FACULTAD DE PSICOLOGÍA

Departamento de Psicología Evolutiva y de la Educación

EFFECTS OF TIMING AND LEVEL OF INTEGRATION OF FORMATIVE FEEDBACK IN QUESTION-ANSWERING FOR CONCEPT LEARNING

TESIS DOCTORAL CON MENCIÓN INTERNACIONAL
PRESENTADA POR:

Carmen Guillermina Candel Sánchez

DIRIGIDA POR:
Dr. Eduardo Vidal-Abarca
Dr. Raquel Cerdán Otero

PROGRAMA DE DOCTORADO EN LECTURA Y COMPRENSIÓN

Valencia, 2019

The present thesis has been supported by the Ministry of Economy and Competitiveness, with the reference number BES-2012-053686.

To my parents Viriato and Carmen, my sister Cristina, academic mentors Raquel and Eduardo, Phd. colleagues Marian and Nacho, and those who were happy to help me learn; Brendan, Amelia, Laura, Alba, Arantxa, Ramiro, Tomás, Sam, the evaluation board and each one of the teachers and students who participated in these studies and made this dream possible.

INDEX

INTRODUCTION

Imagine a learning situation where a group of students are required to answer questions with a scientific text available on a computer. After answering, the system provides them with an elaborative feedback message at different timetables; *a)* Immediately (i.e., *Unfortunately, your answer is not correct; Atmospheric pressure can be measured with a mercury barometer*), *b)* in the form of a Delayed but Summative set of individual messages to each of the questions, or *c),* as a Delayed but unique Integrative message provided after the completion of four learning questions. The integrative elaborative message contains an elaboration of the relevant ideas that the teacher considers as crucial to construct a coherent mental representation from the text, i.e., *Unfortunately, your answer is not correct*; *Atmospheric pressure can be measured with a mercury barometer (hence the commonly used synonym is barometric pressure), which indicates the height of a column of mercury that exactly balances the weight of the column of atmosphere over the barometer,* (example extracted from britannica.com).

The example mentioned above is one of the most effective tasks educators assign students to help them acquire new content knowledge (Kapp et al., 2015). In this type of setting, students are often required to answer a set of multiple-choice questions repeatedly one by one, or as a set of cumulative tasks, in some cases, right after the study of some sort of text (Felker & Dapra, 1975; Hamaker, 1986; Roediger & Karpike, 2015). As we have seen, Formative Feedback (FF) refers to the quality of the students´ responses and can contain different components; as information regarding the correctness or incorrectness of students´ answers or any other relation of ideas, prompts or content elaborations (Golke, Dörfler & Artelt, 2015; Shute, 2008). Although Formative Feedback (FF) has been widely tested during question-answering; only a few studies on feedback effectiveness report interesting findings on students´ comprehension and learning from an available text in these sorts of settings. Up-to-date research on feedback instruction does not yield clear conclusions in regards to what type of timing and level of integration of the elaborative message of Formative Feedback is really effective for the teaching of complex declarative knowledge when students answer learning questions with an available text. This is the main goal of the present work.

On the one hand, research states that some forms of delayed Elaborative Feedback (EF) are effective means for the teaching of complex concept learning (i.e., inference-making). On the other hand, simpler forms of immediate feedback, like verification or correction (KCR, KR), still are considered to work better for memorization tasks such as the recall of words (Van der Kleij, et al. 2015). However, delayed forms of elaborative feedback can include both a summative set of messages delivered individually to each of the questions or, as proposed in the present work, a unique but delayed integrative elaboration provided after the completion of a specific group of questions. Moreover, compared to immediate feedback aiming at helping students correct errors (Llorens et al., 2015; Máñez, Vidal-Abarca, Martínez & Kendeu, 2017), a delayed integrative elaborative message might help students focus on more content-related aspects of the learning material; as in the form of a new re-study opportunity.

Undoubtedly, receiving Formative Feedback (FF) during question-answering is a multidimensional but complex learning process where students are considered to be active constructors of knowledge (Narciss, 2013). In fact, students must deploy a high level of cognitive and self-regulation strategies (Bangert-Downs et al., 1991; Butler & Winne, 1995; Körndle et al., (2004); Kapp & Körndle (2011); Kapp et al., 2015; Narciss, 2013; Vidal-Abarca et al., 2017). For example, planning a course of action to respond correctly to questions while being able to evaluate, reflect and judge their own certitude of correctness (Bangert-Downs et al., 1991; Vidal-Abarca et al., 2017). Thus, students need to process the feedback while being able to adjust to its corrective function according to both, the learning goals (i.e., understanding complex content from text or correcting errors) and the learning task (low or high-level) (Vidal-Abarca et al., 2017). In fact, certain forms of tasks can require different sorts of cognitive processes. While low-level learning questions require the simple location of a term or name in a text, high-level questions demand a more complex type of inferential activity as the interconnection of ideas across paragraphs (Cerdán et al., 2009; Vidal-Abarca et al., 2017).

However, the increment in the number of students per class, the great differences in previous background knowledge and the complexity of the learning material make it difficult for educators to deliver Formative Feedback (FF) that adjusts to student needs. In fact, several studies within the context of reading comprehension and learning have already reported that secondary school students often experience difficulties constructing deep level understandings of content and monitoring their own learning (i.e., Cerdán, Gilabert & Vidal-Abarca, 2011; Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, 2009; Mañá et al., 2009; McGaw & Grotelueschen, 1972; McNamara, Jacovina & Varner, *in press*).

