/false questionnaire mainly measured textual recognition of individual units of information. Students may have known some of these units in advance, though this knowledge did not affect the integration process as measured by the application&transfer practical case.

There was another result which, despite it was not significant, provided interesting information and theoretical validation to our learning levels. Whereas performance scores were negatively correlated with application&transfer scores at -.00, they correlated positively with true/false scores at 0.15. Thus, as we earlier hypothesized, performance accounts for changes due to the experimental manipulation which do not necessarily imply changes in other deeper learning measures. In fact, we found a negative relationship. On the other hand, performance was closer to true/false scores, which were still at a surface learning level. Consequently, performance was indeed inversely related to increases in deeper learning and closer to surface learning measures. This result, though non-significant, provided some support to the claim that deep learning is independent from performance on the task.

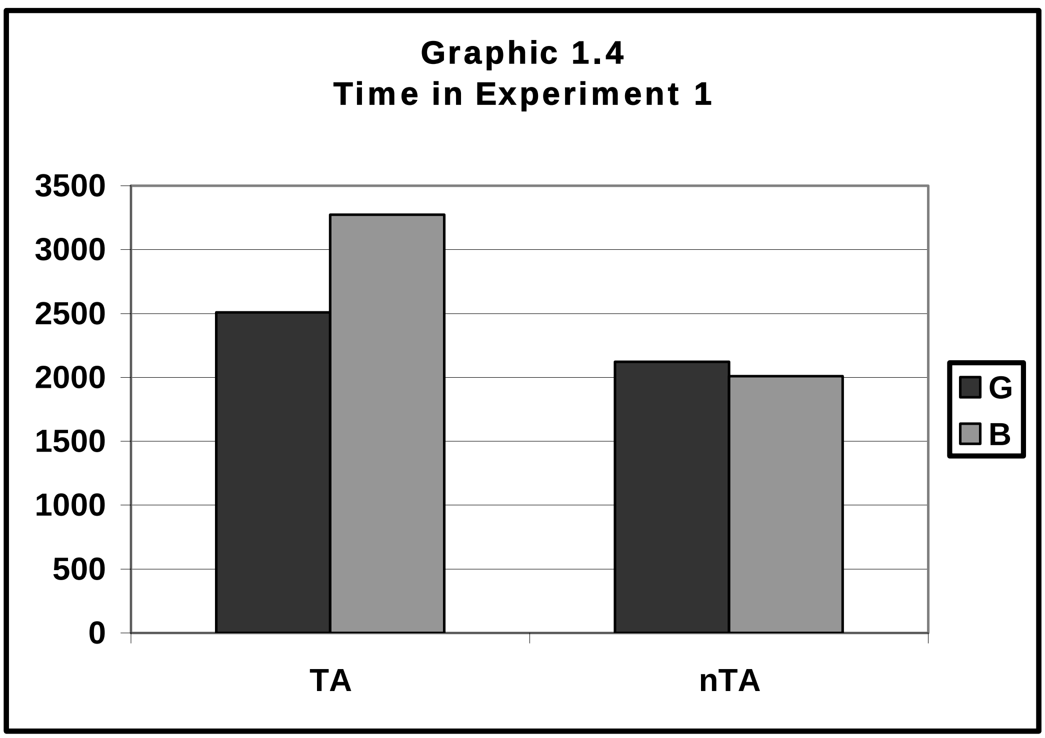
(c) Measures of how students distribute their time in the experimental session.

So far, we have obtained an empirical demonstration that global tasks promote integration of information from multiple sources and deep learning, in contrast to other kind of tasks such as brief questions, which direct the student to a more superficial processing of the texts. We have also concluded that making students think-aloud simultaneoulsy to performing the task seems to hinder higher-level processes and especially affects students answering brief questions, producing a kind of multiplicative effect, which focuses the student on a textual surface level in the multiple sources.

Having these main results in mind, we should start presenting possible on-line evidence which would justify, on the one hand, if a global task produces indeed integration of information from multiple sources and, on the other hand, how thinking-aloud may affect the process of answering the questions and learning from the three texts. For this purpose, we have time and reading-rate based measures which directly focus on how students behaved on-line in the course of the experiment. We will start by presenting a global approach based on time measures of how students distributed their time across actions in the Experiment: *Overall time in the Experiment, Time reading Instructions, Time reading Text, Time reading question/s and, finally, Time Answering the question/s.* With these dependent variables, we again conducted 2X2 Anovas, with independet variables *Think-aloud (think-aloud vs. No-think-aloud)* and *task (global task vs. Brief questions)*. Interpretable differences should be found according to type of task and thinking-aloud or not.

Overall Time in Experiment. We first wanted to have a global approach to how long students had needed to perform the whole experiment. This global time measure would include reading of multiple texts and the question-answering process. If global tasks were the most convenient tasks to work with multiple sources, they should not make the student spend an excessive amount of time performing them. Let us remeber that effectiveness is classically considered as something which promotes the maximum results using the minimum amount of time. On the other hand, we should expect that thinking-aloud would slightly increase performance times (Ericsson & Simon, 1993), because of the specific requirements (i.e, reading and thinking-aloud) that students in the think-aloud condition had.

Results for *Overall Time in Experiment* yielded two main effects. First of all, the global effect of Thinking-aloud was significant. Students thinking-aloud spent more time in seconds in the whole experiment (*M* = 2906.64, *SD* = 856.39) than students who did not think-aloud (*M* = 2067.27, *SD* = 750.11), $F(1, 46) = 14.19$, $p < .05$. Secondly, the interaction was also significant. Thus, brief questions participants needed more time for the experiment (*M* = 3271.81, *SD* = 655.09) than global task participants (*M* = 2508.27, *SD* = 898.24) only in the Think-aloud condition, $F(1, 46) = 3.99$, $p < .05$.



Graphic 1.4. Time in Experiment 1

Thus, results regarding the Think-aloud effect on global performance times followed our expectations, that is, thinking-aloud increased the time needed to perform the whole experiment in comparison to participants performing silently. However, according to the previous learning measures, this increase in performance time did not lead to better deep learning scores, but only some superficially text-based benefit from having probably focused on textual and explicit information, and only in the case of brief questions. Therefore, it seems that thinking-aloud slows down the process of learning from multiple sources, beacuse of the extra demands placed on students reading aloud and verbalizing their thougts. And, curiously, the more time invested in performance does not imply better processing from the multiple sources at all. Instead, it seems that there is a explicit concentration on textual information which may overload working-memory resources, limiting any higher-level processes.

Regarding the interaction effect of thinking-aloud and type of task, we found an on-line confirmation that brief questions are especially affected by thinking-aloud, whereas a global task remains robust to think-aloud effects. Therefore, as can be seen in *graphic 1.4*, there were only differences between tasks when thinking-aloud: brief questions participants thinking-aloud spent much more time performing the experiment than global task participants thinking-aloud. Consequently, it seems that answering brief questions under think-aloud conditions significantly increased the time needed for performance, whereas the global task was not affected by the presence or absence of thinking-aloud.

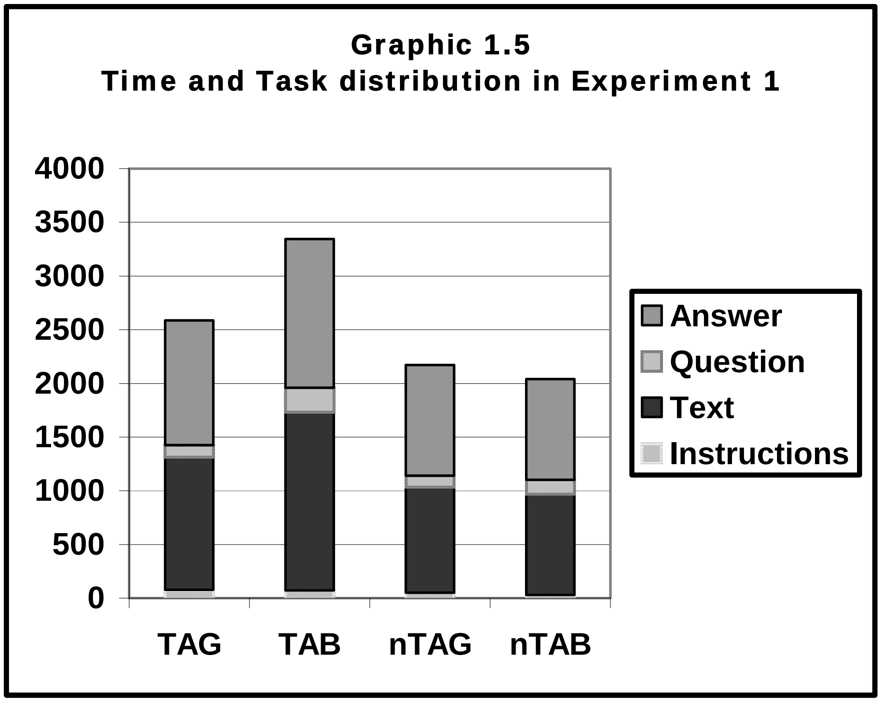
This result is also an on-line conformation that brief questions and thinking-aloud interacted in such a way that participants concentrated much more on textual explicit information, increasing their performance and superficial learning scores due to longer maintenance of textual information in short-term memory. However, and despite this longer time devoted to the experimental phase, no effects were found in deeper learning measures. Consequently, we confirm our claim that thinking-aloud and answering brief questions on explicit textual information in one source had a multiplicative effect which increased students' concentration on superficially-text-based information.

Time in Instructions, texts, question/s and answer/s. The overall measurement presented above splits up in the following tasks which were the units making up the whole experimental sequence. We expected to find differences in how students distributed their time across the experiment depending on task and thinking-aloud or not. On-line Time means and standard deviations can be observed in *Table 1.5* and are visually represented in Graphic 1.5.

TABLE 1.5
ON-LINE TIMES IN EXPERIMENT 1

		Time in Experiment		Time in Instructions		Time in Text		Time in Question		Time in Answer	
TA		M	SD	M	SD	M	SD	M	SD	M	SD
	G	2508,27	898,24	77,00	33,63	1236,38	456,26	112,74	63,19	1159,14	475,96
	B	3271,81	655,09	72,34	43,63	1658,40	382,51	230,52	81,29	1382,89	436,95
nTA											
	G	2121,14	870,47	49,44	21,65	985,34	589,93	104,10	113,09	1031,70	496,73
	B	2009,25	625,48	29,55	17,71	939,33	263,22	132,90	55,23	937,01	356,00

Table 1.5 On-line Times in Experiment 1



*** Note to Graphic 1.5: TAG (Think-aloud-Global task), TAB (Think-aloud-Brief questions), nTAG (no-think-aloud-Global Task), nTAB (no-think-aloud-Brief questions).**

Regarding the dependent measure *Time in Instructions*, which accounts for how long students had spent reading the single-page of instructions for the experiment,we only found a very expected result for Think-aloud. Thus, Think-aloud participants spent a longer time in seconds reading the Instructions for the experiment (M = 74.57, SD= 38.36) than no-Think-aloud participants (M = 39.86, SD= 21.94), F (1,46) = 16.76, p < .05. We consider this result as highly expectable because instructions for Think-aloud participants not only included a description on how to use the software, similarly to the no-think-aloud group, but on some recommendations on how to think-aloud. Therefore, it was completely reasonable that they needed longer to read these instructions.

As regards *Time spent reading the three texts,* which reflected how long students had spent reading and inspecting the multiple sources, we found a significant effect for Think-aloud and a marginal significance for its interaction with Task. Think-aloud students spent longer reading the multiple texts (M = 1456.56, SD= 462.86) than students who did not think-aloud (M = 963.18, SD= 454.46), F (1,46) = 14.83, $p < .05$, a result interpreted in light of additional time demands derived from reading and thinking-aloud simultaneoulsy. And, on the other hand, brief questions students thinking-aloud spent longer time in the texts (M = 1658,40 , SD= 382,51) than global task students thinking-aloud, (M = 1236,38, SD= 456,26), F (1,46) = 3.45, p = 0.07. However, this did not occur in the no-think-aloud condition.

Again, we find additional confirmation that answering brief questions and thinking-aloud had a kind of multiplicative effect which made the students concentrate longer on superficially-text-based aspects of the task. In this case, by spending longer time reading the multiple sources, not for deep processing purposes but sticked to the explicit and isolated ideas in the texts, as evidence by performance an superficial learning measures. Global task students, on the other hand, remained unaltered as regards time reading texts, independent from the presence or absence of thinking-aloud.

When considering the time devoted to process the task demands (i.e., *Time in question*) we found the following significant results. First, think-aloud participants spent longer reading the question (*M* = 174.19, *SD*= 93.47) than no-think-aloud participants (*M* = 117.97, *SD*= 89.54), *F* (1, 46) = 5.10, *p* < .05. This result followed the previous trend in that overall, think-aloud students needed longer for any task in the experiment due to the specific requirements of thinking-aloud.

Secondly, Brief questions participants also spent longer with the questions (*M* = 179.76, *SD*= 83.85) than Global task participants (*M* = 107.90, *SD*= 92.80), *F* (1, 46) = 9.71, *p* < .05. It could be argued that brief questions participants spent longer reading the questions because they had four questions to read and process, in contrast to one unique global task. This is partially true. However, if we consider it in terms of economy of resources, we could argue that the contents that the questions were asking for were exactly the same. Therefore, it was much more economical in terms of cognitive resources to read and answer one unique global task, which finally yielded the best deep learning scores.

Finally, the interaction between Think-aloud and Task was also significant, *F* (1, 46) = 3.57, *p* < .05, following a similar trend to previous results. Hence, Brief questions implied longer times in question (*M* = 179.76, *SD*= 83.85) than a Global Task (*M* = 179.76, *SD*= 83.85), and only in the Think-aloud condition. It seems, thus, that the multiplicative effect we have found in combining thinking-aloud and brief question solving also spread to the time reading the question. We mainly found this interactive effect for the processing of text, though it clearly extends to how students processed the question demands. Thinking-aloud required longer processing times because of thinking-aloud and reading aloud simultaneously and, complementary, brief questions required an iterative reading of four different questions instead of one. All of this explains why thinking aloud in addition to brief questions significantly increased the global time spent reading the question.

The last of the global time measures was the *time spent answering the question*, which would indicate the global amount of time needed to elaborate the answer. In this case, only the Think-aloud effect was significant. Hence, students thinking-aloud needed longer time to answer the questions (M = 1275.88, SD= 459.88) than students who did not think-aloud (M = 986.11, SD= 429.17), F (1, 46) = 5.15, $p < .05$.

This was an additional evidence of how thinking-aloud interfered in the process of answering questions from multiple sources, limiting deep learning. Given that think-aloud participants were instructed to verbalize in the process of answering the questions, this requirement obviously slowed down the answering process. Combining this result with the learning measures evidence, it probably slowed down the process by overloading working-memory resources, which were not available for higher-level processes anymore. On the other hand, there was a trend in that brief questions implied longer answering times than a global task only in the think-aloud condition, thus confirming the *multiplicative effect* also in the time answering the question.

To sum up, global time measurements have given us so far some on-line evidence for the two main questions in Experiment 1. On the one hand, thinking-aloud, which limits deep learning from multiple sources, seems to clearly affect the on-line behavior of students answering questions and reading multiple texts, by generally increasing the time needed to perform the experiment and its divisible components. Moreover, the effects of thinking-aloud dramatically increase in conjuction with brief questions. We have spoken of a kind of *multiplicative effect* which constantly appears in all the time-based analysis. This overall increase in time for students thinking-aloud and performing brief questions (*see Graphic 1.5*) completely explains and matches performance and learning effects. Therefore, the fact that brief questions students thinking aloud had better performance scores, superficial learning scores but no better scores in deeper learning measures give us the key to knowing what kind of processing did take place in the increased experimental time. We have already argued that thinking-aloud seems to have made students concentrate on superficial textual information and, in addition to brief questions which also concentrated on superficial units of information, exponentially increased students' superficial processing of the multiple sources, overloading working-memory and thus hindering any higher-level process.

Contrarily to brief questions and in parallel to performance and learning measures, the global task remains surprisingly unaltered by the presence or absence of thinking-aloud. This result is in fact a strong element to add to the better learning results after performing a global task from multiple sources. However, we wanted stronger empirical evidence to be able to state that learning results are due to the activation of integration processes in the reading of the multiple sources. Despite this global time measurements have provided us some evidence that performing a global task may in fact be more economical and effective (i.e., fewer time processing the question) we believe we need to explore further data to demonstrate that global tasks are more beneficial because they promote integration processes from multiple sources, in contrast to other kind of tasks. This is precisely the aim of the following group of analysis.

(d) *Measures of how students read and integrate information from multiple sources.*

As we have described earlier, the aim of the current measures is to directly assess the degree to which a global task produced a special processing of text which we could name integration. Additionally, we would like to further deepen into how the think-aloud method affected the processing of texts. For this purposes, we present a set of analysis based on reading speed rates which focuses on: (1) the processing time of the question and the processing time of textual information, relevant and non-relevant and, (2) the reading speed ajustment across the experiment (reading speed across experiment).

Processing time of question and textual information. First of all, to analyse how fast or how slow students had read the question, we conducted 2x2 Anovas, with independent variables *Think-aloud (think-aloud vs. No-think-aloud)* and *task (global task vs. Brief questions).* As evidenced in *table 1.6,* no significant differences were found for any of the effects. This result mainly indicates that differences between tasks were not located on how students processed and understood the task demands.

On the other hand, the presence of think-aloud seemed not to affect the processing of the question demands, though again we found a trend for brief questions in the think-aloud condition to increase reading speed rate, thus being processed slower (*Table 1.6*).

TABLE 1.6
READING RATE IN EXPERIMENT 1

		Question reading rate		Relevant reading rate		Non-Relevant reading rate	
TA		M	SD	M	SD	M	SD
	G	0,37	0,09	0,20	0,07	0,22	0,05
	B	0,43	0,10	0,18	0,05	0,20	0,09
nTA							
	G	0,35	0,26	0,21	0,09	0,13	0,05
	B	0,31	0,07	0,11	0,03	0,12	0,04

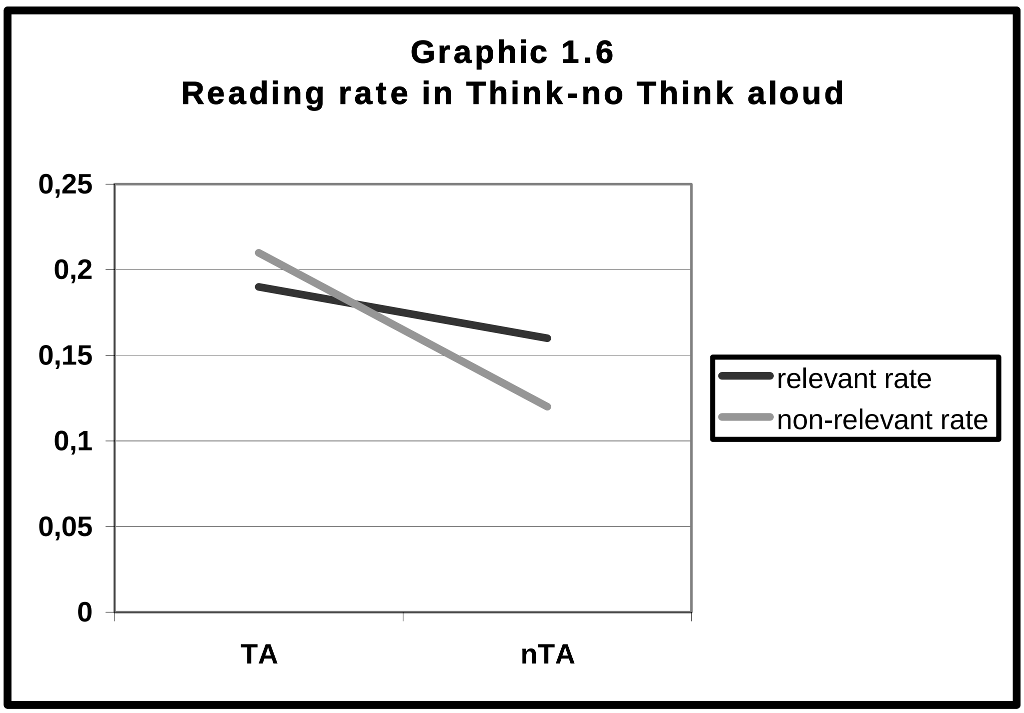
Table 1.6. Reading rate in Experiment 1

When considering *reading speed rate in textual information* interesting data arises. Let us remember that reading speed rate was a repeated-measures variable which included two sublevels: 1) relevant reading speed rate and 2) non-relevant reading speed rate (see *Table 1.6*). With this repeated-measures variable we conducted one Mixed Anova with between-subjets independent variables *Think-aloud (think-aloud vs. No-think-aloud)* and *task (global task vs. Brief questions)* and with the repeated-measures variable *Reading speed rate* (*relevant speed rate* vs. *Non-relevant speed rate*).

First of all, we will report between-subjects effects. The think-aloud effect was significant (*Table 1.6*). Hence, think-aloud participants, as expected, read more slowly (*M* = 0.20, *SD*= 0.06) than no-think-aloud participants, who read faster (*M* = 0.14, *SD*= 0.05), F (1, 46) = 12.09, p < .05. This result further indicates that thinking-aloud seems to affect the processing of texts, slowing down reading and, generally, increasing the time needed to perform the task. On the other hand, the task effect was also significant. Global task participants read textual information more slowly (*M* = 0.19, *SD*= 0.06) than brief questions participants (*M* = 0.15, *SD*= 0.05), F (1, 46) = 4.99, p < .05. This is a first on-line confirmation that the global task produces a significant different kind of processing of texts than brief questions. Indeed, it seems that a global task makes students concentrate on reading more slowly textual information, in contrast to the higher reading speed in brief questions.

Therefore, roughly speaking, the global task seems to induce a more detailed processing of textual information present in multiple sources, whereas the brief questions seem to promote a more superficial processing of the texts. If we made a parallelism to Mannes and Hoyes (1996) findings the global task would induce a more slowly processing of the texts because integration of multiple perspectives would be taking place, whereas this would not be occurring to such an extent with the brief questions. However, we would still need to know if the global task not only induces a more detailed processing of the texts, but the adjustment of reading speed depending on the presence of information needing to be integrated or not. This was in fact the main purpose for including the repeated-measures variable *reading speed rate*. Results are presented below.

Three interaction effects were significant. First, the *Reading rate x Think-aloud* effect was significant. Relevant and non-relevant info was read at a very similar rate in the think-aloud condition. However, this difference increased as soon as thinking-aloud disappeared, decreasing reading rate for non-relevant information (*Table 6 and graphic 1.6*), F (1, 46) = 7.43, p < 0.05.