In an attempt to address these issues, past research has focused on the development of more advanced Computer-Based Systems (CBS). In this regard, Computer-Based Formative Feedback and question-answering procedures are considered to be highly effective for the teaching of reading monitoring skills and the learning of verbal material (Golke, Dorfler & Artelt, 2015; Kapp et al., 2015; Mason & Brunning, 2001). First, CBS allow educators to provide students with different shorts of interactive features like going back and forth between questions (Kapp et al., 2015; Proske, Körndle & Narciss, 2012; Vidal-Abarca et al., 2017). Second, students´ responses and the online question-answering activity can be effectively recorded (i.e., reading times); Third, CBS may be adaptable to the students' responses by providing individualized content messages that can be delivered at specific points throughout the learning process. For example, after a mistake has been committed or right after the completion of four questions (Narciss et al., 2014; Kapp et al., 2015). In light of this, the present work focuses on the latest theoretical models and synthesis in feedback instruction and learning from the text, and proposes to investigate the role of two important variables in feedback effectiveness; the timing of Formative Feedback (FF) and the level of integration of the elaborative feedback message.

Following several studies analyzing how Formative Feedback (FF) mediates on reading performance and the correction of procedural errors in task-oriented reading when accounting the students´ reading skills and level of prior knowledge (e.g., Llorens, Vidal-Abarca, & Cerdán, 2016; Llorens, Cerdán, & Vidal-Abarca, 2014; Máñez, Vidal-Abarca & Martínez, 2016; Máñez, Vidal-Abarca, Martínez & Kendeou, 2017; Martínez, García & Vidal-Abarca, 2018; Vidal-Abarca, Martínez, Ferrer & García, 2017), *Study 1* of the present thesis analyzes how the timing of Formative Feedback (FF) (i.e., immediate versus delayed) affects students´ question-answering performance, the probability of correcting errors, the accuracy of the students´ certitude in their responses and finally, concept learning. Students answered learning questions while a text was fully available after a mistake had been committed and received Immediate or Delayed Feedback every four questions. A control group was prompted for re-reading.

Further, *Study 2* of the present work provides further empirical evidence regarding how timing of Formative Feedback (FF) and the level of integration of the elaborative feedback message affect concept learning during question-answering performance and a posttest. Additional attention is given to the moderation effect of previous background knowledge on performance the potential effects of the experimental manipulation on performance and learning outcomes of those students with high and low levels of previous background knowledge. Within the context of Sciences in secondary school education, in this study, an immediate condition provides FF in the form of verification or error-correction statement and an elaborative message on the correct answer. A second condition provides a delayed summative set of FF messages every four questions. And a third condition delivers a unique delayed integrative elaborative message containing a verification or error-correction statement, plus a unique and a well-structured relation of ideas from the text in the form of a re-study opportunity.

Before presenting both studies, this thesis delves into a general framework on the effects of Formative Feedback (FF) in question-answering settings with an available text on reading performance and learning. As a result, a first chapter of the present thesis displays a well-documented theoretical framework on the definition, types and effectiveness of Computer-Based Formative Feedback procedures in question-answering with an available text (Subsections one and two). Subsection three of the first chapter deepens on the theoretical models and synthesis in the literature accounting for the relevant cognitive and metacognitive effects that are characteristic in this type of instructional settings. First, *The Five-Stage Question-Answering Model* by Bangert-Drowns, et al., (1991) is described.

This model assumes question-answering is compiled of a set of consecutive steps and takes into account the metacognitive role of the learner in the monitoring of the own performance (i.e., the certitude of response). Next, we review the *Multidimensional Interactive Feedback Model by Narciss (2013)* which provides a framework that considers the interaction between Formative Feedback (FF) and the learner as a dual, but interactive loop.

Next, we review the *Model for Question-Answering in Web-LES* by Körndle et al.

(2004), Kapp and Körndle (2011) and Kapp et al. (2015) which offers an analysis of those the Intra and Inter task features characteristic of the question-answering process. The last synthesis by Vidal-Abarca and colleagues (2017) is reviewed in an attempt to fully disclose the most relevant steps and decisions students undertake when answering questions with an available text and receive Formative Feedback (FF). Vidal-Abarca et al., (2017) highlight the importance of the cognitive processing of Elaborative Feedback (FF), the students´ variables (i.e., judgments of learning) and the type of task (low versus high-level) in comprehension and learning from text.

The second subchapter of this dissertation presents evidence in relation to the most important literature in relation to the effect of the timing of the feedback in question-answering for the learning of concepts (*Study 1*). Additionally, we review the recent findings in regards to what type of content of Formative Feedback (FF) is more effective for concept learning. The thesis elaborates on one of the most relevant variables in the present work (*Study 2*), *the level of integration of the feedback message*, which is addressed as the type of components a message may contain (i.e., verification, error-correction or elaborative).

A third subchapter reviews the relevant literature in regard to those individual variables that are explored in the present work (*Study 1* and *2*) and that have been extensively addressed within instructional research (i.e., response certitude and previous-background knowledge). Finally, a fourth subchapter presents the general and specific objectives of the present thesis in regards to the development of the two experimental studies (*Study 1* and *Study 2*) that are compiled in the present work. In chapter two, *Study 1* analyzes the impact of a question-answering procedure that provides Formative Feedback (FF) in an immediate (question by question) or delayed manner (after the completion of four questions) compared to a non-feedback condition where students are induced to reread an available text. Dependent measures include the student´s probability of correcting errors, response certitude estimates, question-answering online processing and performance on a sample of undergraduate university students.