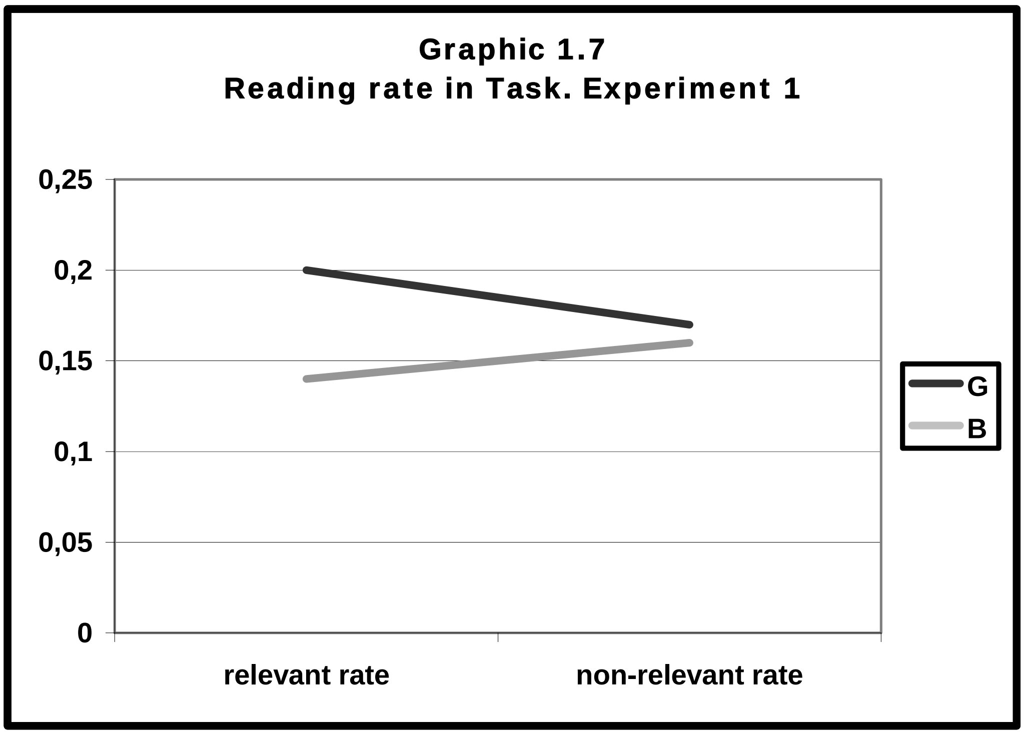


Graphic 1.6. Reading rate in Think-no-Think-aloud

Graphic 1.6 is clearly illustrative of the above mentioned interaction effect. Consequently, it is apparent that thinking-aloud induced an extraneous kind of effect which disappeared as soon as students did not verbalize their thoughts.

Thinking-aloud, thus, uniformizes the processing of texts, as evidenced by reading rates measures, making reading more slowly both for relevant and non-relevant information. This result is in agreement with the overall between-subjects effect that thinking-aloud globally made students read textual information more slowly. Contrarily, when students did not think-aloud the relevant textual information was processed more slowly than the non-relevant information. This would indicate that, overall, participants had an insight into which information should be read in more detail because it was the information needing integration, and which information could be read faster, because it included non-relevant facts. Nevertheless, we should find differences in this insight depending on the kind of task. This is precisely what the next significant interaction clarified.

The *Reading rate x Task* effect was, thus, also significant, $F(1, 46) = 4.98$, $p < 0.05$ (*Table 1.6 and Graphic 1.7*). Therefore, whereas global task participants read relevant information more slowly, brief question participants read both kinds of information at a similar speed rate. Consequently, global task students, appart from globally reading all textual information more slowly, they especially concentrated when they encountered relevant units of information to be integrated to construct a higher-order representation of how bacteria become resistant to antibiotics. On the other hand, brief questions participants seem not to discriminate so much the relevance of the information. One possible explanation to this result is that brief questions induced a more exploratory inspection of texts which made students read all textual segments at a similar speed to search for possible answers for the questions. Indeed, low-level question-answering has been demonstrated to involve more iterations searching for information in texts and a fast revision of textual information, in contrast to high-level question-answering, which promotes a more careful inspection of relevant units of information and fewer text iterations to perform the task (Cerdán et al., submitted).



Graphic 1.7. Reading rate in Task. Experiment 1

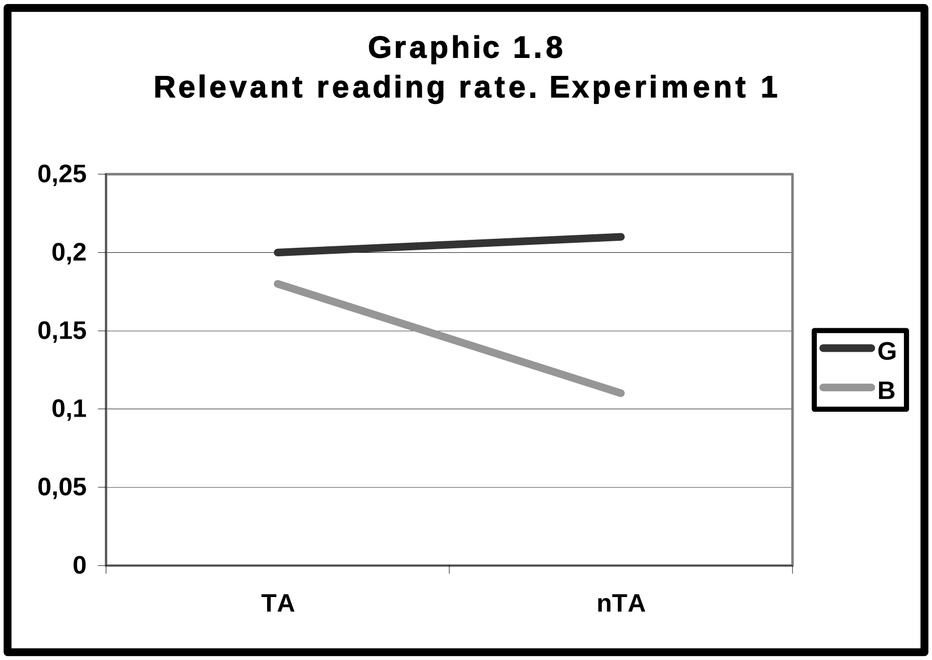
Thus, performing a global task focused students' attention on the relevant units of information to be integrated to fully understand why and how bacteria become resistant to antibiotics. The global task generally increased the reading rate in all textual information, but especially when students read relevant units of information.

If the first evidence was similar to Mannes and Hoyes (1996) finding that students read more slowly when they integrate information from multiple perspectives, this second evidence is also in agreement with other of the main findings in Mannes and Hoyes (1996). Thus, in Mannes and Hoyes (1996) students read textual information even more slowly when this information was new and hence needed to be integrated. This increase in reading slowness was explained in terms of the *reinstatement&integration strategy* taking place to connect old and new information. Similarly, the global task seems to induce integration of the relevant pieces of information making up a higher-order mental model of how bacteria become resistant to antibiotics. This integration process would be apparent in this increase in reading slowness to process and integrate relevant information in detail, which does not happen in brief questions.

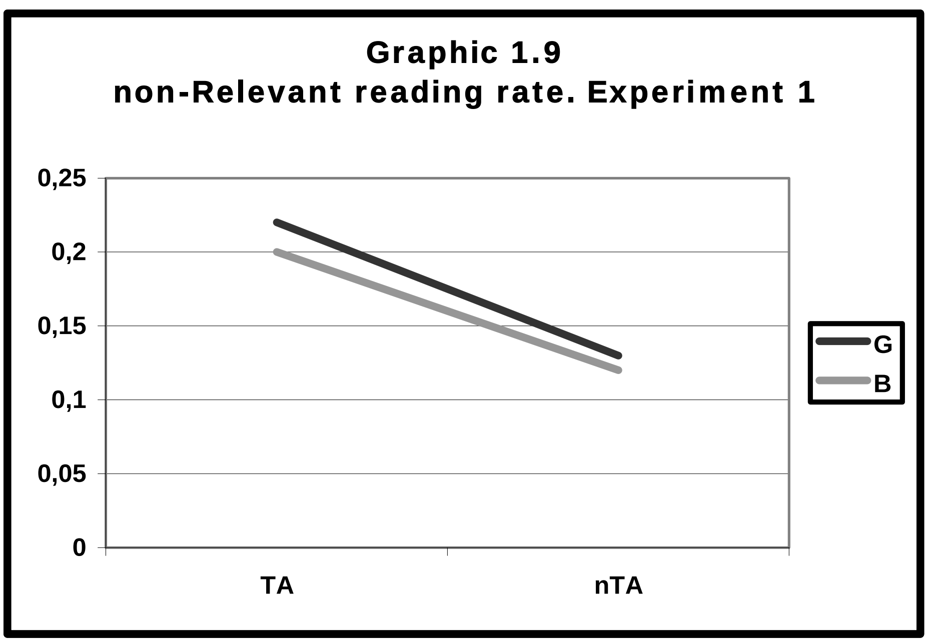
We still have a third significant interaction which reflects how thinking-aloud and task interacted with reading speed rate. Thus, the interaction of *Reading rate x Think-aloud x Task* was also significant, F (1, 46) = 5.79, $p < 0.05$. (*Table 1.6 and Graphics 1.8 and 1.9*). When reading relevant information (see *Graphic 1.8*) thinking-aloud made both global task and brief questions students read at a similar speed. However, as soon as thinking-aloud disappeared, brief questions participants dramatically decreased their relevant reading rate, reading this information much faster. Again and in continuity with data obtained so far, we found the *multiplicative effect* between thinking-aloud and answering brief questions.

In thinking-aloud, brief questions seemed to make student concentrate on relevant information as much as the global task. From this first evidence, we could have inferred that both tasks were equally effective promoting integration of information. However, this effect in brief questions was transitory and mainly due to the effects of thinking-aloud. Thinking-aloud generally increased the reading rate for all the textual information, including also participants answering brief questions. This effect disappeared when there was not thinking-aloud, which helped us conclude that brief questions promoting slow reading of relevant information did not mean information being integrated, but the interference of the thinking-aloud effect. On the contrary, the global task was again robust and remained unaltered to the effects of thinking-aloud. This helps us draw two main conclusions: a) that the increase in relevant speed rate in the global task was probably only due to information being integrated, and b) that the global task is robust to the effects of thinking-aloud in processing the information, which are present in a greater extent in brief questions.

On the other hand, when reading non-relevant information, the overall effect was that both the global task and brief questions students read non-relevant information more slowly in the think-aloud condition than in the no-think-aloud condition (*see Graphic 1.9*). Again, thinking-aloud made students read more slowly non-relevant information, in comparison to the absence of thinking-aloud. On the other hand, it is remarkable that whereas brief questions followed the same pattern as when reading relevant information, global questions participants adjusted their reading speed, reading faster non-relevant information but more slowly relevant information, as we reported earlier.



Graphic 1.8. Relevant reading rate. Experiment 1



Graphic 1.9. Non-Relevant reading rate. Experiment 1

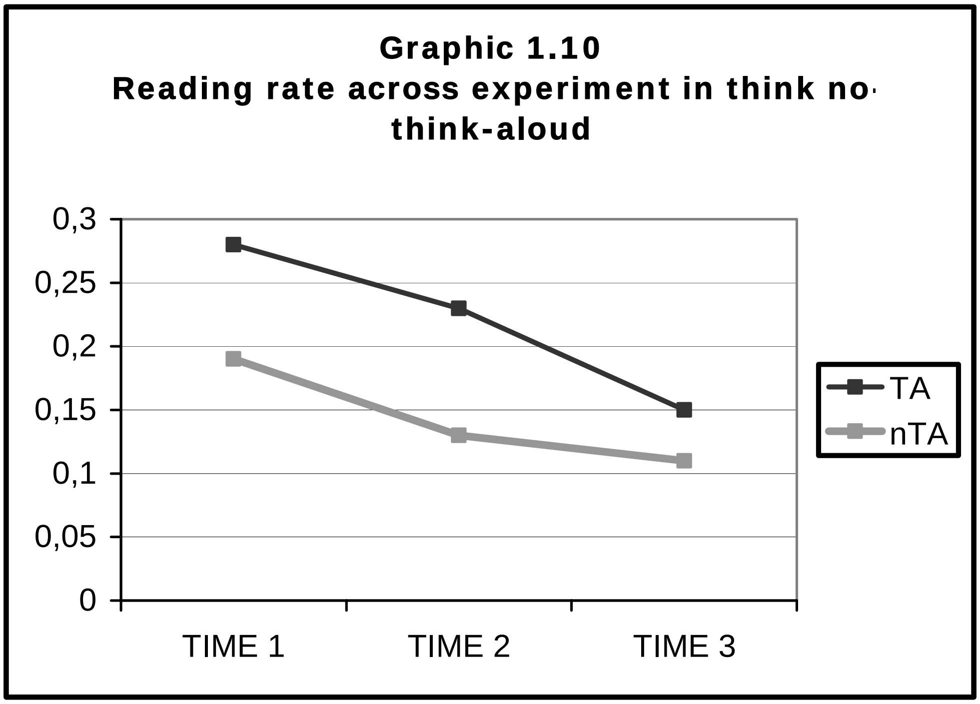
Reading speed adjustment across the experiment (Reading speed across the Experiment). The reading-time-based measures present in this section are aimed at providing on-line evidence that a global task induces integration of information to a greater extent than brief questions. And, complementary, find on-line evidence that thinking-aloud affects the processing of multiple texts. To complete these set of measures, we analysed how students adjusted their reading speed across the experimental time. We had hypothesized that integration would be apparent in students reading more slowly at earlier stages of the experimental time and significantly reducind reading speed at later stages in the experiment. This would indicate that information would be first processed in detail and integrated and then a faster revision of textual units would be undertaken, as if to consolidate the integrated mental model constructed in the first stages of the experimental time.

To measure this reading speed adjustment we had the repeated measures variable *reading rate across experiment*, which was subdivided into *Time 1* (i.e., mean reading rate in the first third), *Time 2* (i.e., mean reading rate in the second third) and *Time 3* (i.e., mean reading rate in the third third). We conducted one Mixed Anova, with between subjects independent variables *Think-aloud (think-aloud vs. No-think-aloud)* and *task (global task vs. Brief questions,* and the repeated measures variable *reading rate across experiment (Time 1, Time 2 and Time 3).* Means and standard deviations for this analysis are presented below (*Table 1.7*).

TABLE 1.7 READING RATE ACROSS EXPERIMENT 1				
	TA	TASK	MEAN	SD
TIME 1	TA	G	0.30	0.09
		B	0.26	0.12
	nTA	G	0.22	0.07
		B	0.16	0.06
TIME 2	TA	G	0.24	0.12
		B	0.23	0.12
	nTA	G	0.14	0.07
		B	0.12	0.04
TIME 3	TA	G	0.16	0.05
		B	0.15	0.08
	nTA	G	0.10	0.06
		B	0.13	0.05

Table 1.7. Reading rate across Experiment 1

There was one significant between-subjects effect, which is similar to the above presented results. Thus, think-aloud participants read globally more slowly (M = 0.22, SD= 0.08) than no-think-aloud participants (M = 0.14, SD= 0.05), F (1, 46) = 4.99, p < .05. This way, the think-aloud effect is strong enough to appear significant in all our different measures. Regarding the repeated-measures effects, we found two main interactions. First, the interaction *reading rate x Think-aloud* was marginally significant. In all time 1, time 2 and time 3 the think-aloud condition had higher reading rates than the no-think-aloud condition, F (1, 46) = 2.91, p = 0.09. (see *Table 1.7* and *Graphic 1.10*).



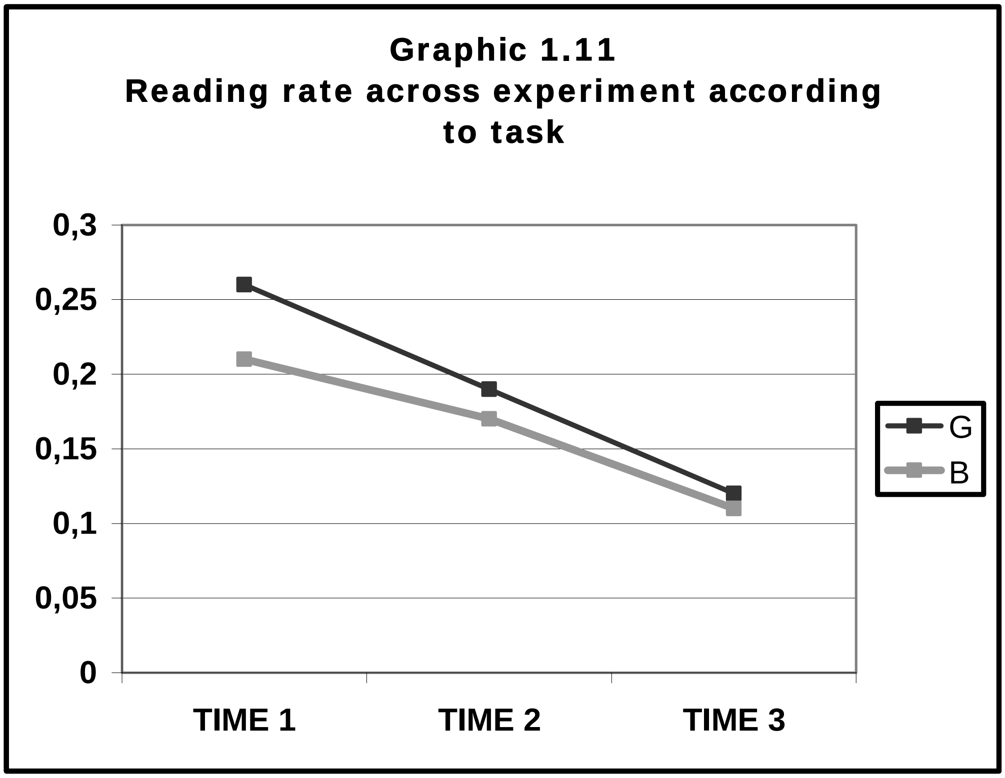
Graphic 1.10. Reading rate across experiment in think no-think-aloud

Moreover, whereas in the think-aloud condition the biggest reading speed decrease occurred from time 2 to time 3, in the no-think-aloud condition it was from time 1 to time 2. Additionally, in the no-think-aloud condition the decrease in reading speed across time was progressive and smooth, whereas it was more drastic in the think-aloud condition. What these data could tell us is that, in the presence of thinking-aloud, students seem to need more time to process and try to integrate information from the multiple sources, all time 1, and always at a higher rate than no-think-aloud students. The decrease in reading rate is only produced after time 2 to time 3, the last third of the reading sequence, and again at a higher rate than no think-aloud students and having a more marked decrease.

Differently, no-think-aloud participants are able to start reducing their reading rate already during time 1, and their reading rate decrease is smoother than think-aloud participants.

Hence, the course of the reading sequence in no-think-aloud participants experiences a progressive and expectable decrease in reading rate, due to the initial acquisition and processing of the textual information (i.e, Time 1) and the subsequent progressive decrease derived from recognising and re-reading information already processed and, probably, integrated. Contrarily, the course of the reading sequence in the think-aloud participants seems to be affected by the additional requirements placed on students because of thinking-aloud. Not only is the overall rate higher in all times, but also the time needed to process and try to integrate information is longer (i.e., all time 1 and part of time 2 to time 3). Consequently, there we have again additional evidence that thinking-aloud interferes with the processing of multiple sources, in this case by slowing down the time needed to process and try to integrate information in the course of the experimental session.

The second of the repeated-measures significant results was the interaction *reading rate x task*, F (1, 46) = 4.12, $p < 0.05$. As apparent in *Graphic 1.11*, global task students read the textual information more slowly in time 1, time 2 and time 3. Moreover, there was a progressive linear decrease in reading rate from time 1 to time 3. Contrarily, brief questions students always read textual information faster than global question students. And, similarly to think-aloud students, they did not start significantly reducing the reading rate from the very beginning, but only after time 2 was this decrease more marked.



Graphic 1.11. Reading rate across experiment according to task

According to our hypothesis for this measure, a progressive decrease in reading rate across the experiment would be another empirical sign of integration taking place. Hence, students would be processing and connecting textual information in their first readings, thus reading more slowly, and from then on, they would read faster because they would review and recognise already processed information.

Generally, it seems this is a common and expectable pattern which appeared both in global task and brief questions students, and also in think-aloud and no-think-aloud participants. All participants significantly decreased their reading speed rate, thus reading faster, across the experiment. Hence, we should say that all participants, especially both global and brief questions participants, engaged in processing and integrating processes to some extent. And this may be partially true. Indeed, the fact that brief question participants were not able to score significantly better in deep learning measures did not imply that they were not able to integrate information, at least to a minimum. However, it seems that the integration processes occurred in a greater extent when performing the global task.

This global task made in fact students read textual information more slowly from the beginning of the experiment to the end and progressively decrease reading rate when re-encountering already processed information in the course of the experiment. Moreover, the global task made students concentrate on reading relevant information more slowly, in comparison to non-relevant information and the global task remained robust to the extraneous effects of think-aloud, which clearly interfered in the process of answering brief questions. All this evidence supports our claim that a detailed processing and integration of information takes place to a greater extent when performing a global task, but not so much when answering brief questions. However, this does not imply that brief questions do not have some kind of beneficial processing effect, at least to clearly identify and understand the main points rose in the documents.

After reaching this preliminary conclusions from the on-line data obtained so far, it is time to present the set of convergent data obtained from the think-aloud protocols. Based on other kind of on-line measurement, we aim at providing complementary evidence to the beneficial effects from answering global tasks and, additionally, reflect how the reading and answering process takes place in both kind of tasks.

(e) *Think-aloud measures*

To analyse students' verbalizations when performing a global task or answering brief questions, we had established a set of categories which would reflect: a) the use of the software and performance of the task in an electronical environment (i.e., *Task and Search* verbalizations), b) Text comprehension and the use of different reading strategies (i.e., *Relevant, non-relevant comprehension and superficial text processing*) and, finally, c) the question-answering and writing processes (i.e., *Writing and superficial writing verbalizations*). We expected that these measures would allow us deepen into the process of performing different task in multiple sources.

With the above categories as dependent variables, we conducted one-way between-subjects Anovas, with independent variable *Task (global task vs. Brief questions).* Mean and standard deviations for each of the think-aloud dependent variables can be observed in *Table 1.8*.