Chapter three presents *Study 2* and deepens the more conceptual aspects of the feedback message, like the level of integration of the elaborative content. *Study 2* specifically explores how *the level of integration of the elaborative feedback message* (as a verification or error-correction statement plus further elaborations on the correct answer) can impact question-answering performance and concept learning when provided in an immediate, delayed summative or delayed integrative manner (i.e., right after the completion of four questions). The first group of students received Knowledge of Correct Response feedback (verification and error-correction, KCR) on their answer plus an elaboration of the correct response question by question. The second group of students received delayed summative feedback that contained KCR and an individual elaboration of the correct answer to each question delivered right after the completion of four questions in a row. A third condition received a delayed integrative elaborative feedback message that contained KCR and a relation of the relevant learning ideas from the text. Its main objective was to help students create a well-structured mental model from the text (Kintsch, 1998). Dependent measures in *Study 2* are the question-answering performance, online behaviour and concept learning.

Additionally, in this study, the moderation effects of Previous Background Knowledge (PBK) and the effects of task type (i.e., low versus high-level questions) based on individual differences on the level of previous background knowledge of students (low and high) are further explored and discussed. Finally, chapter four dives into the general discussion, limitations, educational and research implications based on the results of both empirical studies and the literature reviewed.

CHAPTER ONE

THEORETICAL FRAMEWORK

1. Formative Feedback in Question-Answering and Performance.

1.1. Formative Feedback for Concept Learning: Definition and Types.

Formative Feedback (FF) is defined as any informative message provided by a peer, educator or computer-based system regarding the specific aspects of the students´ performance (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mathan & Koediger, 2005; Narciss, 2014; Shute, 2008). For example, Knowledge of Results (KR) and Knowledge of Correct Responding (KCR) aim at reinforcing the right responses or correcting the incorrect ones (Shute, 2008). Nevertheless, when question-answering implies the learning of complex knowledge, Formative Feedback (FF) can contain specific elaborations aimed at enhancing the acquisition of procedural (i.e., *how to read a text more efficiently*) or declarative knowledge (i.e., *specific knowledge about the developing stage of storms*).

The most effective teachers are well aware of the importance of providing students with meaningful explanations on the relevant information (Rosenshine, 2012). For example, when question-answering implies an available text, providing formative feedback addressing the students´ incorrect responses may be conducive to further text reading and possibly, to error-correction.

However, when students are required to comprehend more complex knowledge, i.e., *how specific atmospheric disturbances may impact snowstorms and related phenomena*, Elaborative Feedback (EF) message may be more useful. For example, an elaboration message clarifying the complex relations between conceptual ideas can be designed according to the type of task at hand (i.e., location of a name compared to relating complex knowledge), and might facilitate the acquisition of meaningful learning. In this line, Hattie and Timperley (2007) concluded that Formative Feedback (FF) can be intended at four key levels; the self, the task, the process or the regulation levels.

Feedback at the self-level refers to individual aspects (i.e., motivation), feedback at the task level highlights and corrects errors, feedback at the process level displays a set of how-to steps to complete the task successfully, and finally, feedback at the regulation level addresses the learners´ ability to monitor their own learning process.

In line with this conceptualization, several authors have classified feedback according to a certain number of components (Hattie & Timperley, 2007; Holding, 1965, Narciss, 2008; Smith & Ragan, 1993, Van der Kleij et al., 2015). For example, according to Narciss et al., (2008, 2014), Knowledge of the Correct Response (KCR) may or may not indicate an error, but quite normally informs about the correct answer either visually or by a short statement, i.e., *Unfortunately incorrect, the correct answer is X*. Knowledge of Response (KR) highlights the students´ incorrect responses by a visual clue without delivering the correct answer. Some subtypes of KR can also include Answer-Until-Correct (AUC), which may provide in some cases Knowledge of Result feedback and the opportunity to make a limited or unlimited number of attempts to answer (Corbett & Anderson, 1991). Additionally, Multiple-Try-Feedback (MTF) provides the opportunity to make a limited number of attempts after a mistake.

A second component corresponds to the elaborative element of a feedback message (EF) (Narciss, 2014; Shute, 2008). Elaborated Feedback (EF) may combine KR and KCR and include several kinds of elaborative information in regards to; *1)* task characteristics (i.e., rules or constraints) (KTC), *2)* concepts necessary for task processing (KC), *3)* errors in the form of Knowledge about Mistakes (KM), *4)* procedural knowledge relevant for task processing (Know-How Feedback, KH), and finally, *5)* metacognitive hints (KMC) (Narciss, 2008; 2014).

In addition to this, as we will review in chapter three of the present thesis, some contextual characteristics of Formative Feedback (FF) such as timing, can have a decisive impact on learning. As Narciss (2014) defines in her review, KR, KCR and EF can be delivered immediately compared to Knowledge of Performance (KP), which provides summative feedback after a certain number of tasks (e.g., the percentage of correct answers). An example of this type of message can easily be found and transferred to question-answering settings where formative feedback include elaborative content like inference-prompts in regards to why a selected response is either right or wrong (i.e., "*Incorrect. The correct answer is reinforcement. The child's crying increased in frequency, so it was reinforcement*", Jaehnig & Miller; 2007).

1. 2. Computer-Based Formative Feedback in Question-Answering with an Available Text.

Question-answering with an available text refers to those academic situations in which an external source (i.e., *teacher, peer or computed-based system*) requires students to complete a purposeful reading task (i.e., *answering a set of comprehension questions*) that implies managing one or more texts (Bråten, McCrudden, Stang, Brante & Strømsø, 2017; Gil et al., 2015; McCrudden, Magliano & Schraw, 2010; Vidal-Abarca et al., 2010). This type of interactive text-based reading is characterized by browsing and scanning a text while presenting questions that may differ in their nature or formats (i.e., multiple-choice versus open-ended). Students are required to self-regulate and monitor their own ability to interact with the relevant information from the text (Llorens et al, 2014; 2016). Due to its myriad of functions, Computer-Based Formative Feedback can support educators to deliver feedback messages in a more effective way (Golke at el., 2015).