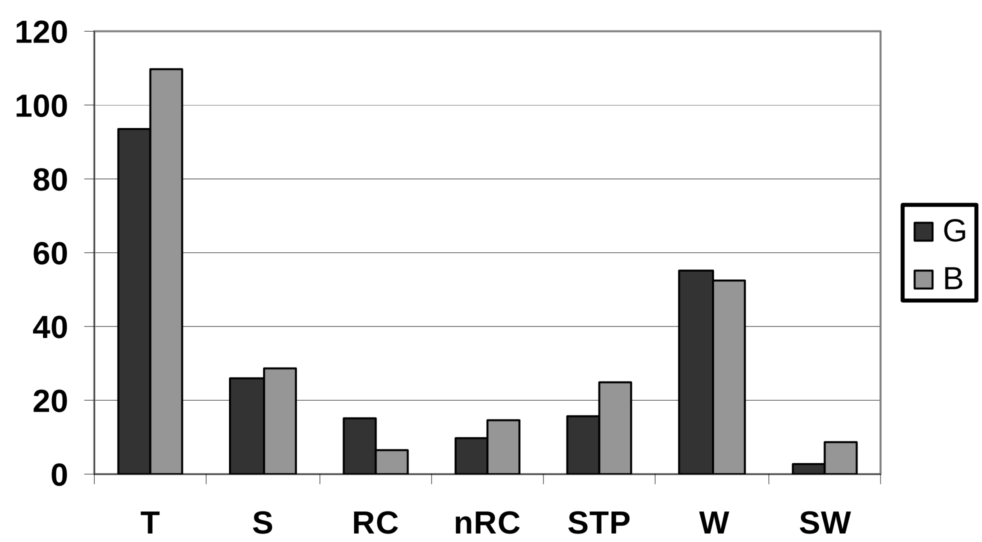
TABLE 1.8 VERBAL PROTOCOL ANALYSIS	T		S		RC		nRC		STP		W		SW	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
G	93.36	37.77	25.81	29.12	15.18	12.57	9.72	10.55	15.72	6.00	55.18	25.63	2.72	2.96
B	109.91	33.45	28.41	13.02	6.25	6.67	14.50	26.35	24.91	37.21	52.41	23.93	8.75	6.21

Table 1.8. Verbal Protocol analysis

Only two main results were significant. First, students performing a global task (*M* = 15.18, *SD*= 12.57) significantly verbalized more on understanding relevant information than those who performed brief questions (*M* = 6.25, *SD*= 6.67), F (1, 21) = 4.64, $p < 0.05$. Secondly, students answering the global task produced fewer superficial writing comments (*M* = 2.72, *SD*= 2.96) than those students answering brief questions (*M* = 8.75, *SD*= 6.21), F (1, 21) = 8.53, $p < 0.05$. These results, though few, are very meaningful for our purposes. Indeed, they follow our hypothesized trend and provide convergent evidence for our claim that global tasks promote a deeper processing of the multiple sources, in comparison to brief questions. Thus, the global task students, driven by the demands of their task, verbalized more to understand relevant information. We had argued before that these verbalizations would be aimed at understanding the main relevant points raised in the documents at a deep level. Hence, this result was in agreement with our expectations for the global task. In fact, students verbalized how they were processing the textual information during the course of the experiment. And this verbalization was focused on deeply understanding the main units of information from the documents. Contrarily, brief question participants did not verbalized so much at this deep level to understand relevant information. As we argued earlier, brief questions promoted a more superficial processing of the sources, as evidenced by learning and time-based masures. Now, we have convergent and further evidence for this claim.

Complementary, the global task also reduced the number of superficial writing verbalizations in the question-answering process, in contrast to brief questions. This is also evidence for the beneficial effects of global tasks not only in reading, but also in the process of performing the task. Superficial verbalizations when writing included the concentration on minimalist dimensions of the question-answering process, such as spelling, but also the wrong selection of the answer or the emission of irrelevant associations. The fact that brief questions participants scored higher in this category might have been due to the nature of the question-answering process in brief questions. Whereas performing a global task only required the process of answering a single but very broad answer, answering brief questions involved a more sophisticated question-answering process to answer four questions. As we earlier argued, answering several questions requires for the reader the engagement in a more complex iterative process (Cerdán et al., submitted) which could have increased the probability of encountering wrong answers and try to make them fit to any of the questions. Therefore, this may be a possible explanation for brief question participants verbalizing on superficial information in a greater extent when answering. However, we strongly believe that part of this effect is also due to the superficial processes brief questions seem to induce when working with multiple sources. In contrast, the global task seems to make students engage in a deeper processing of sources. The rest of the verbal protocol measures, though non-significant, provided further evidence in this sense (*see Graphic 1.12*).

Graphic 1.12
Protocol analysis



*

Note to Graphic 1.12: T (Task verbalizations), S (Search), RC (Relevant Comprehension), nRC (non-Relevant Comprehension), STP (Superficial Text Processing), W (Writing) and SW (Superficial Writing).

Indeed, results for the rest of the verbal protocol categories followed our expectations. Briefly, *task verbalizations* were slightly higher in brief questions participants. We had assumed that verbalizing on the task and on the use of the software would indicate a more superficial processing in the course of the experiment and we had expected to be more present in brief questions. The trend we found somewhat confirmed our expectations. Similarly, we had hypothesized that both the global task and brief questions students would equally verbalize on the *search processes*, as these were inherent to the experimental design in both conditions. We had also hypothesized that it might be that brief question students verbalized more in terms of searching, because of the requirements of brief questions for finding isolated answers. However, results confirmed our first hypothesis that the need for searching was equally present in both conditions.

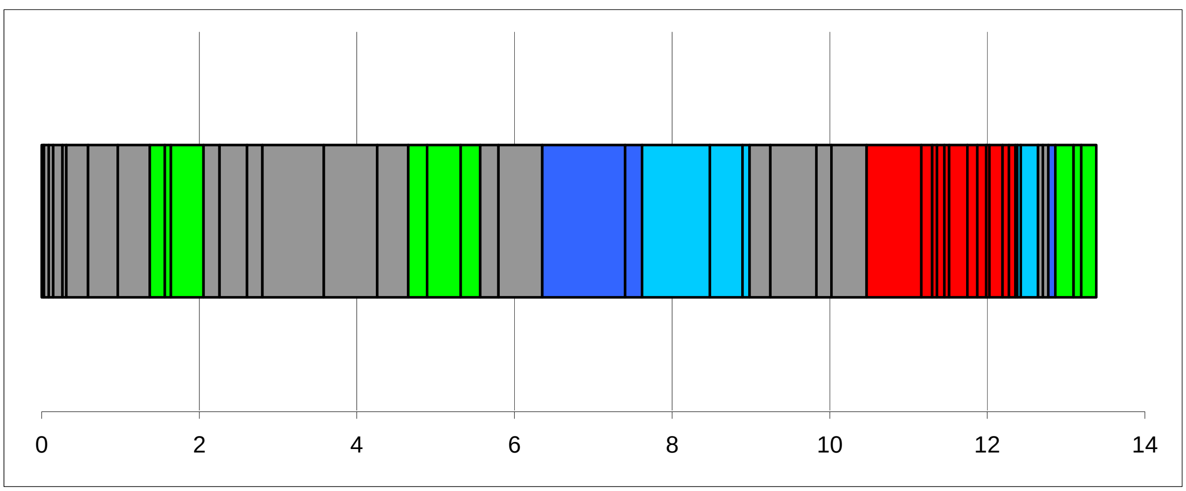
Relevant comprehension was significant, verbalizing more to understand those performing a global task. Complementary, we found the opposite trend in *non-Relevant comprehension*. In this case, it was students answering brief questions who verbalized more to understand irrelevant information. It might have ocurred this way due to the more sophisticated question-answering process in brief questions. Given that brief questions participants needed to complete four questions and find their corresponding answers in different locations of the sources, the probability of focusing on non-relevant information in the inspection process was higher than those with the global task. Regarding *Superficial Text Processing*, the trend we found also confirmed our expectations for brief questions inducing a more superficial processing of texts than the global task. Finally, in the writing process, both groups verbalized almost the same on the *regulation of the question-answering process*, as expected. However, brief question participants significantly produced more *superficial comments* than the global task participants. Hence, providing further evidence to our expectation that brief questions would induce a more superficial process than the global task.

(f) *Visual analysis of the combination of information process*

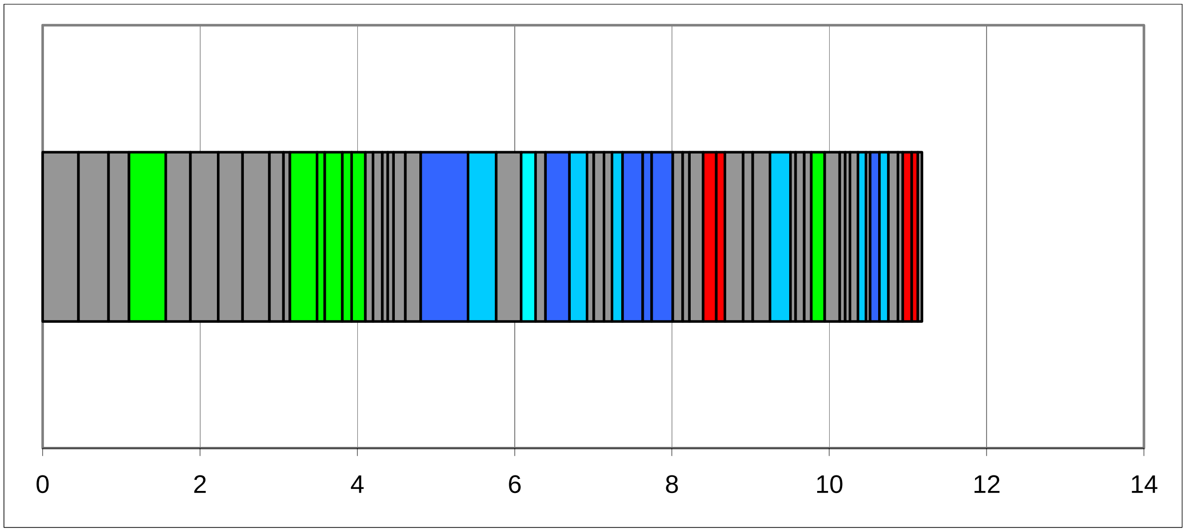
As we explained earlier, we would conclude our set of analysis by conducting and presenting a specific case analysis of how students actually combined information from the multiple sources in their reading sequences across time. We included this case analysis to overcome the limitations of the previous measures, in that they did not accurately indicate if students were really combining the four relevant units of information while reading to perform their task, or, contrarily, they produced a linear reading process which would not include re-readings of relevant segments and connections between different paragraphs. We had hypothesized that combination of units of information across the reading sequence would be more apparent in the global task selected cases, and the linear reading process would be more present in the brief questions participants. Additionally, differences should be found between thinking-aloud or not.

Let us remeber that we had selected four *high-learning cases*, from each of the conditions in Experiment 1 (i.e, Global-think-aloud; Global-no-think-aloud; Brief-think-aloud;Brief-no-think-aloud) and used their reading sequence represented in a *linear one-bar graphic.* Each of the divisions of this *linear one-bar graphic* represents a visit to a textual segment and its width, the reading-rate per word at which this segment was read. Thus, the wider the segment, the more slowly this segment would have been read. Irrelevant information was colored in grey, whereas the four relevant segments were colored as follows: paragraph *P 1-3*, in red*;* Paragraph *G 1-3*, in dark blue and *G 2-1*, in light blue; finally, paragraph *R 2-1* was colored in green. It should be noted that the order at which the first relevant paragraphs appear depend on the texts-presentation-order each student belonged to, as the order of texts was randomized in the experiment.

Global task think-aloud and no-think aloud selected cases. Reading sequence 1.1 represents the reading pattern across time of one participant performing the global task and simultaneously thinking-aloud. Complementary, *Reading sequence 1.2* represents another participant who also performed the global task but did not think-aloud.



Reading sequence 1.1: Global Task&Think-aloud



Reading sequence 1.2: Global Task&no-Think-aloud

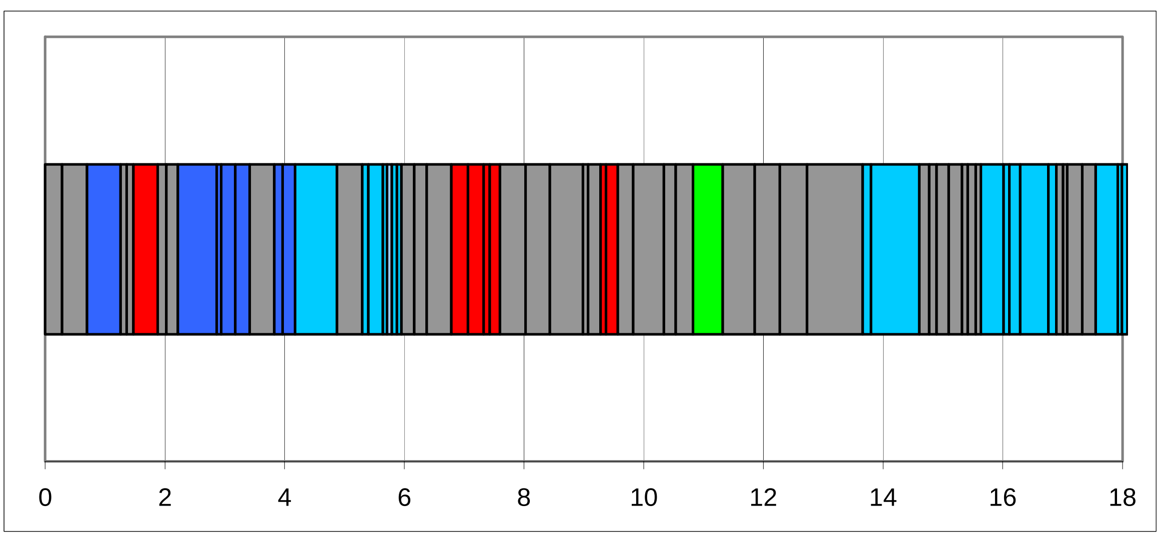
As apparent in *reading sequence 1.1*, the student read the texts and both relevant and irrelevant information linearly during the first and the second third of the sequence. As the student read the texts, he or she would encounter the relevant units of information making up the mental model of how bacteria become resistant to antibiotics. When reading the relevant units, two main aspects are remarkable: first, the width of some relevant segments, which indicates that they were read more slowly and, second, the reading repetitions of the same segment once encountered, indicating that he or she would have noticed that it was key information to process and deserved a special attention. Interestingly, in the last third of the reading sequence we find two main actions performed by the student: a) there was a noticeable adjustment of the reading speed rate, becoming the readings faster, as apparent in the width of the segments;

b) there were re-readings of the relevant segments which were linearly read at early stages of the reading-sequence and in the last third of the reading sequence were further re-read at a quicker speed. Moreover, the student re-read all four relevant segments making up the mental model of how bacteria become resistant to antibiotics, as if trying to consolidate the connections of already processed relevant information.

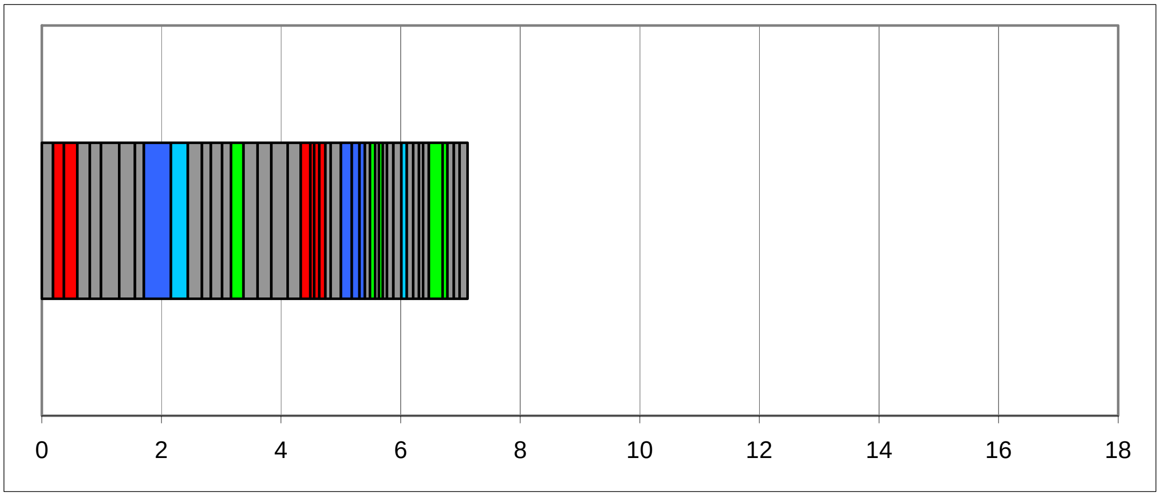
Complementary, the student performing *reading sequence 1.2* also read the textual information linearly at early stages of the reading sequence. He or she also re-read relevant segments once encountered for the first time, probably in an effort to process and understand these relevant units of information. However, differently to *reading sequence 1.1*, we already found combination of information from the second third of the sequence. Thus, segment G 2-1, painted in light blue was re-read several times in combination to segment G1-3, painted in dark blue. This would indicate a first effort to combine and connect information. The last third of the reading sequence also experienced a reduction in reading rate, thus becoming readings faster, and a quicker revision of all four relevant segments as if to consolidate the mental model. Differently to reading sequence 1.1, there were more re-readings of non-relevant segments in the last third, though we think they were casual and due to the efforts to find the four relevant segments to combine in the last third of the reading sequence. It should be noted that reading sequence 1.2 belonged to one student who did not think-aloud. Differences between the reading sequence thinking-aloud and not thinking-aloud are slightly apparent in the width of the segments. Hence, these are generally wider in thinking-aloud, thus indicating that reading was more slowly, a result which is in agreement with previous on-line evidence.

In sum, what we want to remark in this two reading sequences belonging to global task participants is that, whether from the second third or in the last third, the students made an effort to combine all relevant segments for the task and for the integrated mental model of how bacteria become resistant to antibiotics. These had been linear and slowly read at early stages of the experiment. Students had noticed its relevance, devoted more resources to understand and connect them at a deep level (i.e., reading them more slowly, repeating its readings) and they had quicked reviewed them all to consolidate the mental model. We assume global task students were integrating information this way.

Brief questions think-aloud and no-think aloud selected cases.



Reading sequence 1.3: Brief questions&Think-aloud



Reading sequence 1.4: Brief questions&no-Think-aloud

As it is apparent in *reading sequence 1.3*, the reading pattern in this selected brief questions case followed our expectations. Thus, the student apparently read at a quicker speed, as evidenced in the segments' width, from the very beginning of the experiment. He or she would linearly encounter the relevant segments and would sometimes repeat readings, thus showing to have noticed the relevant units of information for the questions. However, readings of relevant segments were generally faster, as if the processing of relevant information were more superficial. Although there was a re-reading of *P 1-3* in the second third of the sequence, we did not find in general a combination of information pattern. In fact, in the second and especially the third third of the reading sequence there was not a quick revision of all relevant segments but a fixated concentration on only one of them. Moreover, reading speed did not decrease in the last third as much as in the global question sequence.

Finally, the amount of readings of non-relevant information was quite high from the second third of the reading sequence, a moment in which a reader who had recognised the relevant units of information should have concentrated mainly on those units, as we found in the global task students.

When we observe *reading sequence 1.4* there is something highly remarkable. There is a drastic reduction in the length of the reading sequence, becoming as shorter as the global task students or even more. What differenciated the above reading sequence to this one was thinking-aloud, which was non-existant in the present reading pattern. Therefore, when thinking aloud and answering brief questions students seemed to increase their reading pattern in length, and as soon as thinking-aloud disappeared, this reading pattern significantly decreased. Again, there seems to be the similar kind of *multiplicative effect* we found in on-line time-based measures which increased the time needed to perform the task in brief question students who simultaneoulsy thought aloud. On the other hand, a combination of information pattern was absent too. Indeed, this student seemed to perform his or her question-answering task in two cycles. The first of them present in the first half of the sequence, in which segments were linearly read and the four relevant segments encountered, there being very few repetitions of relevant segments. The second cycle seemed to occur from the second half of the sequence to the end. In this second cycle, the student again read the relevant segments in the same order as in the first cycle, though only altering the order for one of the segments. Additionally, a considerable amount of non-relevant information was read in this second cycle. In any case, there was never a similar pattern as that found for global task students and combination of information seemed to be quite absent.

In conclusion, combination and apparent efforts to integrate information were only observed in the global task reading sequences, whereas the brief question sequences were mainly linear, with no big decreases in reading rate at later stages of the sequence and no complete revisions of all the relevant segments as if to consolidate a mental model. Moreover, whereas the global task seemed to induce a more detailed processing of relevant units of information, brief questions promoted a faster reading pattern, which would probably induce superficial processing. This superficial processing even focused students' attention on non-relevant segments in a greater degree than in global task students.

Finally, we again found think-aloud interacting with brief questions and, generally, slowing down quite remarkably the process to perform the task in brief questions. All this additional evidence completes a set of measures and analysis, from off-line learning measures to on-line time-based and think-aloud-based, which helps us in reaching one main conclusion: the global task seems to induce combination and integration processes in a greater degree than brief questions. Therefore, it would be the most beneficial task to learn from multiple expository texts.

2.3 Experiment 1: Discussion

We began this dissertation by emphasizing the need to find new methods and tasks by which students would be promted to integrate information from multiple expository texts, similarly as it occurred with specific kind of tasks (i.e., high-level questions) which had been proven especially beneficial to learn from single texts. If in the case of single-text learning the most beneficial tasks were those which made students actively integrate information from several distant segments in the text (Vidal-Abarca, Mengual, Sanjose & Rouet 1996; Vidal-Abarca, Gilabert & Rouet, 1998; Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil) we expected that the most beneficial task in the multiple-texts learning situation should also promote the connections among different units of information.

We understood the expression *tasks beneficial for learning* in the sense that they should be capable of promoting deep comprehension, or the construction of a coherent and integrated mental model from the text/texts (Kintsch, 1998; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). In the single-text situation high-level questions were those tasks found to be the most beneficial for learning, making students actively process and connect distant and different units of information via inference making. This inference making process induced by high-level questions resulted in deep comprehension, as evidenced by final comprehension mesures. In the multiple-texts-situation, and primarily in multiple expository text reading, we hypothesized that the most beneficial task for promoting deep learning would be such task that would also engage the student in connections of information via integration processes.

We considered integration to be a fuzzy but nuclear concept in learning from multiple expository texts. Fuzzy because it had been widely used in research with multiple sources so far, however seldom clarified what it consisted of in terms of mental processes. Nuclear, on the other hand, as it should be the main responsible process in the construction of an integrated mental model from several sources and different units of information on the same topic. Indeed, if the product of reading and learning from multiple sources is the construction of a mental model from the multiple sources and contents (Perfetti, Rouet & Britt, 1999), integration or the connection of different units of information should be the main mental process responible for this construction.

Therefore, the task which would result most beneficial for the construction of a deep and integrated mental model from multiple sources should, on the one hand, promote the integration and combination of multiple sources and units of information and, on the other hand, yield the best scores in deep comprehension measures. This way, it would demonstrate to be the most beneficial task because of inducing the main relevant processes involved in multiple texts comprehension situations, that is, integration of different units of information into a higher mental model from the multiple sources.

To find and test this task which would promote integration of information and the construction of a coherent representation from multiple sources in a greater extent than other tasks, we designed a multiple expository texts learning situation and two different kinds of tasks to be performed from the multiple sources. On the one, a global task which would make students extract, connect and combine the different relevant units of information from each of the documents and integrate them into a higher-order mental representation. On the other hand, four brief questions focused on isolated units of information in each of the sources.

We hypothesized that the task that would induce integration processes and deep learning in a greater degree would be the global task, as it would make the students actively search for and connect the relevant units of information. Contrarily, brief questions would promote an isolated extraction of the main contents in each of the sources and no or very few connections among them, thus limiting the possibilities to build an integrated higher-order mental representation of the multiple sources. Complementary, we used the think-aloud method in half of the sample to obtain convergent evidence for our claim that the global task promotes integration of information**.** Given that the use of this methodology may interfere in the learning process from multiple sources, we decided to test if thinking-aloud interfered or not with performance and learning from multiple sources, and in which direction.

Briefly, our hypothesis for the present dissertation so far could be summarized as follows. First, we wondered if the global task actually promoted the construction of a higher-order integrated mental representation from the multiple sources in comparison to brief questions. If this was the case, we also wondered if deep learning was promoted because of the activation of integration processes. Complementary, we analysed the extent to which thinking-aloud interfered with performance and learning in the multiple sources or not.

Results obtained in Experiment 1 yield very interesting and clarifying results. According to our main hypothesis, the global task actually promoted deep learning and the construction of a coherent and integrated mental representation from the multiple sources. This was apparent in the final learning measures we obtained, especially in the *application&transfer* practical case. We had also hypothesized that the global task, given that it would promote the location, extraction and connection of the main units of information from the documents, would promote a *versatile final learning*. Versatile as it would allow students successfully perform any kind of final learning task, whether more superficial or located at a more text-base level or located at a deep level.

Hence, when measuring the individual identification and understanding of the main units of information making up the mental model of how bacteria become resistant to antibiotics, the global task yield similar results or even higher (i.e., in the no-think-aloud condition) than brief questions. Additionally, we considered a final level to compare between tasks: performance on the task or a measurement of how well students had succedeed in the question-answering process. We had remarked that it should not be considered a learning level in itself, but a transitory effect due to the effects of training during the experimental phase (Schmidt & Bjork, 1992). No global differences were found between tasks, a result which is in agreement with previous results in single-text learning situations (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted).

In conclusion, the global task proved indeed more effective in terms of deep learning and in the construction of an integrated mental model from the multiple sources, especially when considering final learning measurements which directly assess the degree to which students had integrated information and acquired a deep understanding of the main points raised in the documents. Contrarily, brief questions, as hypothesized, yield lower scores than the global task when considering deep learning from the multiple sources. However, in lower learning levels, such as surface construction of the main ideas present in de documents (i.e, true/false questionnaire) or performance on the task an interesting pattern of results appeared which involved the presence of thinking-aloud.

Indeed, the higher scores or similar to the global task scores brief questions students obtained both in performance and surface learning were only apparent when thinking-aloud was present. As soon as thinking-aloud disappeared, brief question students obtained similar scores to the global task or even lower (i.e, true/false questionnaire scores). Interestingly, this interaction effect of thinking-aloud did not appear in deep learning measurements or for the global task. From this evidence we concluded that thinking-aloud and answering brief questions seem to produce a kind of *multiplicative effect* which fosters a surface processing of texts.

On the one hand, thinking-aloud globally appeared not to affect so much surface or performance learning, but clearly hinder high-level learning, as evidenced in the *application&transfer* practical case. On the other hand, brief question answering also limited the possibilities to learn from the multiple sources at a deep level but seemed to be more beneficial for learning at a more superficial level (i.e., location and understanding of main points in documents). Therefore, both thinking-aloud and brief questions answering seemed to focus students' attention on a superficial level of processing from the multiple documents. When presented simultaneoulsy, this superficial processing effect was even more marked.

The explanation of the limiting effects of thinking-aloud to learn from texts at a deep level and its increasing effect when presented with brief question answering was already presented at the beginning of the present dissertation. According to the theoretical evidence existing so far on the effects of thinking-aloud on performance and learning, we could expect both favoring and limiting effects of thinking-aloud. What we have finally found is in agreement with some empirical findings (Wade & Trathen, 1989; Magliano, Trabasso & Graesser, 1999) in that thinking-aloud seemed to clearly affect the product of learning from multiple texts and focus students' attention on a superficial processing of texts.

Thus, it seems that thinking aloud limited the possibility to engage in deep learning processing from the multiple sources, especially in interaction with brief questions. The explanation we raise for this effect is that thinking-aloud may have globally overloaded working-memory because of the *extraneous demands* of performing the tasks from multiple sources and thinking-aloud. Indeed, working memory resources are limited (e.g., Miller, 1956) and if we overload one of its components (Sweller, 1994), that is, the *extraneous load* due to thinking-aloud, fewer cognitive resources would be available for higher-level processes, which in terms of the Cognitive Load Theory distinction is called *germaine load* (Sweller, 1994). In fact, the *Cognitive Load Theory* proposal (i.e, Sweller, 1994; Sweller et al, 1998) argues that working memory resources are mainly divided into three main componentes: a) *extraneous load* or the load placed on working memory due to the demands of the instructional design; b) *intrinsic load*, or the load related to the intrinsic demands of the task (i.e., difficulty), c) *germaine load*, or the load placed on working memory by the instructional design in assisting schema formation.

Hence, if we overload one of its components, such as extraneous load due to the effects of thinking-aloud, fewer cognitive resources would be available for higher-level processes (i.e, germaine load). Moreover, if the intrinsic load increases, because of the demands of learning from multiple sources and answering a more sophisticated question-answering process of four brief questions, it may be that cognitive resources for high-level processes become even more limited. Consequently, there might be a greater focusing on superficial processes, which requires fewer cognitive resources, exactly as it happens with brief-question-answering and thinking-aloud.

Once tested our learning hypothesis both for the task and the effects of thinking-aloud, we wanted to obtain on-line evidence for these effects. For this purpose, we had three sources of on-line information: a) global time and reading-rate time-based measures obtained during performance of the task, b) think-aloud protocol analysis, c) a visual case analysis of how students combined information. We were aimed at confirming that the global task yielded the best learning effects due to integration processes being activated in a greater extent than with the brief questions. Additionally, we wanted to clarify what integration of information from multiple sources consisted of in terms of cognitive processes. Finally, we wanted to find on-line evidence that thinking-aloud affects the process of learning from multiple sources.

Performing a global task created overall no significant differences to performing brief questions when considering global times in the experiment and its divisible components. It slightly implied fewer time reading the question in contrast to brief questions needing longer. However, this was not enough evidence for our claim that global task induced integration processes. When considering reading-rates-based measures clarifying differences appeared. First, the global task made students read all textual information more slowly, in contrast to brief questions reading faster. This first evidence is in agreement with Mannes and Hoyes findings (1996) that when reading to integrate there is an increase in reading slowness. This more detailed processing can be explained in terms of integration processes to be taking place.

Students devote more resources, and hence a more slowly reading, because complementary evidence has to be integrated in short-term memory through *reinstatement&integration* (Mannes, 1994; Mannes & Hoyes, 1996).

A second evidence for integration taking place in a greater extent in the global task was that students performing the global task read the relevant information more slowly than the non-relevant units of information, an effect which did not appeared in brief questions. This result is also in agreement with the second of the main evidence Mannes and Hoyes (1996) provide for integration taking place when reading texts with multiple perspectives. Students read even more slowly when they read new information that they have to connect with old information in a multiple text reading situation. Similarly, global task students, when they encountered the main units of information making up a mental model of how bacteria become resistant to antibiotics, reduced their reading speed in order to process and integrate this information in greater detail.

We also hypothesized that integration processes should also be apparent in how students read the textual information across time. This time, our hypothesis that there should be a more slowly reading at earlier stages of the reading sequence and becoming this reading faster at later stages was valid both for the global task and brief questions. We had hypothesized that the more slowly reading at early stages would indicate efforts to process and integrate information when reading for the first time and the increase in speed at later stages would indicate a revision to consolidate the mental model.

Thus, it seems that integration or a detailed processing of texts also occurs in brief questions to some extent, though not strong enough to create deep learning differences. In any case, the reading speed decrease was linear and progressive for global task students, whereas not so regular for brief questions students. Indeed, brief question students only started to reduce their reading rate from the second half of the reading sequence. This result might indicate not so much integration taking place, but more difficulties in processing the information or an excessive concentration on the textual and isolated ideas in the texts, as evidence by performance and learning measures.

In conclusion, we found that the global task not only created deep learning differences, but also seemed to promote a more detailed processing and integration of the main units of information, in contrast to brief questions. Convergent evidence for this result was also found when analyzing the verbal protocols. Indeed, whereas the global task promoted the production of more verbalizations to understand the relevant units of information, brief questions fostered a more superficial behavior in processing the texts and in the question-answering process, as evidenced by a higher number of superficial comments when writing the answers. This result is parallel to previous think-aloud evidence that students integrating information produce more elaborations to understand information (Mannes and Hoyes, 1996; Strømso, Braten & Samuelstuen, 2003).

Finally, the last of the integration evidence taking place in global tasks in a greater extent came from an informal visual analysis of the reading sequence.

We considered that integration should imply the combination of all relevant units of information making up the mental model presented in the texts. After visually analyzing some illustrative cases for each of the experimental conditions, we concluded that the global task actually promoted the combination of different units of information in contrast to a more linear and isolated reading performance for students answering brief questions. Briefly, whereas global task students slowly read the relevant units of information at earlier stages of the reading sequence and quickly revised them all at the end of the reading sequence, brief question students read textual segments more linearly, without signs of combination, and globally at a faster speed. All of this was final evidende for the global task promoting integration of units of information in a greater extent than brief questions and some light to understand what integration of information consists of in terms of mental processes.

Finally, regarding how thinking-aloud affected the on-line performance of the tasks, thinking-aloud globally increased the time needed to perform the experiment and their divisible components. Moreover, this occurred in a greater extent in interaction with brief questions. Hence, we found on-line evidence that thinking-aloud and brief questions affected the on-line process by requiring more time to perform all tasks. However, this increased time was not devoted to higher-level processes, but to focusing on a superficial text-base level, as evidenced by performance and learning measures.

Again, it seems thinking-aloud and brief-question answering produced an overload in working-memory resources, which was now apparent in increased times for performance.

Thinking-aloud also made overall reading more slowly, as apparent in reading-rate measures. A result which is in agreement with general predictions for thinking-aloud (Ericsson & Simon, 1980; 1993). Moreover, thinking-aloud increased the relevant reading rate for brief questions, making it similar to the global task relevant reading rate. However, this effect disappeared as soon as there was not thinking-aloud. Therefore, thinking-aloud and brief question answering again produced the *multiplicative effect* in reading rate measures. This time, students answering brief questions while thinking aloud were not especially concentrated on processing the relevant units of information at a deep level, buth were sticked to a more slowly reading based on superficial processing only because of thinking-aloud. To conclude, when considering the reading speed adjustment across time we found that students thinking-aloud seemed to need more time to process information across the experiment, as evidenced by a more homogeneous reading during the first half of the experiment and a bigger decrease in reading rate from the second half. Contrarily, students who did not think-aloud seemed to be able to process information more easlily from the very beginning of the experiment.

Consequently, we have also found on-line evidence to be able to state that thinking-aloud seems to clearly affect the on-line processing of texts and performance of the tasks. And, especially, performance of brief questions, exponentially increasing think-aloud effects and thus showing a very clear pattern which completely matches learning evidence. The visual analysis of cases also confirmed our hypothesized *multiplicative effect* in thinking-aloud and answering brief questions.

When we fulfilled our expectations for experiment 1 and the main hypothesis for this dissertation, new research challenges came to our mind. The first of them, how would our designed tasks (i.e, global task and brief questions) behave in a similar experimental design in which there was not the inclusion of thinking-aloud as one of the experimental variables?, would learning and process results for each kind of task replicate?; secondly, what if we included a more extreme kind of task to work with multiple sources which by no means could ever promote integration from multiple documents?; third, if we included this third kind of task and it ended up being the less beneficial task to work with multiple sources, what role would brief questions finally have for learning from multiple sources?.

To answer all these questions we designed a second experiment, exactly similar to Experiment 1, however changing in two aspects: a) there would be no thinking-aloud participants, given that it seems to limit the possibilities to learn from multiple sources at a deep level; b) there would be an additional task to be performed from the multiple sources, which would foster students' attention on very concrete isolated ideas in the experiment, thus eliminating any possibility to integrate information from the documents.

For this purpose, we designed *very specific questions* whose answer was only located in one sentence or two belonging only to one document. They were the most literal questions we could have ever created, parallel to the low-level questions we used in past research (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil). Thus, similarly to brief questions, they would focus students' attention only on one of the documents. But specific questions were more concrete in terms of the location of the answer, which only required finding it in one of the sentences of a paragraph, and in terms of the mental processes they induce, because they would only promote search and location processes. Indeed, whereas brief questions sometimes needed the production of inferences to connect information only in one text, this was never necessary in the case of specific questions. In any case, only *copy&paste* processes were required to give a correct answer. For all these reasons, neither integration of information nor the understanding of the main points raised in the documents would be possible.

3. EXPERIMENT 2: PERFORMING A GLOBAL TASK vs. ANSWERING BRIEF or SPECIFIC QUESTIONS WITH MULTIPLE EXPOSITORY TEXTS: THE ROLE OF TASK ON INTEGRATION PROCESSES

Experiment 2 has thus two main objectives. The first of them, try to replicate results obtained in Experiment 1 and consolidate our main conclusions raised in Experiment 1 for types of tasks promoting more integration or not in working with multiple sources, in the absence of students thinking-aloud. Second, include a more extreme task which should not allow students learn and integrate information from multiple sources at all and, this way, clarify the role of brief questions in multiple expository texts learning situations. For this purpose, students read the same three texts and performed three different kind of tasks (i.e., Global task vs. Brief or Specific questions), again by means of the software Read&Answer (Martínez, 2003).

3.1 Method

Participants

Twenty-three students participated in Experiment 2. Similarly to Experiment 1, they were all enrolled in a Psychology undergraduate programme at the University of Valencia and participated for course credit. They had also a mean age of 20 years. The reasons for selecting this kind of sample were the same as in Experiment 1: to have university students who were familiarized with learning from expository texts and thus would have few problems in being confronted with a multiple expository text learning situation.

They were randomly assigned to each of the experimental conditions in Experiment 2. This way, 8 students were assigned to the *Global task* condition; another 8 students were included in the *Brief questions* condition; and, finally, the rest 7 students were assigned to the new *Specific questions* condition.

Materials

Materials were exactly the same as in Experiment 1. However, there was an slight change in the tasks matching the main contents to be integrated which deserves a detailed explanation. In fact, as we have presented earlier, we created a new kind of task which would be much more specific as the brief question task. Nevertheless, the contents which this new task would have associated would be exactly the same as the global task and brief questions.

If a detailed table of the main contents to be integrated in the experiment was the basis for the design of the global task and the corresponding four brief questions, it was also the starting point to elaborate 10 specific questions which would also match the 22 main ideas in the table of contents. Therefore, additionally to the global task and the four brief questions, we also proposed *10 specific questions* which would cover the isolated units of information, each question covering no more than two ideas, and would be contrasted now in experiment 2. A detailed description of all the kind of questions and their corresponding contents is presented below in the same *hierarchichal table of questions across contents* as that presented in

Experiment 1, but now including a more elementary level with 10 specific questions.

HIERARCHICAL TABLE OF QUESTIONS ACROSS CONTENTS IN EXPERIMENT 2			
QUESTIONS: CORRESPONDING IDEAS AND TEXT SEGMENTS		**IDEA**	**TEXT**
GLOBAL	***Explain how bacteria resist the effects of antibiotics and which biological mechanisms explain this phenomenon and its transmission to other bacteria.***	**22 IDEAS**	**P,G,R**
BRIEF 1	*Which characteristics of bacteria have an influence on the development of bacteria resistance to antibiotics?*	1,2,3,4	R 2-1
SPECIFIC 1	Which are the main characteristics of bacteria?	1,2	R 2-1
SPECIFIC 2	Which is the main factor that facilitates the development of resistance?	3,4	R 2-1
BRIEF 2	*Which biological mechanisms permit bacteria become resistant to antibiotics?*	5-13	P 1-3 **OR** G 2-1
SPECIFIC 3	When did bacteria begin to develop resistance to antibiotics?	5	P 1-3
SPECIFIC 4	How many resistance mechanisms can we find?	6,7,8	P 1-3 **OR** G 2-1
SPECIFIC 5	What is genetics mutation?	9,10	P 1-3 **OR** G 2-1
SPECIFIC 6	What are Plasmids and what do they do to produce resistance to antibiotics?	11,12	P 1-3
SPECIFIC 7	In which case is multirresistance developed?	13	G 2-1
BRIEF 3	*Can resistance be transmitted to other bacteria?, under which circumstances?*	14-16	P 1-3 **OR** R 2-1
SPECIFIC 8	Why only one random mutation of a gene can have a big impact?	14	R 2-1
SPECIFIC 9	To which groups of bacteria is resitance transmited?	15,16	P 1-3 **OR** R 2-1
BRIEF 4	*How can bacteria resist to antibiotics?*	17-22	G 1-3
SPECIFIC 10	By which ways respond bacteria to the letal effects of antibiotics?	17-22	G 1-3

Briefly, as it can be observed in the *hierarchical table of questions across contents* in Experiment 2, the *10 specific questions* directly make students concentrate on the individual idea units, including each of the questions no more than 2 idea units. Brief questions associated to the 4 main areas split into several specific questions. Only in the case of brief question 4 we created no more than one corresponding specific question, specific question 10. The reason for this was that the contents for this area were specific enough to be splitted into smaller questions. To make brief question 4 and specific question 10 vary according to our hypothesis, we tried that the wording of brief question 4 made the students process more deeply what he or she was being asked for, whereas the wording for specific question 10 directly overlapped the wording of the paragraph were the answer was located, so that the student could only copy the information with no need to process and understand.

The elaboration of these *10 specific questions* also implied some learning effects hypothesis. Hence, because specific questions would focus students' attention only on isolated ideas present in single documents and only requiring copy&paste processes to extract them, integration and deep learning from multiple sources would by no means be possible. On the other hand, specific questions would also not allow surface learning of the main contents present in the documents, contrarily to brief questions. Because specific questions would only promote concentration on individual sentences from each document and the display of *copy&paste* processes to answer correctly, they would not even help students in understanding the individual relevant points raised in each of the documents, which is precisely what brief questions were supposed to be helpful at and what was measured with the final recognition and understanding of individual items test. Let us remember that both brief and specific questions required students to focus on single documents. However, whereas brief questions promoted some inference making at least among a group of ideas present in a single text, specific questions only required the location of one or two specific ideas, without never connecting them.

Finally, we hypothesize that the best performance effects would be present in *specific questions*. Given that specific questions focus students's attention not on integrating units of information, but on considering them individually, even at a more atomic level than brief questions, they would be especially helpful to extract these units of information for the task, thus increasing performance. Complementary, on-line time-based evidence should also match this learning hypothesis for specific questions, there not being signs of integration of information and deep processing at all.

Procedure

Similarly to Experiment 1, participants in Experiment 2 were also tested over two sessions. Again, session 1 was used to obtain the control measures scores and to train students to use the software *Read&Answer* to fully understand the experimental requirements. Session 2, on the other hand, was used to perform the experiment and obtain the final learning measures. Similarly to the case of participants in experiment 1 not thinking-aloud, students performed their task in groups of three or four students, each of them sitting in individual computers and having no limits of time to complete their task. Overall, participants required almost two hours to complete the experiment, though remarkable differences appeared depending on the kind of task they had been assigned to.

Dependent measures

We used the same dependent measures as in Experiment 1. In short, *control measures* to obtain evidence that groups did not significantly differ in a set of measures that could contaminate subsequent analysis; an *assessment of type of task effects on performance and learning from multiple expository texts*, by using performance and learning measures; measures to analyse *how students globally distribute their time in the experimental session*; measures to test *how students read and integrate information from the multiple sources*; and, finally, an informal visual case analysis of the *combination of information process*. All this measures and their corresponding analysis should help us in determining the role of brief and specific questions in learning from multiple expository texts and, on the other hand, try to confirm the beneficial effects obtained in Experiment 1 for the global task.

3.2 Experiment 2: Results

(a) Control measures

Similarly to experiment 1, we analyse the set of measures that could contaminate subsequent results by conducting one-way Anovas, with independent variable *Task* (*Global task vs. Brief questions vs. Specific questions*). Dependent measures were: *Background knowledge, Time reading words, Time reading Pseudo-words, mistakes in words, mistakes in Pseudo-words and, finally, writing speed on the keyboard*.

As we expected and in agreement with experiment 1 results, no significant differences were found among the experimental conditions (*see Table 2.1*).

TABLE 2.1 CONTROL MEASURES IN EXPERIMENT 2												
	B. Knowledge		Time Words		Time Pseudoword		Mistakes Words		Mistakes Pseudoword		Writing Speed	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
G	9.25	2.82	24.87	4.97	38.62	7.31	0.13	0.35	1.13	1.55	57.75	13.59
B	10.57	2.23	25.86	3.18	41.86	5.43	0.14	0.38	0.43	0.53	73.71	20.93
s	9.63	3.70	26.50	7.15	41.50	9.99	0.00	0.00	1.63	0.92	66.50	26.32

Table 2.1 Control measures in Experiment 2.

Because we wanted to contrast results obtained in Experiment 1 and 2, we should guarantee that both samples did not significantly differ in the present control measures. For this purpose, we conducted one-way Anovas with independent variable *Experiment* (Experiment 1 vs. Experiment 2) and dependent variables, the above control measures. As we hoped to find, no significant differences were found for any of the control measures. Hence, we had a guarantee that both the sample in Experiment 1 and the sample in Experiment 2 were similar and, thus, comparisons could be made between results in both experiments.

(b) Assessment of type of task effects on performance and learning from multiple expository texts.

To test the learning effectiveness of each kind of task in this Experiment 2, we again considered the same three learning levels as in Experiment 1.

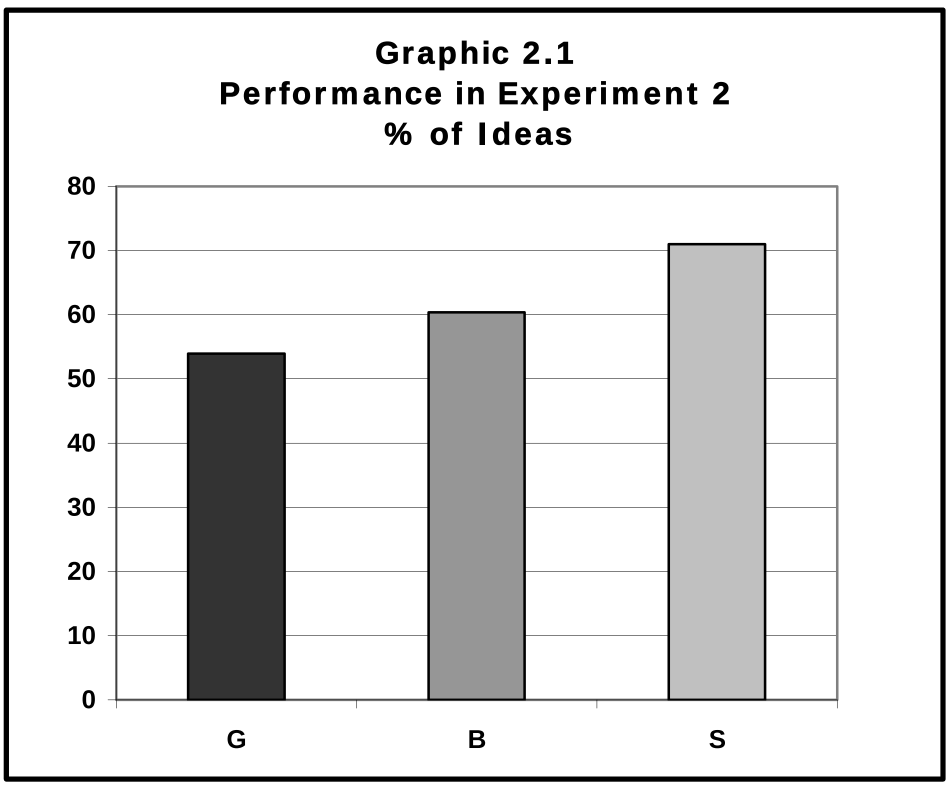
Thus, *performance* on the task or the measurement of how well students had succeeded in extracting the main ideas for the task, being this a transitory training effect more than a learning level in itself. Second, to test the surface construction of a mental model with the main ideas present in each of the documents we used the *true/false recognition&understanding of ideas questionnaire*. Finally, to assess the degree to which students had integrated the main units of information present in the documents into a higher-order mental model we used the *application&transfer practical case.*

Having these measures as dependent variables, we conducted one-way Anovas, with independent variable *Task* (*Global task vs. Brief questions vs. Specific questions*). As we did in Experiment 1, we also conducted Pearson correlations among these learning measures and also including previous background knowledge to test if they were actually measuing learning, though at different levels.