Amongst the advantages that Computer-Based Systems (CBS) bring to question-answering, is the chance to provide individualized messages that combine different forms of content (i.e., procedural or elaborative) that can be delivered at specific points throughout the learning process (Narciss et al., 2014; van der Kleij, Eggen, Timmers, & Veldkamp, 2012). Computer-Based Systems (CBS) also allow educators to record online information like the students´ scores and self-monitoring processes, like the time students spend reading a text.

It is of importance to mention that research in reading comprehension and learning from the text considers learning with being associated to two main levels. First, high-level concept learning is considered to be the result of the effortful interconnection of the relevant ideas from a text into a coherent mental model. Low-level learning refers to the mere memorization of literal ideas, dates or names from a text (Cerdán et al., 2008; Cerdán et al., 2009; Rouet, 2006; Goldman & Durán, 1988; Kintsch 1988, 1998; van Dijk & Kintsch, 1983; Vidal-Abarca et al., 2010; Vidal-Abarca et al., 2017). In this light, recent research within the context of comprehension with an available text, provides interesting findings in regards to how Formative Feedback (FF) enhances the students´ processing and learning (Máñez, Vidal-Abarca & Martínez, 2019; Máñez, Vidal-Abarca; Martínez & Kendeou, 2017; Llorens et al., 2014; Llorens et al., 2015; Llorens et al., 2016; Vidal-Abarca et al., 2017).

For instance, Llorens, Cerdán and Vidal-Abarca (2014) used the Computer-Based software named *Read&Answer* to record the students´ actions and outcomes. In this study, these authors manipulated the content of Formative Feedback (FF) by modifying its components. In the first condition, students received Knowledge of Response feedback (KR) about the incorrect response Knowledge of the Correct Response (KCR). A second condition provided students with Knowledge of Response feedback (KR) and Knowledge of the Correct Response (KCR) plus an additional elaborative feedback message (EF) with specific indications on the desirable search strategies to answer correctly. This content specifically referred to those *what* (i.e., what information to search) and *when* (i.e., when is conveniently effective to search the text) decisions, as these monitoring processes have been strongly related to successful self-regulation and comprehension of the relevant content within a text by recent literature (Rouet, 2006; Vidal-Abarca, et al., 2014; Vidal-

Abarca, et al., 2011; Vidal-Abarca et al., 2010). In addition, the message also included a specific recommendation on how to proceed in subsequent questions. In fact, when students provided a right answer, but they had not followed the proper strategic decisions, the feedback message delivered advice (i.e., *In the following questions, think if you should search the text to ensure a correct answer*). In case the answer provided by the student was correct, the software provided a confirmatory message instead (i.e., *Right! During the search you have read information which is necessary to answer correctly. This way to answer has been effective. It will certainly help you to answer other questions*). The experimenters compared both feedback procedures to a third one which provided placebo feedback messages only in regard to the order of the questions (i.e., *You have answered question number 5*).

Results reported those students receiving KR+KCR feedback in regards to their *strategic search decisions* obtained higher performance scores, decided to reread the text to a higher extent and were able to extract more relevant information to answer compared to students receiving KR+KCR and *placebo feedback*.

In another study, Llorens, Vidal-Abarca & Cerdán (2016) designed a feedback procedure to specifically promote the transfer of strategic reading skills into a new reading situation. However, in this study experimenters provided students with an elaborative search & revisit feedback, which allowed students to search the text at their own will, or received select & revisit feedback which forced them to specifically revisit and select the relevant information to answer every question. In the second phase, students were required to read a different text and answer a set of learning questions and received no feedback. Results reported that both feedback groups revisited the questions more often than those students receiving neutral feedback. Additionally, students who received specific select & revisit feedback scored higher in performance and read more relevant information than those students in the neutral and search & revisit conditions.

In a similar but more recent study, Máñez, Vidal-Abarca & Martínez (2019) divided a sample of secondary school students among a control and formative feedback condition. Students read a text, responded to questions and provided a selection of the information they considered as relevant to answer. *Read&Answer* provided Knowledge of Response (KR) feedback on the correctness or incorrectness of their answer and the accuracy of their selection. Additionally, feedback students received a Hint (KH) about the proper selection strategy to follow in the subsequent questions.

Once students had closed this message, students received another Knowledge of Correct Response (KCR) feedback on the question and on their selection compared to an ideal selection that was visually depicted on the screen. Students had the option to access the text where the ideal selection appeared highlighted on the screen. Participants in the control group received placebo feedback messages informing only about the number of the question they just responded (i.e., *"You have answered question number 1"*). Authors found that students in the formative feedback group reported higher reading performance scores compared to the control group ($p = .027$) and included less non-relevant information in their answers.

Beyond the effectiveness of Formative Feedback (FF) on procedural learning (i.e., reading skills), the role of more content-wise aspects has been also addressed. In fact, Vidal-Abarca, Martínez, Ferrer & García, (2017) asked students to read the text at their will and answer low and high-level learning questions. After providing the first response to a question, the first group received Knowledge of Response (KR) feedback. The second group received either Knowledge of Response followed by an elaborative message explaining the correct answer (EFcorrect), or Knowledge of Response (KR) feedback followed by an elaborative message explaining why the selected option was incorrect (EFmistake). Students were able to search the text after a second attempt after which KR feedback was provided to all the conditions.

Students receiving only KR revisited and reread the text more often when answering the low-level learning questions compared to those students in the two Elaborative Feedback (EF) conditions. In addition to this, receiving an elaborative feedback message on the correct answer induced students to search the text a greater extent, especially after a mistake was committed. In addition, Vidal-Abarca et al., (2017) observed that receiving an elaborative feedback message on their mistake helped students perform better on the high-level questions.