Performance results. We found no significant results for the performance measure (*see Table 2.2*), though the pattern of results clearly follows our predictions for each kind of task (*see Graphic 2.1*). Indeed, specific questions were the ones having a higher performance score, following brief questions and the global task. As we argued earlier, specific questions focused students attention on the very elementary units of information in each of the documents. Moreover, the solving process was easier because students only needed to locate this atomic units of information and almost copy them from the documents. Hence, the probability to have better performance scores was higher, as apparent in the present analysis.

TABLE 2.2 PERFORMANCE IN EXPERIMENT 2		
	% Ideas in Answers	
	M	SD
G	53,97	26,10
B	60,38	16,94
S	71,02	22,65

Table 2.2. Performance in Experiment 2



Graphic 2.1. Performance in Experiment 2

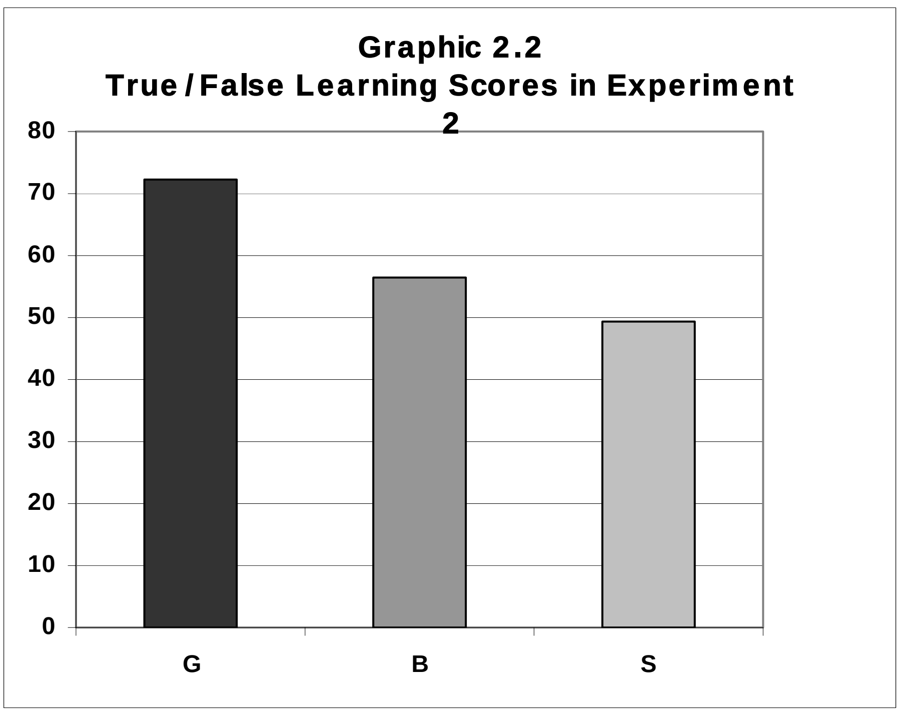
On the other hand, the global task and brief questions were closer in performance scores, in comparison to specific questions. The global task and brief questions yielding similar performance scores was a result we already obtained in Experiment 1 and, appart from indicating that differences between these two task are not located at this shallow level, it may also indicate that both tasks may have something in common and not shared with specific questions.

True/False questionnaire results. Results for the surface construction of a mental model based on the identification and understanding of the main points present in each document yielded a significant effect for task (see Table *2.3*), F (2,20) = 3.71, $p < .05$. To exactly now were differences were located among tasks, we conducted *Tukey honestly significant difference post-hoc tests at 0.05 of significance*. Results indicated that both the global task and brief questions were similar in the true/false learning scores and both significantly differed from specific questions, having specific questions the lowest scores.

Therefore, both the global task and brief questions obtained significantly similar true/false scores and always higher than specific questions. It should be noted that the difference between the global task obtaining higher true/false scores and brief questions was almost similar to the same difference obtained in Experiment 1 in the no-think-aloud condition. Thus, despite the global task yielded the best surface scores, brief questions did not differ so greatly from the global task at this learning level. However, specific questions were the ones producing the worst results, as we had expected.

TABLE 2.3 FINAL LEARNING MEASURES IN EXPERIMENT 2				
	% True/False Questionnaire		% Application case	
TASK	M	SD	M	SD
G	72.22	17.05	61.11	25.19
B	56.34	19.62	34.92	11.87
S	49.30	14.97	30.55	9.84

Table 2.3. Final Learning measures in Experiment 2



Graphic 2.2 True/false Learning scores in Experiment 2

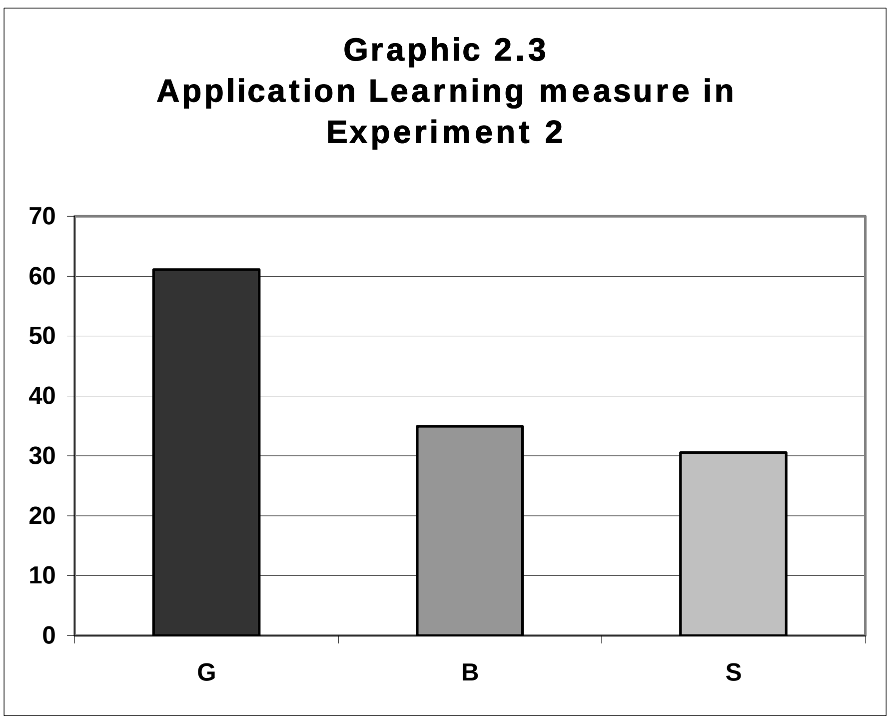
From this surface-level results we can conclude that even at this surface-learning level the task that seemed to be the most effective was again the global task. However, results did not differ so greatly from brief questions scores. It might be possible that the reason why the global task had the highest scores at this level was because a precise identification and understanding of the main points present in the documents was needed to actually integrate these main points into a higher-order representation. Therefore, by constructing a precise text-base, higher-level integration processes had greater chances to take place. We would say that this surface-level construction would be a prerrequisite for higher-level integration processes to occur.

On the other hand, brief questions were able to yield very similar results at this level because they were precisely designed to make students identify the main units of information making up a mental model of how bacteria become resistant to antibiotics. However, with fewer chances to go beyond a construction of an interrelated mental model, as it happened with the global task. Finally, as we expected, specific questions could not even help students in correctly understanding the main ideas in the documents. The mental processes specific questions induced were located at such a shallow level (i.e., copy&paste) that no information processing activity had the chance to ever take place. This is the main reason why, whereas both the global task and brief questions seemed to be equally effective at this learning level, specific questions failed to help students in constructing this superficial mental representation.

Application&Transfer practical case results. Finally, when we considered the measure that would assess the degree to which an integrated mental model from the multiple sources had been built, results were significant for the independent variable task, F (2,20) = 7.27, p < .05. (*Table 2.3*). Results were also completely parallel to those found in Experiment 1, in the no-think-aloud condition for the global task and brief questions, respectively. Hence, the global task scored again the highest at this learning level, being followed by brief and specific questions, in this order. To test were significant differences were located among tasks, we again conducted *Tukey honestly significant difference post-hoc tests at 0.05 of significance*. Results confirmed our hypothesis: the global task was the one scoring the highest at this level and both brief and specific questions significantly differed from the global task, having lower *application&transfer* scores (*see Graphic 2.3*).

As apparent in Graphic 2.2 and 2.3, whereas brief questions were closer to the global task when measuring a surface construction of a mental model from the sources, this time the difference was much bigger. Indeed, this time brief questions were closer to specific questions in the lower ability to promote deep learning and the construction of an integrated higher-order representation from the multiple sources. Therefore, we can confirm our main hypothesis for the present dissertation that the most beneficial task to learn from multiple sources is a global task, which focuses students' attention on the related documents and promotes integration processes.

Contrarily, whereas brief questions were equally effective as the global task to promote a surface construction of the main ideas raised in the documents, they proved now less effective in fostering deep learning and integration from multiple sources. Finally, following our expectations, specific questions could not promote integration from multiple sources either. In fact, it was the task that had the lowest scores at this level, as we had hypothesized. All this learning effects for each kind of task confirmed our general hypothesis for both Experiment 1 and 2 and should have, similarly to Experiment 1, their corresponding on-line behavioral correlates, as we will present in subsequent analysis.



Graphic 2.3. Application Learning measure in Experiment 2

Pearson Correlations among learning measures. Similarly to Experiment 1, we conclude this set of measures on the learning effects of different kind of tasks when working with multiple sources by including a complementary analysis to test the degree to which the three learning levels we hypothesized were related or not, and if relations among variables were in agreement with Experiment 1 results. For this purpose, we conduted *Pearson Correlations* among the following variables: *Performance scores, True/false questionnaire scores, application&transfer scores and, finally, previous background knowledge scores.* Results can be observed in *table 2.4* . Correlations marked with an asterisk were significant at 0.05.

TABLE 2.4
PEARSON CORRELATIONS AMONG OFF-LINE MEASURES
EXPERIMENT 2

	True-False	Application	Background Knowledge	Performance
True-False	1.00	0.41*	-0.12	-0.23
Application	0.41*	1.00	0.15	0.01
Background Knowledge	-0.12	0.15	1.00	0.17
Performance	-0.23	0.01	0.17	1.00

Table 2.4. Pearson correlations among off-line measuresin Experiment 2

Just as we found in Experiment 1, the true/false questionnaire and the application&transfer practical case were significantly related at 0.41. Thus, it provided further evidence for both learning measures assessing final learning, though at different levels, as we have constantly hypothesized. On the other hand, we had expected to find the same trend as in Experiment 1, for performance scores to be inversely related to application scores, but positively related to true/false scores. We did not find the previous trend. Indeed, the trend we found this time seemed to be nearly the opposite. In any case, performance scores did not go beyond the 0.01 positive relationship with learning measures, which again gives us some empirical support that performance is something different from learning, either at a more superficial or deeper level.

(c) Measures of how students distribute their time in the experimental session.

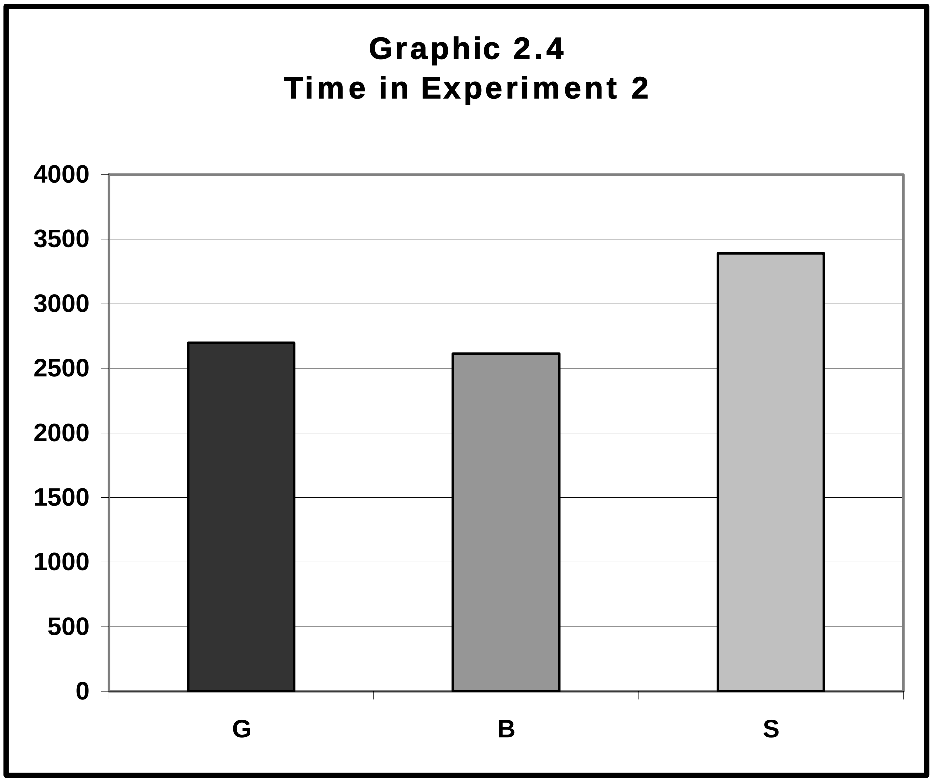
As we did in Experiment 1, now we will start presenting on-line time-based measures to track how students in each kind of question behaved when reading the texts and performing their assigned tasks. This on-line evidence, similarly to Experiment 1, should help us in concluding which kind of tasks promote integration from multiple texts or not, to which degree and why. We will also start by presenting a a global approach based on time measures of how students distributed their time across actions in the Experiment: *Overall time in the Experiment, Time reading Instructions, Time reading Text, Time reading question/s and, finally, Time Answering the question/s.* With these dependent measures we conducted one-way Anovas, having as independent variable task (*Global vs. Brief vs. Specific*).

Overall Time in Experiment. Statistical results for overall time in Experiment were, sadly, non-significant. And we regret this lack of significance because the trend observed in data exactly followed our predictions for each kind of task and was in continuity with results obtained in Experiment 1 (*see Table 2.5*).

	TABLE 2.5 ON-LINE TIMES IN EXPERIMENT 2									
	Time in Experiment		Time in Instructions		Time in Text		Time in Question		Time in Answer	
	M	SD	M	SD	M	SD	M	SD	M	SD
G	2697,80	654,41	51,13	29,31	1260,46	212,97	129,80	34,57	1307,53	554,64
B	2613,37	829,09	65,04	30,36	1146,67	377,14	218,81	99,49	1247,88	472,62
S	3390,52	917,26	58,51	28,00	1788,91	615,97	219,43	94,75	1382,17	286,28

Table 2.5. On-line Times in Experiment 2

In fact, similarly to results obtained in Experiment 1, the global task and brief questions seemed to require a similar amount of experimental time to be performed from the multiple sources. Mean experimental times for both tasks were even nearly the same as those obtained in the no-think-aloud condition in Experiment 1. On the contrary, specific questions were the tasks that required the highest amount of time to be performed (see *Graphic 2.4*). Considering the low learning results simultaneously to this increased time present in specific questions, we could say that specific questions were in fact the less economical kind of questions to work with multiple sources.



Graphic 2.4. Time in Experiment 2

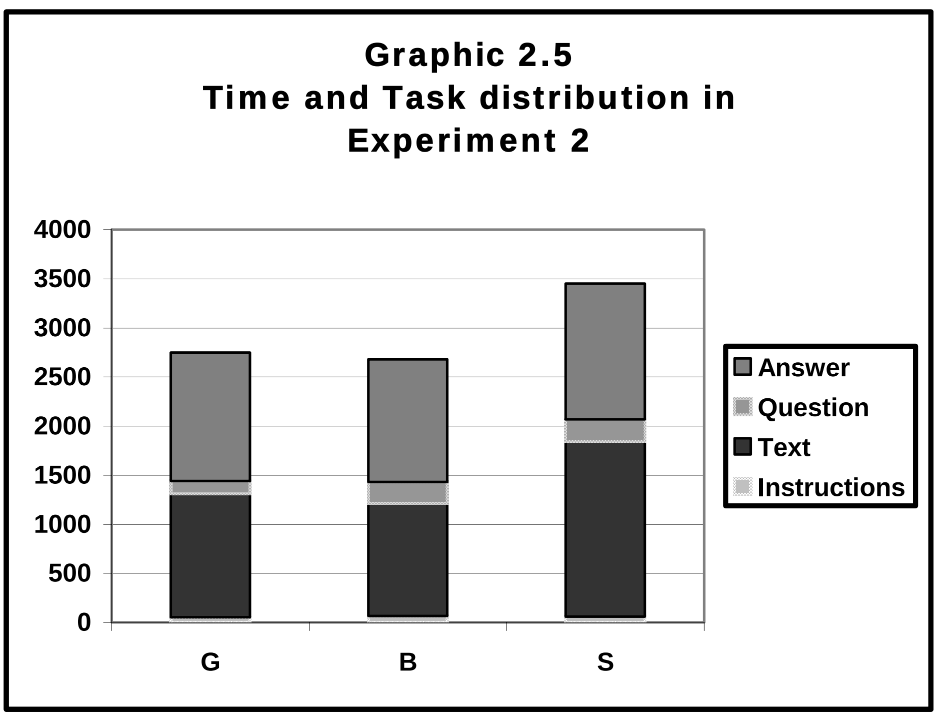
Time in Instructions, texts, question/s and answer/s. With these measures we analyzed how students performing each kind of task distributed their time across the experiment, expecting to find differences depending on the kind of questions that would clarify our learning results (*Table 2.5*). We only find one significant difference for the dependent variable *Time in Text*, F (2,20) = 4.73, p < .05 and one marginal effect for *Time reading the questions*, F (2,20) = 3.18, p = .06. Regarding the first result, whereas time in text was very similar for the global task and brief questions, it was much higher in the case of specific questions.

To contrast were significant differences were located, we again conducted *Tukey post-hoc tests* at 0.05 significance. Results indicated that, in fact, the main difference was located in specific questions, in comparison to both the global task and brief questions (*see Graphic 2.5*). Results for the global task and brief questions needing similar global times in reading the multiple sources were exactly the same as in Experiment 1 and provided us with on-line evidence why both task may have been equally effective when considering a surface level of learning. Thus, if the time invested to inspect and process the textual information was nearly similar, this result may explain why the measurement of the identification and understanding of the main ideas present in the documents yielded similar results for both tasks.

In contrast, specific questions made students spend much more time in reading the sources. Due to the low-learning effects obtained for specific questions and to the higher performance scores, we assume that this increased reading time was not used to process and understand information, but to find and locate the concrete answers in the three sources, to be able to extract them for the questions. A more sophisticated, though non-economical, question-answering process which actually required students to be a longer time performing the tasks (*Graphic 2.5*).

For the time reading the question, we also conducted *Tukey post-hoc tests* to find out were differences exactly yielded. Significant results at .05 were obtained for the global task needing less time to process the demands of the task, in comparison to both brief and specific questions needing longer. This result is parallel to that obtained in Experiment 1 for the global task.

Hence, it seems that performing a global task is indeed a more effective kind of task as performing brief or specific questions, as it does not require the student to invest an excessive amount of time processing the demands of the task. It might be argued that both brief and specific questions needing longer time to read the questions was due to the higher amount of questions in these two conditions. As we argued earlier, we should consider this result in terms of economy of resources. Despite the global task was only one, whereas brief and specific questions were more than one, the contents that the questions were asking for were exactly the same. Therefore, the global task proved more effective in making students waste no additional time to process the task demands.



Graphic 2.5 Time and Task distribution in Experiment 2

In sum, both the global task and brief questions students needed a similar amount of time to perform the experiment and its divisible components, though the global task made students spend less time processing the task demands. These results were completely parallel to those obtained in Experiment 1. In contrast, specific questions were the kind of task needing overall a higher amount of time for the experiment, and, especially, to read and inspect the multiple sources. We argue that specific questions were the most non-economical kind of tasks, as they made students spend the highest amount of time in the experiment and inspecting the sources, though yielding the lowest learning results.

Now, similarly to experiment 1, we would need a more precise kind of on-line evidence to be able to state that specific questions were the most inefficient tasks to learn from multiple sources because they did not promoted integration of information at all, whereas the global task produced integration of information in a greater extent than brief and specific questions. The following set of analysis would allow us clarify this questions, as well as clearly define the status of brief questions in learning from multiple sources.

(d) *Measures of how students read and integrate information from multiple sources.*

Therefore, similarly to experiment 1, we present an additional set of on-line evidence to analyse in detail the integration of information process in multiple sources based on reading speed rates. We also considered: (1) the processing time of the question and the processing time of textual information, relevant and non-relevant and, (2) the reading speed ajustment across the experiment (reading speed across experiment).