Due to the increasing importance of the individual variables in the current theoretical models on feedback research (Bangert-Downs et al., 1991; Narciss, 2013) , in another study, Máñez, Vidal-Abarca, Martínez & Kendeou (2019) examined how text availability, prior background knowledge (PK) and corrective feedback (KR) affected the decision to access elaborated feedback (EF), while exploring how these variables had an indirect effect on self-regulation and question-answering performance. These authors found that the relationship between students´ prior knowledge and learning during training was significantly moderated by the number of times students accessed content elaborated feedback during a testing phase. Specifically, students with low levels of prior background knowledge, who decided to consult the Elaborative Feedback (EF) more often, benefited significantly more in comparison to the ones that consulted feedback to a lesser extent.

In summary, the studies previously mentioned so far clearly indicate how the content of Formative Feedback (FF) is a crucial factor to enhance reading performance and monitoring during question-answering with an available text (Llorens et al., 2016; Máñez, Vidal-Abarca & Martínez, 2019). Formative Feedback (FF) seems to play a key role not only in the way students assess the relevant information from the text, but takes a first step into how individual differences can be a strong precursor for procedural and content learning during question-answering. Individual differences are important factors in the modulation of feedback effectiveness (Smits, Boon, Sluijsmans, & Van Gog, 2008). The study by Vidal-Abarca et al., (2017) provides empirical findings on how the delivery of corrective feedback affects students reading online behaviour when answering different types of questions (i.e., low and high level) while having the text available. In fact, Vidal-Abarca et al., (2017) observed that receiving an elaborative feedback message on the students´ mistake helped them perform better on the high-level questions.

Next, the study by Máñez et al., (2019), shows how previous background knowledge (PK) moderates learning and provides interesting information in regards to the cognitive and metacognitive factors that underlie the processing of Formative Feedback for error-correction. From these studies we can conclude that an available text is also an important context variable for learning (Ozuru et al., 2007; Schroeder, 2011). In fact, it allows the student to interact on a cyclical basis with the learning material, the feedback messages and the learning questions. Nevertheless, the aforementioned studies do not address the extent to which a Formative Feedback that varies in its timing of presentation (i.e., immediate or delayed) and the level of integration of the feedback message might impact concept learning when students answer questions with an available text. First, delivering immediate feedback KR and KCR on a question per question basis can impel students to search the text with more emphasis, especially after making a mistake. This procedure could have a remarkable impact on the students´ ability to search the text. In fact, it could have a remarkable impact on the students´ certainty to respond correctly. In addition, receiving Formative Feedback (FF) in an immediate manner (i.e., question after question) compared to a delayed manner (i.e., every four questions) may be decisive to the way students process feedback content, read a text initially and monitor their own learning.

Moreover, this effect could be completely different when, instead of providing KR or KCR only, the system provides an Elaborative Feedback message. As we have reviewed, Elaborative Feedback (EF) can include more complex relations in regards to why certain responses are correct or incorrect (Butler, Godbole, & Marsh, 2013; Vidal-

Abarca, Martínez, Ferrer & García, 2017). As the present work intends to confirm, an integrative elaborative feedback message that provides students with a coherent mental model from the learning content might help students comprehend information from text to a larger extent. For this reason, the second objective of the present thesis is to further investigate how the level of the integration of the information contained in a feedback message (i.e., KR, KCR and EF) may have a crucial impact on the way students acquire concept knowledge during question-answering. The moderation effect of previous background knowledge and the Individual differences and the role of the type of question (low and high-level) are also taken into account and further discussed.

Therefore, in order to fully understand the cyclical and dynamic nature of Formative Feedback (FF) in question-answering settings, in the third section of the present chapter, we provide a general framework with the relevant literature on the existing theoretical models and synthesis in feedback processing that account for these cognitive and metacognitive effects. The second subchapter explores the relevant literature on the role of feedback timing and type of content for the teaching of declarative knowledge in question-answering settings. In subchapter three we review two of the most important variables that belong to the reader and that have reported to be strong modulators for feedback effectiveness and concept learning so far; *response certitude* and *previous background knowledge*. Subsequently, the general objectives of the present work will be summarized in subchapter four. *Study 1* (chapter two) 1 and *2* (chapter three), will be presented. In chapter four, a final general discussion of results and further implications for the future are fully displayed.

1.3. Cognitive and Metacognitive Factors of Formative Feedback in Question-Answering; Theoretical Models.

Formative Feedback (FF) procedures can be designed to induce the acquisition of procedural and conceptual knowledge in various ways; for example, while some procedures may advise students on selecting, correcting and elaborating on errors (Anderson, Conrad, & Corbett, 1989; Corbett & Anderson, 2001; Máñez, Vidal-Abarca & Martínez, 2019; Máñez, Vidal-Abarca; Martínez & Kendeou, 2019; McMartin-Miller, 2014; Llorens et al., 2014), some others may focus on providing specific guidance and scaffolding on their error-correction skills (i.e., *models of desired performance*) (Mathan & Koedinger, 2005; Narciss, 2006). This type of feedback procedures consider the importance of a verification statement, i.e., confirming correct answers or EF, however, they mainly focus on providing the metacognitive guidance students need to succeed through the practice of error-detection and correction. Advocates for this view state that a feedback message should focus the students´ attention on the correction of errors, instead of on the correct answers. In this line, these authors highlight the importance of providing students with the opportunity to reason about the causes and consequences of their errors (Moreno & Valdez, 2005; Narciss, 2008). However, beyond the verification and elaboration statements for the treatment of errors, Formative Feedback components can be combined in a wide range of forms and loops (i.e., number of tries), which are often referred to as *Informative Tutoring Strategies* (ITF) (Narciss, 2013). For example, a feedback procedure may present students with a second try after a mistake has been committed, or with an elaboration on why a response is either correct (Butler, Godbole and Marsh; 2013) or incorrect (Anderson, Conrad, & Corbett, 1989; Corbett & Anderson, 2001). Feedback can then be provided immediately after providing a response or with some delay (Kapp, Proske, Narciss & Körndle, 2015; Narciss, 2014).