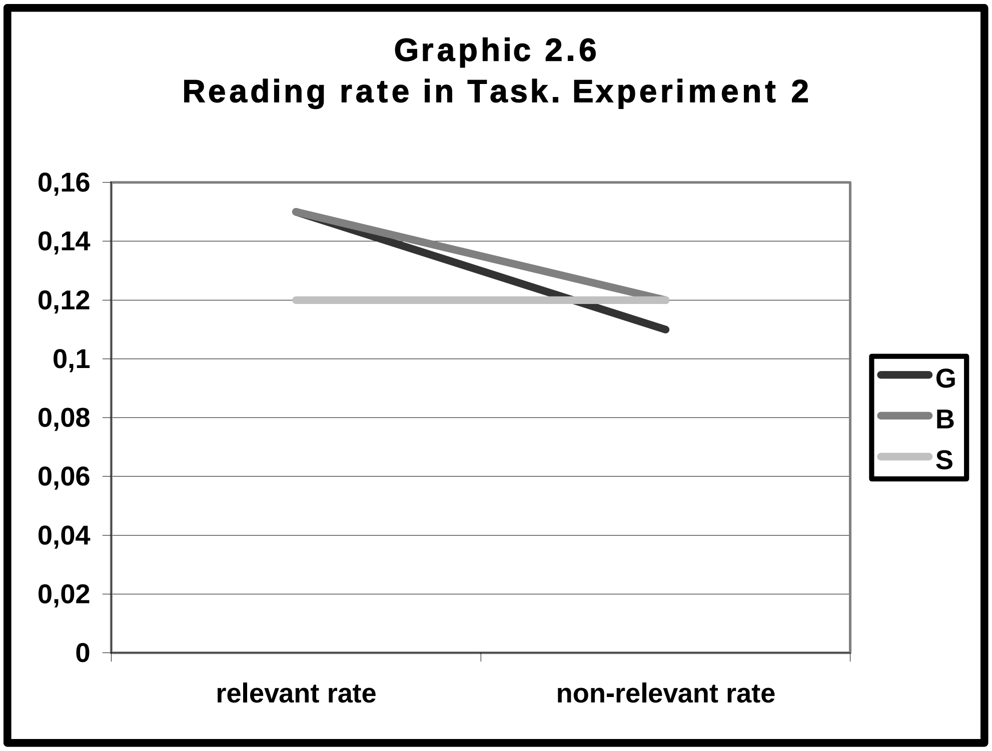
Processing time of question and textual information. As we did in Experiment 1, to analyse how fast or slow students had processed the demands of their task, we conducted one-way between-subjects Anova, with independent variable *task* (*Global vs. Brief vs. Specific*). The dependent variable was the mean reading rate per word in reading the questions. Results were non-significant, exactly the same as in Experiment 1. Hence, we again confirm our finding that differences among tasks were not located on how fast or slow students processed and understood the task demands (*see Table 2.6*).

TABLE 2.6
READING RATE IN EXPERIMENT 2

	Question reading rate		Relevant reading rate		Non-Relevant reading rate	
	M	SD	M	SD	M	SD
G	0,41	0,12	0,15	0,04	0,11	0,01
B	0,38	0,16	0,15	0,06	0,12	0,05
S	0,48	0,13	0,12	0,03	0.12	0.05

Table 2.6. Reading rate in Experiment 2

When we considered the repeated-measures variable *Reading rate*, to analyse how relevant in comparison to non-relevant information was processed, interesting differences appeared among tasks. With this repeated-measures variable we conducted one Mixed Anova with between-subjets independent variable *task (Global vs. Brief vs. Specific)* and with the repeated-measures variable *Reading speed rate* (*relevant speed rate* vs. *Non-relevant speed rate*). Results were significant for the repeated measures interaction Reading rate x Task, F (2,20) = 5.83, $p < .05$. Mean relevant and non-relevant reading rates can be observed in *table 2.6* and in *graphic 2.6*.



Graphic 2.6. Reading rate in task. Experiment 2

Several conclusions can be drawn from these results and can be clearly observed in *graphic 2.6*. The first of them is that the global task continued to be the task that promoted a better adjustment in the reading speed of both relevant and non-relevant information. Hence, whereas relevant information was read more slowly, non-relevant information was read much faster.

Similarly to experiment 1, we argue that this result may be due to a more detailed processing of relevant information and integration among units of information taking place (Mannes & Hoyes, 1996). On the opposite situation, specific questions created a completely flat reading pattern, not discriminating at all the reading of relevant or non-relevant units of information.

We hypothesize that this reading pattern was due to the question-answering process promoted by specific questions. Thus, specific questions made students only locate very atomic units of information from the documents, with no need to understand but only copy the information found and matching the wording of the question. For this reason, specific question students displayed an homogeneous reading pattern which did not adjusted reading speed to process important information. Indeed, no deep processing activity was needed to perform the task. Finally, in the absence of thinking-aloud interfering in overall results, as it happened in Experiment 1, the brief questions reading pattern seemed to vary to some degree to that found in Experiment 1. In fact, relevant reading rate was now similar to that obtained for the global task. However, the adjustment between relevant and non-relevant reading rate was still higher for the global task and lower for brief questions. Indeed, the non-relevant reading speed was exactly the same for brief and specific questions, but much lower for the global task.

Consequently, the global task continued to be the task that promoted a greater adjustment in the reading of relevant and non-relevant information. Brief questions, although they promoted a similar relevant speed rate, they had a higher non-relevant rate, similar to that for specific questions.

Thus, the adjustment occurred in a greater extent for the global task. Finally, specific questions presented a completely flat reading pattern which showed no evidence of efforts to process and integrate information at all. The fact that brief and specific questions had similar and higher non-relevant reading rates may be explained by the question-answering process when the students have to locate answers in the texts for four or more questions. As we have discussed earlier, the need to search for answers may make students even read non-relevant information at a considerable speed.

From all these results we want to emphasize that the role of brief questions appears more clearly as we present the diverse results obtained in Experiment 2. Thus, in the absence of thinking-aloud interfering in overall results, brief questions approach to the global task in the ability to foster a detailed processing of the main units of information present in the documents. For this reason, surface learning results were almost similar. However, brief questions still failed at promoting deep comprehension of the multiple sources and integration of information, as it was apparent in the deep learning measures. Brief questions still not adjusting reading rate according to relevance so much as the global task provides, again, evidence that brief questions still differ with the global task in their capability to promote deep level learning processes from multiple sources.

Reading speed adjustment across the experiment (Reading speed across the Experiment). As we did in Experiment 1, we wanted to complete this set of on-line evidence by analysing how students adjusted their reading speed across the experimental time. These analysis would shed light on the kind of reading pattern promoted by each kind of task, similarly as it did in Experiment 1. We again used the repeated-measures variable *reading rate across experiment*, which was subdivided into *Time 1* (i.e., mean reading rate in the first third), *Time 2* (i.e., mean reading rate in the second third) and *Time 3* (i.e., mean reading rate in the third third). We conducted one Mixed Anova with between-subjects independent variable *task (Global vs. Brief vs. Specific)* and with the repeated-measures variable *reading rate across experiment (Time 1, Time 2 and Time 3).* Means and standard deviations for this analysis are presented in *table 2.7*.

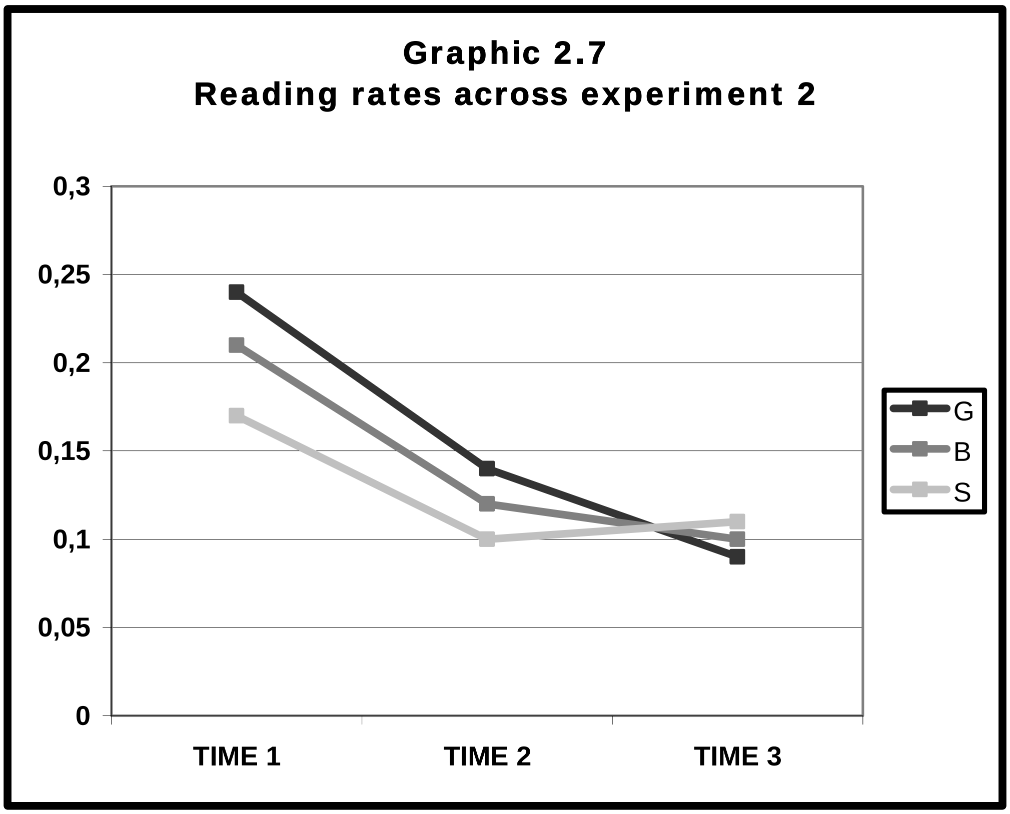
TABLE 2.7
READING RATE ACROSS EXPERIMENT 2

	TASK	MEAN	SD
TIME 1	G	0.24	0.05
	B	0.21	0.11
	S	0.17	0.03
TIME 2	G	0.14	0.06
	B	0.12	0.03
	S	0.10	0.03
TIME 3	G	0.09	0.03
	B	0.10	0.03
	S	0.11	0.06

Table 2.7. Reading rate across Experiment 2

Results yielded a marginal effect for the interaction of Reading rate across Experiment and task, F (4,40) = 2.26, p = 0.07. (*Table and graphic 2.7*). Hence, the global task was the task that promoted higher reading rates in time 1 and 2, however the lowest in time 3.

Brief questions followed the global task by having lower reading rates in all time 1, time 2 and time 3, similarly as we found in Experiment 1. Finally, specific questions created a curious reading pattern having the lowest reading rate in time 1 and 2, however the highest in time 3, exactly the opposite to the global task.



Graphic 2.7. Reading rates across Experiment 2

First of all, though between-subjects effects for task were non-significant, it should be remarked that the task promoting overall a more detailed and slowly reading of textual information, especially in times 1 and 2, was again the global task. Therefore, the global task presents a consistency in results from experiment 1 in promoting a more detailed processing of textual information. As we interpreted in Experiment 1, this reading slowness should be explained in terms of information being deeply processed and, probably, integrated in the multiple texts reading (Mannes & Hoyes, 1996). And, mainly in the first half of the experiment, when textual information is read for the first times and a mental model from the multiple sources is being built. On the contrary, both brief and specific questions promoted overall faster reading times, mainly in time 1 and 2, confirming our hypothesis that the more specific the question, the more superficial processing of texts it promotes.

Moreover, both specific and, in this order, brief questions had the highest reading speed rates in time 3, whereas the global task had the lowest. Thus, the global task, as we also found in Experiment 1, promoted a reading pattern based on a detailed reading in the first half of the experiment and a progressive linear decrease in reading speed, thus becoming faster, from the second half of the experiment to the end, having the fastest reading speed at the end of the experiment. Additionally, this decrease throughout the experiment was quite homogeneous, in comparison to brief and specific questions. By analysing complementary on-line evidence, we had concluded in experiment 1 that the more slowly reading in times 1 and 2 had been probably used to process and combine relevant information in detail, whereas de fast readings in the last third were mainly used to consolidate the mental model built during the first readings. The same interpretation was completely valid for Experiment 2 and we considered this result to be strong evidence for an integration reading pattern taking place.

Brief questions, on the contrary, increased the reading speed in the last third above the global task, a result which had not occurred in Experiment 1 probably due to the interference of thinking-aloud in the analysis. In Experiment 2, thus, our expectations for a different reading pattern for brief questions appear more clearly. And they appear in the form of a more superficial processing pattern, with fewer chances for integration to take place. A more superficial processing pattern in that information was read much faster at earlier stages of the experiment and, at the end of the experiment, it was even read more slowly than in the global task. Hence, there not being such a clear final revision pattern to consolidate a mental model as in the global task.

Regarding specific questions, they promoted the most superficial reading pattern from all three tasks, according to our hypothesis for this task and completely opposite to the reading pattern in the global task. Therefore, they promoted the fastest readings all in time 1 and 2 and the biggest increase in reading speed in time 3. The reason why both brief and specific questions increased reading rates in the last third, thus not promoting such a clear revision pattern as in the global task, might have been due to the nature of the question-answering process. Because answering brief or specific questions required the reader to consider isolated tasks having isolated units of information to be located for the questions, it could have been possible that students read at a considerable speed relevant units of information for a question, even approaching the end of the experiment.

In general, results followed our expectations for each kind of task and provided additional confirmation that the task promoting integration of information and a more detailed processing of textual information was again the global task. On the other hand, it should be noted that all three tasks promoted a decreased reading pattern, from time 1 to time 3. We argue that this is a generalized reading pattern derived from the progressive familiarization with the texts in the course of the experiment.

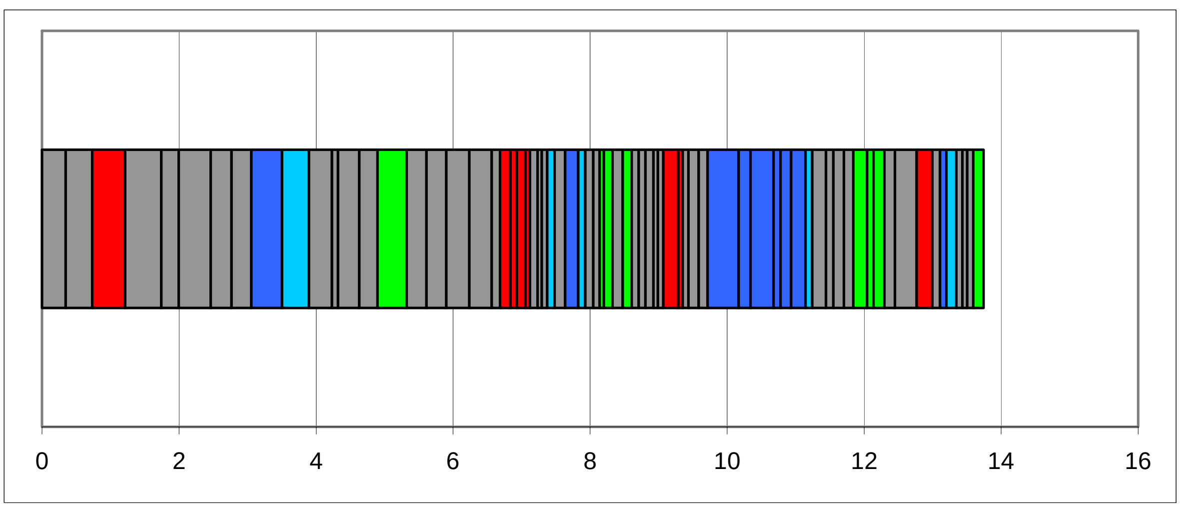
(e) *Visual analysis of the combination of information process*

In previous analysis we have obtained on-line evidence for the learning effects found for each kind of question and re-confirmed the beneficial effects of global tasks, by promoting a more detailed processing of textual information and integration processes in a greater extent than other tasks. Now, similarly as we did in Experiment 1, we present a complementary analysis to further enrich results presented so far and try to understand what integration of information consists of in terms of mental processes being displayed when reading multiple sources. It consists of a visual analysis of how students actually combined information from the multiple sources in their reading sequences across time.

This time, we have only selected three cases, one belonging to each condition, to test if combination of information patterns found for both the global task and brief questions do replicate in this experiment, and, on the other hand, show visual evidence that specific questions are the tasks promoting a more superficial processing of texts lacking of combination of units of information.

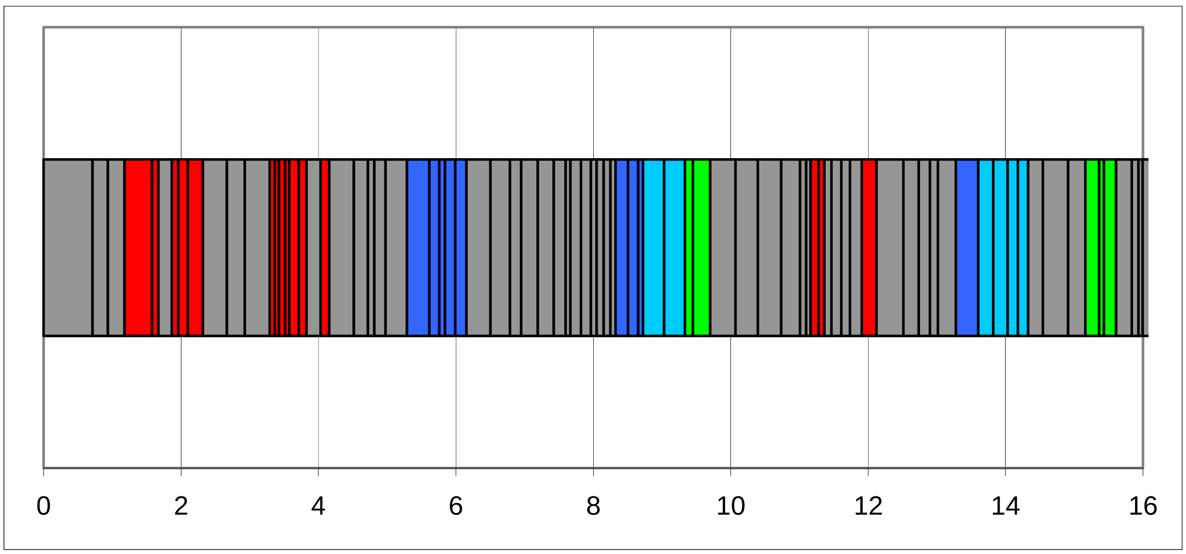
Global task selected case. In Experiment 1 we had found that global tasks students actually presented reading patterns in which combination of information was apparent. Mainly because relevant information was slowly read in the first stages of the reading sequence and, in the last third, there was a dramatic decrease in reading speed rate, becoming readings much faster. Interestingly, this last decrease in reading speed was used to revise all relevant segments making up a mental model of how bacteria become resistant to antibiotics. We interpreted this pattern as information being deeply processed and integrated in the first stages of the reading sequence and the mental model initially built in the first readings being consolidated in the last revision of all the relevant segments. We argued that this was evidence of integration taking place. Integration of information understood as the processing, connection and combination of all the relevant points to understand why bacteria become resistant to antibiotics.

We exactly found the same pattern in our selected case in Experiment 2 (*Reading sequence 2.1*). As it can be observed in reading sequence 2.1, the reading pattern presents three main reading cycles and a final revision. Each cycle contains a linear revision of all the relevant segments present in each of the texts. Cycle 1 was located in the first third of the reading sequence, in which all textual information was slowly read, because it was the first readings of the textual information. The student encountered the main relevant units of information for the first time and he or she carefully read them to process the information in detail. Cycles 2 and 3 occur in the second third of the reading sequence. In each of the cycles, all relevant segments were linearly revised, at a faster speed. The amount of non-relevant information was also reduced. Finally, there was a final fast revision of all relevant segments in the last third of the experiment. Hence, similarly to experiment 1 results, the reading pattern showed a more detailed processing of relevant information during the first readings and a final faster revision of all relevant segments making up a mental model of how bacteria become resistant to antibiotics. We hypothesize again that during the first and second third of the reading sequence information was being integrated to make up a mental model of bacteria resistance and the final revision of all segments was a kind of consolidation of this mental model.



Reading sequence 2.1: Global task

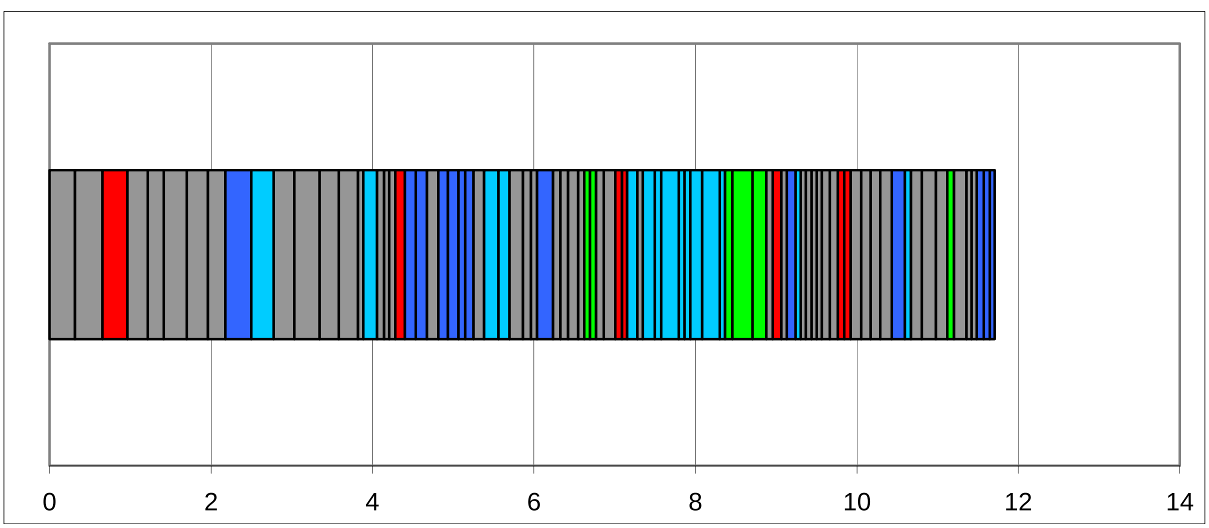
Brief question selected case. For the brief questions selected cases in Experiment 1 we had found a linear reading pattern, generally at a faster speed than the global task, and with no signs of combination of information. Results in the brief question selected case in the present experiment show a similar pattern (*Reading sequence 2.2*).



Reading sequence 2.2: Brief question

Thus, our selected case presented a more homogeneous reading pattern, being reading speed more similar in all the sequence and faster than in the global task example. Moreover, this student revised the relevant segments for the four questions linearly in two clear cycles. Cycle 1 taking place in the first half of the reading sequence, in which the student started reading the three texts and encountered the relevant segments for the first time. Cycle 2, in the second half of the sequence, in which the student almost replicated reading cycle 1, revising the relevant segments in the same order and at a similar speed. Cycle 2 also included a remarkable amount of non-relevant information. In sum, the brief question example seemed to replicate that analysed in Experiment 1 and it provided evidence to the hypothesized superficial processing associated to brief questions and the fewer presence of combination of information processes in this kind of questions.

Specific question selected case. Finally, previous learning and on-line evidence had demonstrated that specific questions were the less beneficial tasks in learning from multiple sources. They promoted a highly homogeneous reading pattern, without discriminating relevant and non-relevant information and presenting the fastest readings of textual units of information from all the kinds of tasks. Integration of information could not be present at all, because specific questions only promoted an iterative *copy&paste* process to answer 10 concrete questions. The visual analysis derived from the specific selected case confirms these findings (*reading sequence 2.3*).



Reading sequence 2.3: Specific question

What is visually apparent in reading sequence 2.3 is that the reading process was the fastest from the three kind of tasks. Similarly to brief questions, it was also homogeneous. Even more homogeneous than brief questions, as previous on-line measures had also demonstrated. On the other hand, the reading pattern was much more chaotic in the revision of relevant units of information. Although some regular iterations could be observed, there were goings back and forth reading the same segment. Had it occurred in the global task, we would have interpreted it in terms of combination of units of information. Because the process to answer the 10 independent specific questions only required the independent location of each answer, we interpret it as search iterations to look for the specific units of information.