These sorts of *Tutoring Feedback Strategies* are common procedures not only in the context of Mathematics or Physics, but also in Question-Answering settings. In fact, the delivery of Formative Feedback (FF) during question-answering brings experimenters more advanced interactive tools to test a wide range of feedback combinations, such as providing a multiple-try attempt to answer a question after providing corrective feedback. Consequently, the cognitive and metacognitive demands that the processing of Formative Feedback can require from students are complex and similar to those described by relevant literature in self-regulated learning (e.g., Azevedo & Cromley, 2004; Bangert et al., 1991; Butler & Winne, 1995; Narciss et al. 2014; Pintrich, 2000; Kapp et al., 2015; Vidal-Abarca et al., 2017).

Question-answering is a multidimensional learning activity, and the literature in feedback effectiveness embraces the application of Formative Feedback procedures as a series of components that can be distributed and combined according to; *a)* learning goals (i.e., strategic steps or knowledge building phases), *b)* feedback content (i.e., ranging from informative, corrective, reinforcing and elaborated), *c)* feedback distribution and *c)* type of task. Overall, the learner is considered as a constructor of knowledge and his individual characteristics play a decisive role in the moderation of the acquisition of knowledge (e.g., Bangert-Drowns, et al., 1991; Butler & Winne, 1995; Körndle et al., 2004, Kapp & Körndle, 2011; Kapp et al., 2015; Narciss, 2013).

According to a classical model by Bangert-Drowns, Kulik, Kulick and Morgan´ (1991), delivering Formative Feedback (FF) while testing induces students into a complex decision-making process. First, the learner´s background knowledge is the first mediator of this interaction. Second, searching and retrieval memory strategies are activated by the task characteristics and are strongly influenced by the learning goals. At a third and fourth stage, the learner is prompted to provide a response while is required to evaluate his performance through the monitoring of his own adjustment processes (i. e., *the degree of* *certainty of the response*).

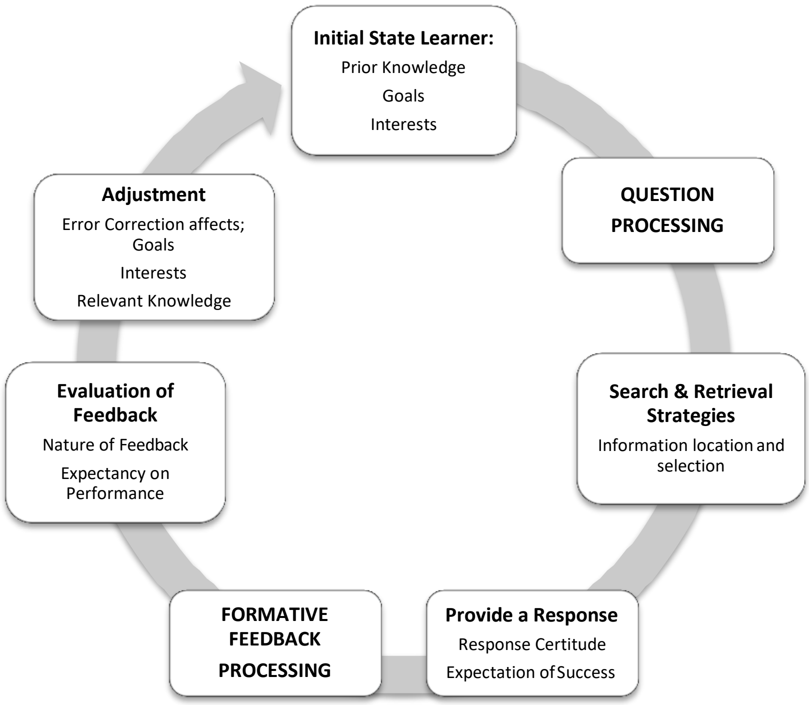


Figure 1. *The Five Stage Model by Bangert-Downs et al., (1991), from Dempsey, Driscoll & Swindell (1993). From Interactive Instruction and Feedback (p, 40), (Eds) Dempsey and Sales.*

As Bangert-Drowns, et al., (1991) state, the processing of a feedback message leads students into the consideration of those error-correction strategies that eventually will influence the acquisition of procedural or conceptual knowledge in a decisive way. This model provides interesting insight into the interaction between a feedback message and the learner. However, it does not provide further clarifications about how the type of task and the internal validation of the learner affect specific self-regulation processes (i.e., as the decision to search a text and what information is relevant to answer), question-answering performance and task processing.

In a later Model, Butler and Winne (1995) proposed a frame for the analysis of the effects of Formative Feedback (FF) that accounted for the self-regulation process more in depth. The Model assumes that learners generate an internal evaluation of their own performance, learning goals and their own strategies applied during the learning activity. An external feedback message provides the necessary information the learner needs to proceed with the learning process, and assumes that learners are able to monitor their own task engagement to a certain degree. The learner needs to adjust to the feedback requirements, when needed. Feedback content may agree with or disregard the internal evaluation of the learner. In case of disagreement, the feedback creates a discrepancy between the learners´ knowledge and beliefs. An efficient Formative Feedback (FF) message must be the one that provides students with new learning strategies, re-establishes motivation, enhances knowledge and supports the self-regulation process of the learner at every step during the learning process.

Butler and Winne (1995) provide a very interesting perspective on how the learner evaluates and adjusts its own performance. Nevertheless, the inclusion of Computer-Based Systems (CBS) has allowed experimenters to analyze and describe question-answering processes in a more interactive manner (Proske, Körndle & Narciss, 2012). For example, Körndle et al. (2004), Kapp and Körndle (2011) and Kapp et al., (2015) proposed a Model that describes the Q-A cycle as a set of several stages that directly relate feedback effectiveness to the cognitive and metacognitive process that pertain to the learner. According to this Model, in the *orientation stage* students activate prior-knowledge and analyze the digital environment in order to set the operative conditions under which students will operate. In the *planning stage* students select the relevant study strategies they will follow according to the demands of the learning context. During the *processing stage,* learners interact with instructional questions which provide them with the clues they need to monitor the searching, selecting and processing of information.

Next, a Formative Feedback message that provides students with information regarding their performance helps students acknowledge and adjust their learning process. For this, a two trial feedback algorithm can guide students through the necessary cognitive and metacognitive processes that are needed to correct errors (i.e., a new search may be considered when failing at a particular question). So students need to devote their cognitive and metacognitive resources to the construction of knowledge and the monitoring of memory resources to answer accordingly (i.e., read the text or a specific feedback message more carefully). During the *evaluating stage*, students are ready to use feedback clues to assess their progress and to decide whether or not a specific reading strategy fulfills task requirements.

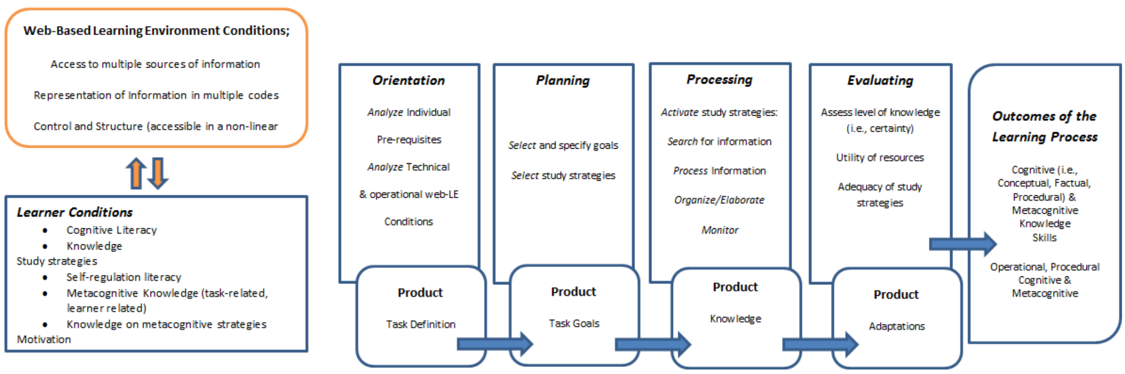


Figure 2. *Question-answering in web-LES; Intra and Inter Task Features Model. Extracted from Kapp, et al., (2015). Adapted from Self-regulated learning in web-LES. Learners´ conditions, web-LE characteristics and study tactics, based on Narciss (2013) and Winne and Hadwin (1998).*

As the Model extracted from Kapp, et al., (2015) describes in the picture, the cognitive and metacognitive processes involved in question-answering in web-LES environments are strongly dependent on the learner conditions, which at the same time, set up the initial stage of the learning process. This model takes into account the aforementioned internal variables of learners, but it does not provide a full description of the interaction between the *tutoring feedback procedure* and the internal control system of the learner. In order to do so, The ITF-Model by Narciss (2013) gathers some other important aspects like self-regulation (i.e., Butler & Wine, 1995), systems theory (e.g., Ramaprasad, 1983), tutoring procedures, formative assessment, error analyses and task analyses theories (*for further reference see* Narciss, 2013).

In this vein, the *Interactive Tutoring Feedback Model* by Narciss (2013), takes a stand on the matter and considers the interaction between Formative Feedback (FF) and the learner as a multidimensional loop which takes part within the context of an interactive learning setting (i.e., electronic environment). Feedback effects are determined by the quality of the instructional activity in terms of its nature and the situational conditions of the instructional setting (e.g., instructional goals, learning content and tasks). First, the Model also considers individual characteristics to a larger extent compared to the other Models (e.g. previous knowledge, certainty of response, self-regulation skills, literacy skills, meta-cognitive strategies, motivational dispositions and strategies).

The ITF-framework assumes that, in order for feedback to induce learning a *standard or reference level* has to be determined. This is displayed by the learner´s *internal standard or reference* described as subjective representation of competencies and those related standards a learner has at the moment of learning (i.e., previous background knowledge). Similarly, an *external standard or reference* is determined by the *external representation of competencies* that construct a feedback message. The two external and internal standards are influenced by the situational features of the instructional context (i.e., learning goals, standards of knowledge, nature of the learning tasks and feedback distribution).

Second, an efficient tutoring feedback strategy is the one that continuously assesses the actual state of competencies by both the learner (*internal assessment)* and the external feedback source (*external assessment*). Third, a controlling information processor *or controller* regulates the process of competence acquisition for both loops. As a result, the *external controller* generates an *external feedback message* that may confirm or disregard the learner response. If a gap between the current level of competencies and the standard of reference is detected, the external feedback message provides some evaluative information (i.e., a work-out example, explanations or content elaborations, amongst others).