In any case, the most remarkable feature was the extreme fast reading of all textual segments from the beginning of the reading sequence and the lack of a final reading speed adjustment to revise all segments of relevant information. For all these reasons, we also obtained complementary evidence to be able to state that specific questions were indeed the most inefficient task to learn from multiple sources, in contrast to the global task promoting a detailed processing of textual units and combination of information pattern. As an intermediate category, we had brief questions promoting an isolated understanding of the main points making up a mental model of bacteria resistance.

3.3 Experiment 2: Discussion

The reasons why we conducted a second complementary experiment to the first one were the following: a) to try to replicate results obtained in Experiment 1 regarding the effects of a global task and brief questions in learning from multiple sources, all in the absence of thinking-aloud interfering in results; b) to test the inclusion of a more concrete kind of task from which we expected that no integration processes could be displayed; c) to clarify the role of brief questions in comparison to other tasks to learn from multiple sources. Having these objectives in mind, we conducted Experiment 2, eliminating the presence of students thinking-aloud and including a third kind of task (i.e., 10 specific questions) which we hypothesized would promote a much more superficial inspection of the multiple sources than brief questions.

Overall results confirmed our main hypothesis for the present dissertation that the global task was indeed the most beneficial to promote learning from the multiple sources at a higher-level. The results pattern we obtained, both for learning measures and the behavioral on-line correlates, was exactly similar to that obtained in Experiment 1. Hence, this similarity was strong empirical evidence confirming our main hypothesis for this study.

First of all, the global task had the lowest, though non-significant, performance scores from all three kinds of tasks. In contrast, both the surface learning measure and, especially, the deep learning measure, yielded the highest scores. When considering the surface learning measure, though the mean score was still higher than brief questions, both kind of tasks proved equally beneficial for the identification and understanding of the main ideas present in the documents. However, the global task proved again to be the most beneficial task to construct a higher-order mental model from the multiple sources, as apparent in the *application&transfer* practical case. The explanation we gave for this result is that a correct and detailed processing of all the main units making up a mental model from the three sources seems to be a prerrequisite for the construction of an integrated mental model. For this reason, the task that was more beneficial for this higher-level construction still obtained better scores when measuring the processing of the isolated units of information, though not significantly varying from the brief question scores at this level.

On the other hand, the fact that the lowest performance scores were present in the global task may lead to two main interpretations: a) that the global task was indeed the most difficult task to perform, because students seemed to have more difficulties in extracting all relevant units of information for the task, b) that, despite this initial difficulty in performance, it proved to be the most beneficial task to learn from the multiple sources at a deep level. Therefore, we might agree with previous research that introducing difficulties for the learning can enhance long-term learning (Mannes & Kintsch, 1987, McNamara, Kintsch, Songer & Kintsch, 1996; Mannes, 1994). In any case, this trend for the global task having lower performance scores was not so clear in Experiment 1, because of the presence of thinking-aloud, and was non-significant in Experiment 2. For this reasons, we have to be careful with the conclusions we draw from these results.

In sum, results obtained in Experiment 2 were a confirmation for the beneficial effects we hypothesized for the global task in fostering deep learning from multiple sources. Experiment 2 also provided on-line conformation that the global task produced an on-line performance pattern indicating a detailed processing of information and integration taking place and being completely similar to that obtained in Experiment 1. In short, the global task continued to make students process information globally more slowly than any other task, thus indicating a more detailed processing taking place. Moreover, global task students discriminated and adjusted their reading pattern according to the relevance of information. When considering the reading speed rate across the experiment, global task students presented a very specific pattern: whereas they would read more slowly in the first half of the experiment, there was a dramatic decrease in reading speed rate, thus becoming readings faster, in the last third of the experiment. Similarly as we did in Experiment 1, we interpreted this pattern as information being deeply processed and connected in the first half of the experiment and information being quickly revised in the last third. The visual case analysis confirmed our hypothesis once again in Experiment 2 that global task students used the quick readings in the last third of the reading sequence to read and combine all relevant segments making up a mental model of how bacteria become resistant to antibiotics. All this on-line evidence was, therefore, additional confirmation obtained in Experiment 2 for our main hypothesis in the present dissertation that the global task would be the most beneficial task to integrate information and learn from multiple sources at a deep level.

Similarly, the brief questions' learning and on-line pattern also replicated results from Experiment 1. As we expected, performance scores were slightly higher than in the case of the global task, probably because the presence of a much more specified question-answering process helped students to extract the isolated units of information for the task. Moreover, we had hypothesized that performance results should be this way, as brief questions had been specifically designed to focus students' attention on the isolated units of information present in the three documents. In any case, results were not so clear in Experiment 1 and non-significant in Experiment 2.

On the other hand, as we have discussed above, surface learning results brought brief questions closer to the global task in their ability to correctly identify and understand the main points making up a mental model of bacteria resistance. Indeed, both brief questions and the global task had significant similar scores in this measure, and both of them higher than specific questions. However, brief questions failed at promoting deep learning from the multiple sources, producing similar low scores as specific questions. Therefore, our conclusion obtained in Experiment 1 that brief questions were especially effective to foster an isolated understanding of the main points present in the documents did replicate in Experiment 2.

In addition, we also found similar on-line evidence in Experiment 2 to be able to state that brief questions were not so beneficial for promoting integration of information and deep processing of the multiple texts. However, they were equally effective as the global task to make students understand and identify the main points present in the documents. Thus, brief question students read all textual information faster than the global task, though more slowly than specific questions. This result might indicate that brief questions would be located at an intermediate position in terms of capability to make students process textual information in detail. On the other hand, brief question students read relevant information at a similar speed than global task students, though non-relevant information was read more slowly than global task students. This interesting reading pattern was not found so clearly in Experiment 1, probably due to interference effects of thinking-aloud. Brief question students reading relevant information at a similar speed than global task students could indicate why both groups were equally effective in identifying and understanding the main points of the documents. However, there should have been more in global task students that made them obtain higher deep learning scores. We hypothesize that the greater adjustment in reading relevant and non-relevant information present in global task students, the more homogeneous and descendant reading pattern across time and the combination of information pattern derived from the visual analysis is the key evidence for our claim.

Further evidence to be able to state that brief questions were not so effective in promoting integration and combination of information as the global task was present in that they did not produced such a clear descendant reading pattern across time. Indeed, contrary to the global task, reading rates were higher in the last third of the experiment than the reading rates for the global task.

By visually analysing how students combined information across time, we found out that not only this increased reading rate in the last third was not used to combine information, but it was mainly used to fixate in the reading or readings of one or two relevant segments. Hence, we also confirmed that the reading pattern present in brief questions clearly reflected what brief questions had been designed for, which was primarily the isolated understanding of the main points present in the documents.

Finally, we had included a third kind of task, expecting that it would be the less beneficial task to learn from multiple sources, even more than brief questions. As expected, the new specific questions produced a clear learning and on-line pattern, which was completely the opposite to the global task. First, though again non-significant, they produced the highest performance scores. This trend appeared as we had expected. Specific questions made students only locate very atomic units of information from the documents. Moreover, the location of answers process should be easier, as most of the times the wording of the question matched the wording of the answer. For all these reasons, students answering specific questions were expected to extract the units of information from the documents more easily than students answering brief questions or the global task.

Despite having the highest performance scores, specific questions produced the lowest learning scores from the multiple sources, both at a more surface or deeper level. As we had hypothesized, specific questions only making students *copy&paste* information from the documents, with little need to process information in detail, would have few probabilities of making students learn from the multiple sources. Hence, becoming the less efficient task to make students learn from multiple documents, instead of brief questions. As previous research has demonstrated, questions that only make students search and locate very atomic units of information have been demonstrated to have nothing to do with comprehension processes being activated (Cerdán et al, submitted; Cataldo & Oakhill, 2000).

For all these performance and learning effects we also had on-line evidence. Thus, we found a trend for specific questions making students spend much more time in the experiment than any other task. Because students had to answer 10 questions and locate very specific units of information, they may have needed a longer time to inspect the sources. This longer time produced as effect higher performance, or the inmediate success in the task. However, the longer time was not used to process information in detail, as evidence by the low surface and deep learning measures. When we analysed the components of the experiment in detail, we found that specific questions especially produced longer times in reading the textual information and processing the demands of the ten questions. All this overall time measures made us conclude that specific questions are indeed inefficient tasks to perform from multiple sources, requiring more time than any, though yielding the worst learning results.

Moreover, specific questions produced no adjustment at all in the reading of relevant and non-relevant information. In fact, the relevant reading pattern was completely flat. We argued before that this might have been due to the search process induced by specific questions, in which very atomic units of information have to be located, with no need to understand. Finally, the reading pattern across time and the visual combination of information pattern provided further evidence that the reading process for specific questions was much more superficial, with no signs of integration of information at all.

In conclusion, Experiment 2 results have clearly fulfilled the goals we raised for this last experiment. First, we have been able to consolidate and replicate results obtained in Experiment 1, hence providing stronger empirical support for our claim that the most beneficial task to learn and integrate information from multiple sources is the global task. On the other hand, the inclusion of more specific questions yielding the worst learning results has enabled the reconsideration of the status of brief questions, as questions that are as effective as a global task to understand a set of relevant points present in multiple sources, however less effective in promoting the construction of a higher interrelated mental model. The strength of the global task for this precise last purpose makes it the most convenient task to foster integration processes from multiple expository texts.

4. GENERAL DISCUSSION

Let us go back to the starting point in the present dissertation, when we brought out the need for improving the quality of students' learning from text by using specific learning tasks that had been especially effective in creating long-term learning (Vidal-Abarca, Mengual, Sanjose & Rouet 1996; Vidal-Abarca, Gilabert & Rouet, 1998; Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted). Those learning tasks were questions that make the students connect several and distant units of information in a text via inferences and, because they mainly promote the most relevant processes in constructing a deep mental representation from text (Kintsch, 1998; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983), the learning outcome is increase in deep comprehension and the possibility for students not only to remember the textual contents they worked with to answer the question, but most importantly, the ability to apply this knowledge to other contexts and make new deductions upon this knowledge.

Moreover, answering questions that promoted the connection among ideas via inferences was especially effective, as it implied less time in performance, compared to students who answered more specific questions (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted). Additionally, differences with other kind of tasks were only apparent when deep comprehension measures were considered, but not when performance or surface recall was compared. This led us to consider an important distinction between performance and learning (Schmidt and Bjork, 1992) and to establish the quality of learning only when using measures that directly focus on the construction of a deep and interrelated representation of text.

After reviewing the main results obtained so far on learning through questions from text, we concluded that it was worth making students work with tasks that directly promote deep learning, as they would allow the student be versatile in the use of the acquired knowledge, from simple recognition or remembering, to the application and reasoning in new contexts. And, despite answering questions that need the connections of multiple units of information via inferences seemed to be more resource demanding in terms of cognitive processes, the final outcome was worth the performance effort. Therefore, we agreed with other authors (Schmidt and Bjork,1992; Mannes and Kintsch, 1987; McNamara, Kintsch, Songer and Kintsch,1996) that introducing difficulties for the learning during performance can enhance long-term learning, which is precisely what occurred when students answered high-level questions, in comparison to low-level questions (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, submitted).

We continued our discussion emphasizing the need for research on new ways to improve learning from multiple sources. We argued that, taking into account the increasing relevance of this new perspective in handling with texts, educational psychologists should concentrate on testing learning activities which would foster deep comprehension from multiple sources. In other words, we should find new methods and tasks that prove effective in increasing students' abilities and strategies to obtain and integrate information from multiple sources.

After highlighting this need, we considered that some clarification regarding the concept of multiple sources, integration and the construction of an integrated mental model from several documents should be presented. Therefore, we reviewed the existing literature on multiple sources and concluded that some interesting points had been developed. However, many of these results could not probably apply to other kinds of texts, which were the focus of our interest.

Specifically, we wondered how integration and learning from multiple sources would take place when working with expository texts. These texts were susbtantially different to history texts, which had mainly been used so far in research on multiple texts. On the other hand, we also remarked that there should be tasks that would promote learning from multiple texts, similarly to the case of single-text learning (Perfetti, Rouet & Britt, 1999) and in continuity with findings in studies of integration from multiple historical documents, which had found that argumentative tasks were those that increased integration the most (Wiley & Voss, 1999).

We considered that the main activity in learning from multiple sources would be the integration of complementary information or corroboration across sources (Wineburg, 1991). Therefore, those activities that would promote integration of information and the construction of a higher order representation from all sources (Perfetti, Rouet & Britt, 1999) would be the most effective. Thanks to a strong empirical support on what integration consists of in terms of processing cognitive activity (Mannes, 1994; Mannes & Hoyes, 1996) we were able to shed some light on the concept of integration and present empirical findings and behavioral correlates which have been obtained so far in the context of expository text.

Most interestingly, Mannes (1994) and Mannes and Hoyes (1996) demonstrated that the design of learning situations that enhanced integration of information were the most effective for learning. Therefore, we found strong empirical support that integrating information as the ability to contrast and combine information from different documents was the main factor explaining learning from multiple expository texts. This is a key finding for educational research on this issue. Knowing what to promote is indeed a clear step forward in the design of learning tasks that can improve the quality of learning from multiple expository texts.

Based on this assumption, we designed two experiments to test if a global task, designed to promote integration and combination of information from multiple expository texts explaining why bacteria become resistant to antibiotics, was more effective than answering 4 brief questions or 10 specific questions. Results across the two experiments completely confirmed our expectations. The global task, designed to foster the construction of an integrated mental model from the multiple documents, proved indeed to make students learn and integrate information from the multiple sources at a deep level. On-line evidence, in accordance to our hypothesis and following previous research (i.e., Mannes & Hoyes, 1996), also confirmed that the global task promoted a more detailed processsing of the sources than brief or specific questions.

In contrast, brief questions were not so effective to promote deep learning from the multiple sources. However, when we considered the identification and understanding of the main units of information present in the documents and the capability to extract these units during performance, results were similar to those obtained with the global task. This made us conclude that differences between these two kinds of tasks were especially located when measuring the construction of a higher mental model from the three sources, but not so much when considering more surface levels of processing. A result completely parallel to that obtained for high and low-level questions (Cerdán et al., submitted). Therefore, brief questions proved to be a quite valuable task if we want to promote an isolated comprehension of relevant ideas presented in several documents.

Because brief questions were still designed to favor some kind of inference making in a single document and because they would not be extrictly similar to the low-level questions we had previously used in single-text learning (Vidal-Abarca, Mengual, Sanjose & Rouet 1996; Vidal-Abarca, Gilabert & Rouet, 1998; Cerdán et al., submitted) we included in experiment 2 a third kind of task to contrast to the global task and brief questions all together. As we hypothesized, specific questions, which made students concentrate on very atomic ideas present in the three documents and fostered the most superficial mental processes to extract this information, proved to be the less efficient task to work with multiple sources.

In fact, they showed a performance, learning and on-line pattern which was completely the opposite to that obtained for the global task. Hence, whereas the global task made students learn at the deepest level from the sources because of integration of information being activated, specific questions made students learn the least, by only inducing search and copy processes in working with the multiple documents. And search for information to answer very specific questions has been demonstrated not to foster comprehension processes at all (Cerdán et al, submitted).

Consequently, our research purposes for the present dissertation have been achieved by again demonstrating that those tasks that are the most beneficial to learn at a deep level, either from one text or from multiple texts, are those that make students actively process information. Though this active processing might be more difficult in the performance phase, it is worth the effort for the learning outcome (Mannes & Kintsch,1987; McNamara, Kintsch, Songer & Kintsch,1996). For all this reasons, if we want students to learn and integrate information from multiple expository sources, we should try to use similar kind of tasks.

On the other hand, we should not reduce the value of brief and specific questions for other kind of learning purposes when working with multiple sources. In fact, brief questions have proved especially useful to make students identify and understand individual points present in several sources. There are indeed many learning situations in which integration of information is not so important as a correct understanding of a set of points or arguments. In this precise cases, brief questions or similar kind of tasks would work perfectly well. Finally, even specific questions could serve learning purposes in working with multiple expository texts. For instance, if we want students to be competent searchers for concrete units of information in a set of documents, as it might be the case in some professional areas or even in university contexts.

Other of the main purposes of the present dissertation was to clarify what integration of information consists of in terms of mental processes. Integration of information seems to explain how multiple documents are combined among them. Integration of information should be the mental process explaining the construction of a higher mental model from multiple documents, either historical or expository (Perfetti, Rouet & Britt, 1999) and, finally, integration of information should account for the good learning results obtained in essay tasks after reading multiple historical documents (Rouet, Britt, Mason & Perfetti, Wiley & Voss, 1999) or our better learning results for the global task.

Similarly as Mannes (1994) and Mannes and Hoyes (1996) proposed, we have mainly used time-based measures to demonstrate that if integration is taking place when reading the documents, this should be apparent in how students read this information. In the present dissertation, and especially when students performed the global task, we found that students who were supposed to integrate information from the three sources read globally more slowly than any other group of students and they adjusted their reading according to the relevance of information, increasing reading slowness when reading the relevant units of information.

As Mannes and Hoyes (1996) concluded, this reading adjustment may be explained by a need to integrate information in short-term memory when complementary units of information are simultaneously activated. The *reinstatement* of previously read information and the i*ntegration* with the current information being processed seems to explain why additional processing time was needed for all this integration mental processes. On the other hand, integration of information was also apparent in global task students combining the different units of information in their reading sequence, in contrast to a more linear reading present in brief of specific students. All this complementary evidence was, thus, an obvious indication of why the global task was more effective than the others and a clarification of what integration of information consists of.

Therefore, we have been able to shed some light on which mental processes should be under the denomination of integration, in continuity with other authors (Mannes, 1994; Mannes & Hoyes, 1996). We have also tried to understand in detail which should be the main cognitive process in the construction of a higher order mental model from multiple sources (Perfetti, Rouet & Britt, 1999) and, finally, we expect to have provided strong on-line empirical support to be able to state that learning from multiple sources can be fostered in a greater extent if this integration mental processes are activated, for example by using some kind of tasks. We do not belive that the mere exposure to multiple texts situations is effective in itself to promote integration mental processes or complex reasoning activities (Perfetti, Britt & Georgi,1995).

As the present research has demonstrated, not all kinds of learning activities are equally effective in making students integrate information from multiple sources. In conclusion, the kind of task clearly seems to mediate the extent to which documents are connected and a higher order integrated mental model built from several sources (Perfetti, Rouet & Britt, 1999).

Integrating information from multiple sources seems to be, as apparent by on-line results, a highly-demanding cognitive task, despite finally having the better learning outcomes (e.g., Mannes & Kintsch,1987; McNamara, Kintsch, Songer & Kintsch,1996). This high intrinsic demands, added to the extraneous cognitive demands derived from making students think-aloud, may explain why thinking-aloud hindered learning from multiple sources at a deep level and fostered a superficial concentration on textual information in the documents. Thinking-aloud effects were even more marked when answering brief questions, which should have increased the intrinsic demands of the task, because of the requirements of four questions to be answered instead of one.

According to the *Cognitive Load Theory* (e.g., Sweller, 1994) cognitive resources in short-term memory are limited and splitted up into the demands of the task by itself (i.e., intrinsic load), the demands derived from the instructional design (i.e., extraneous load) and the demands created by the activation of higher-level processes (i.e., germaine load). Thinking-aloud added to the requirements of reading information from multiple documents and the need to solve one or, especially, four brief question may have increased both the extraneous and intrinsic load in short-term memory, thus limiting the possibilities of higher-order processes to be displayed (i.e., germaine load).

Because of this, deep learning was not possible when students thought-aloud, in contrast to those performing silently. We assume that this result is somewhat controversial, as empirical evidence on the learning effects of thinking-aloud has not been consistent as regards to the possible neutral (Crain-Thoreson, Lippman & McClendon-Magnuson, 1997; Fletcher, 1986), beneficial (Loxterman et al.,1994;Coté, Goldman & Saul,1998) or limiting Wade & Trathen,1989; Magliano,Trabasso & Graesser,1999) effects of thinking-aloud. We would like to remark that thinking-aloud affecting the learning outcome was not due to an specific instruction to verbalize to explain (Ericsson & Simon, 1993), nor to give reasons or justifications during performance. In fact students were strictly indicated to verbalize whatever thoughts came to their mind in the process of reading texts and answering the questions, and only the thoughts that would crossed their mind and no others.

We would also like to indicate that previous research on the effects of thinking-aloud was conducted in the context of single-text learning. Because working with multiple documents seems to be a more cognitive demanding task, thinking-aloud might in fact interact with the additional demands made by the need to extract and integrate information from multiple documents and, thus, show the limiting effects on final learning we have found in the present study. In any case, thinking-aloud interacting with the increasing demands of specific learning contexts and limiting the possibilities to learn at a higher level should be further contrasted in future research.

To finally conclude, we are satisfied to have conducted research on multiple sources and investigated the role of specific kinds of learning tasks based on previous research on integrating information from multiple historical documents (e.g., Wineburg, 1991; Perfetti, Britt and Georgi,1995; Rouet, Britt, Mason & Perfetti, 1996; Rouet,Favart, Britt & Perfetti, 1997; Perfetti, Rouet and Britt,1999; Wiley & Voss, 1999;Britt & Aglinskas, 2002). None of this would have been possible without some researchers starting to concern about how comprehension processes displayed in one document would be applied to reading and comprehending multiple texts. Despite the limitations we raised earlier, mainly due to the existing differences between history and expository texts, which are the focus of our interest, we are convinced that the present study is clearly enriched by past research on multiple sources. Similarly, future research on integration and the use of multiple sources should not only not be left apart, but on the other hand, it should profit from past findings to help us clarify what integration of information from multiple sources consists of in terms of mental processes.

5. REFERENCES

Andre, T. (1979). Does answering higher level questions while reading facilitate productive reading? *Review of Educational Research, 49*, 280-318.

Britt, M.A. & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition & Instruction*, *20*(*4*), 485-522.

Britt, M.A., Rouet, J-F., Georgi, M.C. & Perfetti, C.A. (1994). Learning from history texts: From causal anlysis to argument models. In G. Leinhardt, I.L. Beck, & C. Stainton (Eds.), *Teaching and learning in history* (pp. 47-84). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cataldo, M.G. & Oakhill, J.(2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology*, *92*, 791-799.

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R. & Gil, L. (submitted). Tracking question-answering behavior in electronic learning environments.

Chi, M.T.H., de Leeuw, N., Chiu, M-H., & Lavancher, Ch. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.

Coté, N., & Goldman, S. R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. En H. Van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp.169-193). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cote, N., Goldman, S.R., & Saul, E.U (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*, 1-53.

Crain-Thoreson, C., Lippman, M. Z., McClendon-Magnuson, D. (1997). Windows on comprehension: reading comprehension processes as revealed by two think-aloud procedures. *Journal of Educational Psychology*, *89*, 579-591.

Ericsson, K. Anders & Simon, Herbert A. (1980). Verbal reports as Data. *Psychologycal Review, 87*, 3, 215-251.

Ericsson, K.A. & Simon, H.A. (1993). *Protocol Analysis. Verbal reports as data.* The MIT Press, Cambridge. Massachusetts.

Estes, W.K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369-377.

Fletcher, C. R. (1986). Strategies for the allocation of short-term memory during comprehension. *Journal of Memory and Language*,*25*, 43-58.

Goldman, S.R & Durán, R.P. (1988). Answering questions from Oceanography Texts: Learner, Task and Text Characteristics. *Discourse Processes, 11*, 373-412.

Goldman, S.R & Saul, E.U. (1990) Applications for tracking reading behavior on the Macintosh. *Behavior Research Methods, Instruments & Computers*, *22*(*6*), 526-532. Graesser, A.C. & Franklin, S.P. (1990). QUEST: A model of question-answering. *Discourse Processes*, *13*, 279-303.

Graesser, A.C. & Franklin, S.P. (1990). QUEST: A model of question-answering. *Discourse Processes, 13*, 279-303.

Graesser, A.C., Lang, K.L. & Roberts, R.M. (1991). Question Answering in the Context of Stories. *Journal of Experimental Psychology, 120* (3) 254-277.

Guthrie, E.R. (1952). *The psychology of learning*. New York: Harper and Row.

Hamilton, R.J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research, 55,* 47-85.

Hartley, J. & Davies, I.K. (1976). Preinstructional strategies: the role of pretests, behavioral objectives, overviews and advance organizers. *Review of Educational Research, 46*, 239-265.

Hull, C.L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.

Kintsch, W. & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

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

Langer, J.A. (1985): Levels of questioning: An alternative view. *Reading Research Quaterly, 20,* 586-602.

Long, D. L., & Bourg, T. (1996). Thinking aloud: Telling a story about a story. *Discourse Processes*, *21*, 329-339.

Loxterman, J.A., Beck, I.L., & McKeown, M.G. (1994). The effects of thinking-aloud during reading on students' comprehension of more or less coherent text. *Reading Research Quaterly*, 29, 352-367.

Magliano, J.P. & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Discourse Processes*, *21(3)*, 251-283.

Magliano, J.P., & Graesser, A.C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics*, *20*, 193-232.

Magliano, J.P., Trabasso, T., & Graesser, A.C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91,* 615-629.

Mannes, S & Hoyes, S.M. (1996). Reinstating Knowledge during reading: a strategic process. *Discourse Processes*, *21*, 105-130.

Mannes, S. (1994). Strategic Processing of Text. *Journal of Educational Psychology*, *86,* 577-588.

Mannes, S.M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4*, 91-115.

McDonald, D.R. (1987). Drawing inferences from expository text. Unpublished doctoral dissertation. University of New Mexico, Las Cruces, NM.

McGingley, W. (1992). The role of reading and writing while composing from sources. *Reading Research Quaterly*, *27*, 227-248.

McNamara, D.S., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information*. Psychological Review*, 63, 81-97.

Nash, J.G., Schumacher, G.M., & Carlson, B. W. (1993). Writing from sources: A structure-mapping model. *Journal of Educational Psychology, 85,* 159-170.

Nisbett, R.E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *8*, 231-259.

Paas, F., Renkl, A. & Sweller, J. (2003). Cognitive Load Theory and instructional design: recent developments. *Educational Psychologist*, *38*, 1-4.

Perfetti, C.A., Britt, M.A., & Georgi, M.C. (1995). *Text-based learning and reasoning: Studies in history*. Hillsdale, NJ: Erlbaum.

Perfetti, C.A., Rouet, J-F & Britt (1999). Toward a Theory of Documents representation. In van Oostendorp, H. & Goldman, S.R. (Eds.). *The construction of mental representations during reading* (pp. 99-123) Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pressley, M. & Afflerbach, P.(1995). Verbal protocols of reading: The nature of constructively responsive reading. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rickards, J.P. (1979). Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research, 49*, 181-196.

Rouet, J-F., Britt, M.A., Mason, R.A., & Perfetti, C.A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*, 478-493.

Rouet, J-F., Favart, M., Britt, M.A. & Perfetti, C.A. (1997). Studying and using multiple documents in history : effects of discipline expertise. *Cognition & Instruction*, *15*(*1*), 85-106.

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Vol 2, Reading, writing and language learning*. Cambridge, England: Cambridge University Press.

Schmidt, R.A. & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Skinner, B.F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.

Spivey, N.S., & King, J.R. (1989). Readers as writers composing from sources. *Reading Research Quaterly, 24*, 7-27.

Strømsø, H.I., Bråten, I. & Samuelstuen, M.S. (2003). Student´s strategic use of multiple sources during expository text reading: a longitudinal think-aloud study. *Cognition and Instruction, 21* (2), 113-147.

Suh, S. & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols and recognition priming. *Journal of Memory and Language*, *32*, 279-300.

Sweller, J. (1994). Cognitive Load Theory, learning difficulty and instructional design. *Learning and Instruction*, *4*, 295-312.

Sweller, J. (2003). Evolution of human cognitive arquitecture. In B. Ross (Ed.), *The Psychology of Learning and Motivation*. San Diego: Academic Press.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive arquitecture and instructional design. *Educational Psychology Review*, *10*, 251-296.

Tolman, E.C. (1932). *Purposive behavior of animals and men*. New York: Century.

Trabasso, T. & Magliano, J.P. (1996a). Conscious understanding during text comprehension. *Discourse Processes*, *21*, 255-288.

Trabasso, T. & Suh, S. (1993). Understanding Text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes*, *16*, 3-34.

Trabasso, T., & Magliano, J.P.(1996b). How do children understand what they read and what can we do to help them? In M. Graves, P. Van den Broek , & B. Taylor (Eds.), *The first R: a right of all children* (pp. 160-188). New York: Columbia University Press.

Trabasso, T., van den Broek, P., & Lui, L. (1988). A model for generating questions that assess and promote comprehension. *Questioning Exchange, 2,* 25-38.

van Dijk, T.A. & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Vidal-Abarca, E., Gilabert, R. & Rouet, J.-F. (1998). The role of question type on learning from scientific text. *Meeting on Comprehension and production of scientific texts*. Aveiro (Portugal), July 8-10, 1998.

Vidal-Abarca, E., Mengual, V., Sanjose, V. & Rouet, J.-F. (1996). 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, September 4-6.

Wade, S.E. & Trathen, W. (1989). Effect of self-selected study methods on learning. *Journal of Educational Psychology*, *81*, 40-47.

Whitney, P., & Budd, D. (1996). Thinking-aloud protocols and the study of comprehension. *Discourse Processes*, *21*, 341-351.

Wiley, J. & Voss, J.F. (1999). Constructing arguments from multiple sources: tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*, 301-311.

Wineburg, S.S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorical evidence. *Journal of Educational Psychology, 83*, 73-87.

Wixson, K.K. (1983). Postreading question-answer interactions and children's learning from text. *Journal of Educational Psychology, 75* (3), 413-423.

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

Zwaan, R.A. & Brown, C.M. (1996). The influence of language proficiency and comprehension skill on situation-model construction. *Discourse Processes*, *21*, 289-328.

6. APPENDIX

A. MULTIPLE TEXTS.

B. PREVIOUS BACKGROUND KNOWLEDGE TEST.

C. TRUE/FALSE LEARNING QUESTIONNAIRE.

D. APPLICATION&TRANSFER PRACTICAL CASE

- Appendix materials are presented in the original versions.

A. MULTIPLE TEXTS.

"Perspectivas de la resistencia bacteriana a los antibióticos. ¿Una nueva barrera a la globalización?" (G. Sánchez L., profesor titular de la Universidad Nacional de Colombia)"

De acuerdo con la OMS, de los 52 millones de personas que mueren en el mundo cada año, 17 fallecen por enfermedades infecciosas. Algunas infecciones que parecían erradicadas o controladas como el paludismo, la fiebre amarilla, el dengue y la tuberculosis, están resurgiendo. El Comité del Consejo Nacional de Investigación (NCR) de los E. U. afirma que "la resistencia de las bacterias a los antibióticos se ha acelerado enormemente en las últimas décadas tanto en animales como en los humanos". De igual manera, en la Conferencia de la Unión Europea llevada a cabo en Copenhagen (Dinamarca) en 1998, se advierte que en Europa han aumentado los casos de resistencia producidos por Staphylococcus, Enterococcus, Pseudomonas, Enterobacteriaceae (Klebsiella y Enterobacter) a nivel hospitalario, y entre la comunidad Streptococcus pneumonie, Haemophilus influenzae, Moraxella catharralis, Salmonella, Neisseria. En Medicina Veterinaria es bien conocida la baja eficacia de los tratamientos antibióticos frente a la mastitis causada por Staphylococcus aureus.

Cuando aparecieron los antibióticos muchas bacterias desarrollaron mecanismos de resistencia, principalmente de dos maneras:

* Por mutación genética, mediante cambios en los cromosomas que se oponen a la acción de los antibióticos, por ejemplo, modificando el sitio de acción de las tetraciclinas.
* Por creación de piezas de ADN extracromosómico (plásmidos) que codifican mecanismos de resistencia contra los antibióticos, por ejemplo mediante la creación de la enzima penicilinasa contra la penicilina.

Lo grave de este nuevo material genético es que puede ser transmitido a generaciones subsiguientes de bacterias o transmitirse entre bacterias que incluso pueden pertenecer a especies diferentes (de E. coli a Salmonella, por ejemplo).

Independientemente de cómo las bacterias desarrollen sus mecanismos de resistencia a los antibióticos el efecto es que seleccionan las bacterias resistentes. El proceso de selección es simple: cuando las bacterias se enfrentan a un antibiótico, las bacterias no resistentes al antibiótico mueren, pero las células resistentes sobreviven, lo cual ocurre con más facilidad si la cantidad de antibiótico presente es pequeña. Ese es el problema adicional que presentan los residuos de antibióticos en los alimentos de origen animal (leche, carne, huevos). Al ponerse en contacto con las bacterias, intestinales por ejemplo, éstas no mueren pero son capaces de desarrollar mecanismos de resistencia.

Existen cerca de 30 antibióticos comunes para el tratamiento de infecciones en animales y en humanos. Además, bacterias como salmonella y E. coli pueden pasar a los humanos a través de los alimentos de origen animal. Si estas bacterias generan mecanismos de resistencia a los antibióticos cuando están en el animal, esa resistencia puede transferirse a los humanos a través de los alimentos.

Ya sea que los genes resistentes se originen por mutación o por piezas extracromosómicas de ADN (plásmidos), estos genes se han convertido en un problema mundial ya que pueden viajar por todo el mundo a través de personas, animales o alimentos contaminados. Así, cepas de bacterias que se hicieron resistentes en un país pueden aparecer en otro y diseminarse. Se sabe, por ejemplo, de casos de tuberculosis resistente a varios fármacos en E. U. cuyo origen está en otros países, no obstante las estrictas normas existentes en E. U. sobre el uso de medicamentos.

Es lógico que este problema constituya una preocupación internacional debido al impacto sobre la migración de gentes y el comercio de alimentos. El uso irresponsable de antibióticos con animales en un determinado país, la falta de normas al respecto, o el descuido en su aplicación pueden producir impedimentos o barreras no arancelarias para la comercialización de alimentos de origen animal.

Ante la creciente preocupación que está despertando la resistencia bacteriana, se celebró una Conferencia de la Unión Europea en Copenhagen en Septiembre de 1998 sobre "La amenaza microbiana" de la cual surgió un documento conocido como "Las recomendaciones de Copenhagen". Las recomendaciones de la conferencia deberán ser seguidas por la Unión Europea. Las principales recomendaciones sobre uso apropiado de antibióticos fueron las siguientes:

* Los tratamientos deben limitarse a infecciones bacterianas.
* El tratamiento debe ser dirigido al agente causal.
* El tratamiento debe ser dado en dosis óptimas, administrado a los intervalos y duración de tiempo apropiados.
* Se recomienda no utilizar antibióticos para promover el crecimiento de los animales debiendo sustituirse esta práctica por otras alternativas.
* La mejor manera de corregir prácticas inadecuadas en el uso de agentes antibacterianos es la formación tanto de médicos como de veterinarios y de público en general.
* Los antibióticos utilizados con fines terapéuticos deberán ser adquiridos únicamente a través de prescripciones médicas.

“Genética de la resistencia bacteriana” (Centro de Investigaciones sobre Enfermedades Infecciosas / INSP).“

Uno de los problemas de más morbilidad y mortalidad en los centros hospitalarios es la resistencia bacteriana. Esta se presenta cuando el médico necesita combatir las enfermedades infecciosas producidas por bacterias pero ellas no responden al tratamiento de antibióticos. A principios de los años 40 surgió el uso de la penicilina para combatir estas enfermedades. Al poco tiempo de emplear este fármaco empezaron a surgir bacterias resistentes al medicamento con lo cual ya no fue posible combatirlas. Con el paso del tiempo, las compañías farmacéuticas han diseñado un gran número de moléculas antibacterianas para inhibir el crecimiento de los agentes causantes de las enfermedades infecciosas; estos medicamentos han requerido grandes inversiones económicas y largos periodos de investigación.

Desgraciadamente para los humanos, pero no así para las bacterias, estas han logrado generar diferentes alternativas para evitar el efecto letal de los antibióticos convirtiéndose en cepas resistentes. Estas alternativas son básicamente tres. La bacteria: 1) puede modificar químicamente el antibiótico o degradarlo, 2) puede modificar el sitio blanco celular de acción del antibiótico y 3) puede impedir la entrada del antibiótico o expulsarlo del interior celular.

La resistencia bacteriana obedece a cambios en la información genética de la bacteria (mutación) o la adquisición de nueva formación genética a partir de otras bacterias mediante "plásmidos" (una molécula ADN super-helicoidal que codifica para información genética). En general, estos elementos genéticos no indispensables para las bacterias pueden conferir información adicional que les permite contar con ciertas ventajas con respecto a otras, tales como formación de tumores en algunas plantas, degradación de compuestos orgánicos (como el petróleo), producción de toxinas, resistencia a metales pesados y/o antibióticos. En ocasiones los plásmidos que confieren la resistencia a antibióticos la producen para varios de éstos generándose así la multirresistencia.

La batalla contra la resistencia bacteriana es inacabable, ya que dicha resistencia aparecerá cada vez que se use un inhibidor del crecimiento bacteriano (antibiótico). Sin embargo, existen formas de controlarla, por ejemplo: 1) usar vacunas para prevenir las infecciones, 2) en los hospitales, crear comités de vigilancia epidemiología, esto es, la supervisión fiable y eficaz de resistencia a las bacterias que se presenten en los pacientes infectados, 3) uso adecuado de los antibióticos, ya sea en el campo médico para humanos (automedicación e inadecuada prescripción) o veterinario (uso de antibióticos como factores de crecimiento en los animales).

“Resistencia antibacteriana” (National Institute of Allergy and Infectious Diseases. National Institute of Health of USA).”

Los agentes infecciosos farmacorresistentes, los que no mueren ni son inhibidos por los compuestos antibacterianos, son una preocupación cada vez más importante para la salud pública. Infecciones como la tuberculosis, la gonorrea, la malaria y del oído en la niñez son algunas de las enfermedades que están siendo difíciles de tratar debido a la aparición de bacterias resistentes. Muchos médicos afirman que hay infecciones bacterianas que pronto pueden ser intratables.

Además de su efecto nocivo sobre la salud pública, la aparición de bacterias resistentes ocasiona costes cada vez más altos en el tratamiento médico. Tratar infecciones resistentes requiere usar fármacos más costosos o más tóxicos y puede hacer que pacientes infectados permanezcan más tiempo en los hospitales. El Instituto de Medicina de la Academia Americana de Ciencias ha estimado que el coste anual de tratar infecciones resistentes a los antibióticos en los Estados Unidos puede alcanzar los 30.000 millones de dólares.

Un factor clave en el desarrollo de la resistencia bacteriana es la capacidad de los organismos infecciosos para adaptarse rápidamente a nuevas condiciones medioambientales. Las bacterias son organismos unicelulares que, comparados con los pluricelulares, tienen un número pequeño de genes. Incluso una sola mutación al azar de un gen puede tener un impacto grande ya que, puesto que la mayoría de las bacterias se multiplican muy rápidamente, el número de éstas puede incrementarse con gran rapidez. Así, una mutación que ayude a una bacteria a sobrevivir en presencia de un antibiótico se transmitirá rápidamente al resto de la población bacteriana. Las bacterias también adquieren genes codificantes para la resistencia por transferencia directa de otros miembros de su propia especie o de bacterias de otra especie.

La adaptabilidad natural de los gérmenes se complementa con el uso masivo y a veces inadecuado de antibacterianos. Las condiciones ideales para la aparición de bacterias farmacorresistentes se dan cuando se prescriben antibióticos para resfriados y otras enfermedades comunes para las cuales no están indicados, o cuando los pacientes no completan el tratamiento prescrito. Los hospitales también proporcionan un ambiente fértil para las bacterias farmacorresistentes. El contacto cercano entre pacientes y el uso excesivo de antibacterianos potencian en el patógeno el desarrollo de la resistencia.

Los científicos y los profesionales de la salud coinciden en que para disminuir la incidencia de la resistencia antibacteriana será necesario mejorar los sistemas de monitorización de los brotes de infecciones farmacorresistentes y llevar a cabo un uso más racional de los antibióticos. También reconocen el papel crítico que juega la investigación para responder a este problema. Por ejemplo, los estudios sobre fisiología bacteriana ayudan a los científicos a entender los procesos biológicos que utilizan las bacterias patógenos para resistir la acción de ciertos fármacos. Este conocimiento puede conducir al desarrollo de nuevas estrategias para superar o para impedir estos procesos.

La investigación en genética molecular y en bioquímica permite identificar las rutas y funciones claves de la multiplicación de las bacterias. Los rápidos adelantos de la tecnología en la secuenciación de genes están haciendo más rápida y fácil la localización de las moléculas implicadas en estas rutas, las cuales podrían a su vez convertirse en dianas de nuevos fármacos antibacterianos. La investigación básica de este tipo ha generado ya resultados prácticos. Por ejemplo, los estudios sobre las bases moleculares de la resistencia a los fármacos en parásitos han conducido a:

* desarrollar las herramientas moleculares para identificar parásitos farmacorresistentes
* identificar la base genética de la resistencia y las alteraciones bioquímicas resultantes en varias especies de parásitos;
* identificar los métodos para revertir la resistencia; y
* sintetizar fármacos eficaces contra cepas de malaria farmacorresistentes.

B. PREVIOUS BACKGROUND KNOWLEDGE TEST.

*A continuación se presentan una serie de afirmaciones sobre bacterias. Indica con una **F** las que consideres falsas, con una **V** las verdaderas y con un **NS** en las que no sepas que contestar.*

1. Las bacterias son organismos pluricelulares.
2. Todas las bacterias son patógenas.
3. Existe la posibilidad de que la resistencia de las bacterias a los antibióticos sea trasmitida a los humanos a través de alimentos de origen animal.
4. El mal uso de antibióticos en animales y humanos favorece la aparición de la resistencia de las bacterias a estos medicamentos.
5. El ADN extracromosómico no es material genético.
6. La resistencia a los antibióticos se puede trasmitir de animal-a-persona, pero no de persona-a-persona.
7. Bacterias no patógenas pueden llegar a ser patógenas.
8. La trasmisión de material genético entre bacterias ocurre sólo entre un ser vivo y sus descendientes.
9. Los antibióticos son fármacos que destruyen o matan bacterias.
10. Las mutaciones pueden ser inducidas por agentes externos pero no generarse espontáneamente.
11. El grado de capacidad mutante que poseen las bacterias varía de unas a otras.
12. Una mutación consiste en un cambio en el material genético de las células de un ser vivo.
13. Una cepa de bacterias es un conjunto de bacterias descendientes todas ellas de una misma bacteria..
14. Los seres vivos tienen características genéticas que les hacen inmunes al ataque de determinadas bacterias.
15. Los genes son compuestos de ADN.

C. TRUE/FALSE LEARNING QUESTIONNAIRE.

Indica si las siguientes afirmaciones son verdaderas (V) o falsas (F).

1) La alta capacidad de las bacterias para adaptarse a nuevos ambientes es un factor clave para el desarrollo de la resistencia a los antibióticos *y para disminuir la tasa de mutación de las bacterias.*

2) Las bacterias transfieren su resistencia a otras generaciones y también a otros grupos de bacterias de las que no son descendientes.

3) Cuando las bacterias son atacadas por los antibióticos, pueden hacerse resistentes a ellos *porque se acostumbran a evitar sus efectos perjudiciales*.

4) Los plásmidos son piezas de ADN *de los cromosomas*

5) Los plásmidos codifican mecanismos de resistencia a los antibióticos

6) Las bacterias se convierten en resistentes cuando cambian o adquieren material genético que hace que los antibióticos no les dañen

7) Es necesario que ocurran *varias mutaciones* en el material genético de las bacterias para que se conviertan en resistentes a un antibiótico determinado

8) Cuando la cantidad del antibiótico que se toma es pequeña se facilita la aparición de bacterias resistentes porque habrá bacterias no-resistentes que sobrevivan al antibiótico, con lo cual disponen de más tiempo para mutar y convertirse en resistentes

9) El rápido incremento de bacterias resistentes se explica porque las no-resistentes son eliminadas por los antibióticos, disponiendo las resistentes de más recursos para reproducirse y multiplicarse rápidamente, y trasmitiendo la resistencia a sus descendientes

10) Las bacterias transfieren la resistencia a otras generaciones de bacterias procedentes de ellas mediante *plásmidos* mientras que transfieren la resistencia a otras bacterias con las que no están filogenéticamente relacionadas mediante *mutación.*

11) Bacterias no patógenas pueden llegar a ser patógenas y resistentes a los antibióticos.

12) Un factor que contribuye a *frenar* la transmisión de la resistencia es que las bacterias se reproducen muy rápidamente.

13) Para que las bacterias desarrollen mecanismos de resistencia a varios antibióticos es *necesario que experimenten varios cambios genéticos*, pero no uno solo

14) *Cualquier* cambio genético en las bacterias produce un aumento de su resistencia a los antibióticos

15) Si las bacterias no se multiplicaran rápidamente ni tuvieran gran capacidad de mutación, la aparición de la resistencia sería un problema de salud mucho menor

16) Las bacterias producen siempre resistencia sólo a antibióticos específicos.

17) Cuando una bacteria se vuelve resistente, adquiere ciertas ventajas frente a otras bacterias que no lo son.

18) Las bacterias emplean diversos mecanismos para combatir a los antibióticos, como por ejemplo impedirles la entrada al interior celular.

D. APPLICATION&TRANSFER PRACTICAL CASE

Imagina que has enfermado de amigdalitis y acudes al médico para que te prescriba un tratamiento. Este te informa que el agente causante de tu enfermedad es bacteriano y que, por ello, habrás de tomar un antibiótico durante una semana.

Comienzas con tu tratamiento, siguiendo las indicaciones de tu médico. Pero a los tres días te encuentras ya perfectamente y decides dejarlo.

a- ¿Crees que el antibiótico ha logrado erradicar a todas las bacterias causantes de tu enfermedad, a pesar de que te encuentres perfectamente? ¿Por qué?

b- ¿Cómo ha podido suceder?

c- ¿Qué implicaciones puede tener que hayas dejado el antibiótico de cara a combatir esa enfermedad?