Fourth, the learner´s *internal controller* is required to process the external feedback message. In fact, a learner needs to compare, *a)* his desired level internal current state of competencies (internal feedback), *b)* his desired level of competencies with the external feedback message and, *c)* his internal feedback with the external feedback message. Based on the results of these comparisons, learners generate their *internal control actions*, such as the decision to read a text or correct an error. This process may have a great impact on the cognitive, metacognitive and volitional resources of the learners. Finally, in a fifth step, learners have to select and apply the *internal control actions* and perform the selected strategies to achieve the desired level of competences (e.g., error correction strategies; revision of information). When learners are able to select the adequate control action(s), their level of competence improves. This improvement should result in the reduction of the gap between the desired and the actual state of competency.

As such, the ITF-Model by Narciss (2013) emphasizes the role of the internal controlling processes of the learners as well as the importance of the learners´ accuracy when monitoring their own learning. However, the Model does not account for any other cognitive and metacognitive effects which are characteristic of question- answering and learning with available text, such as what information is relevant to answer and when to refer back to the text.

In this vein, Vidal-Abarca et al., (2017) developed a theoretical synthesis that fully describes how Formative Feedback and question-answering tasks require students to interact with an available text, process feedback messages and adjust their reading behavior to meet task requirements. First, these authors highlight three factors that are key in the formation of a *task model* (Rouet, 2006). When students answer questions, they engage in a set of ***QTW** cycles.* Cycles are question-answering sequences that imply a first reading of a text and the **Q**uestion, the decision to searching the **T**ext and the final action to

Write or provide an answer (QTW cycles). All of this takes part while students receive formative feedback in their performance. However, as authors claim in their approach, different sorts of learning tasks may also impact the way students process feedback and learn from a text. For example, high-level processing (Kintsch, 1998) is often induced by questions that require complex processing (i.e., making inferences). In fact, deep learning results from the integration of the relevant information with the students´ prior knowledge in a coherent and constructive manner. Processing at a low-level implies associative learning (i.e., the location of verbatim).

The second key factor in Vidal-Abarca et al., (2017) synthesis refers to the self-regulatory Q-A mechanisms students engage in during task processing, i.e., searching decisions. Based on Rouet TRACE Model (2006), these authors take into account a set of cyclical steps that are crucial in the interaction between the learner, the questions and the text. Firstly, the student needs to create an *initial task model* when reading the task (e.g., a learning question) (1a). To this, the student needs to integrate and interpret the propositions included in the question. After a first evaluation, the student activates those monitoring processes that are needed to decide whether or not he needs to search the text (1b) to provide a response. In case the student decides not to search, a response is created (1c) and provided. If the student decides to keep on searching the text, a new QTW cycle takes place (2a) and the student needs to scan the text (2b) in a new attempt. At this stage, authors mention the importance of the *Judgments of Learning*, or JOLs, as an effective manner to gather information on the metacognitive assessment students perform during the learning process.

These sequential cycles also allow the student to update his task model (2c). As Vidal-Abarca et al., (2017) highlight, a third stage of this process occurs when external feedback is provided. In this regard, Formative Feedback acts an external evaluation source that invites students to update their responses given (3a), according to feedback characteristics (i.e., content).

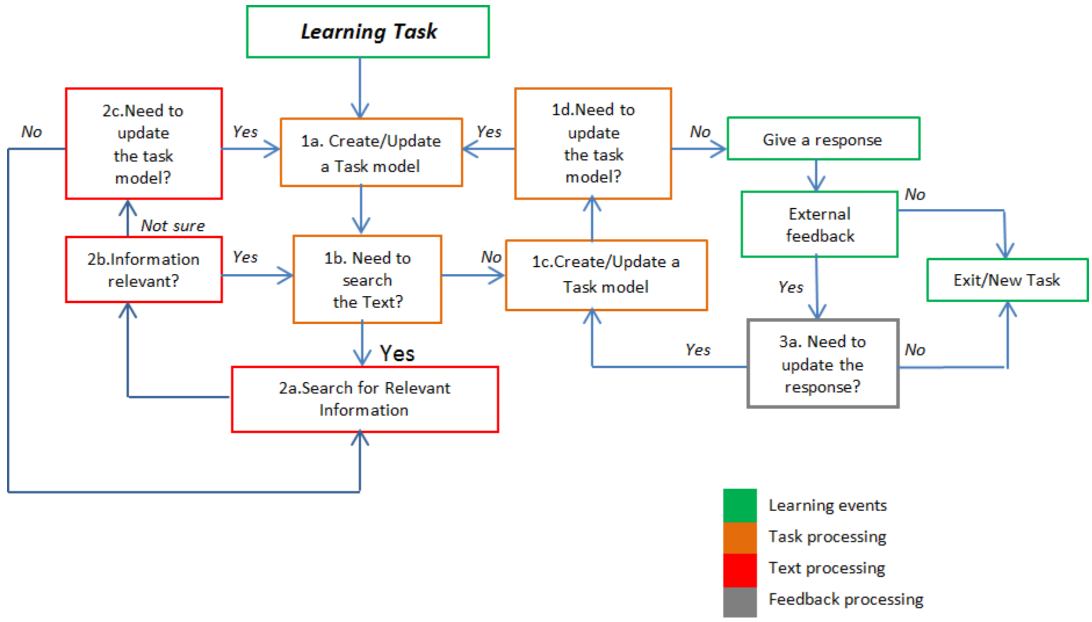


Figure 3. *Question-answering task with an available text and feedback processing. Extracted from Vidal-Abarca et al, et al., (2017).*

Nevertheless, beyond its corrective or reinforcing function, Vidal-Abarca et al., (2017) clearly conclude that a beneficial feedback message needs to contain the right information in order to be adequately processed. According to Vidal-Abarca et al., (2017), Elaborated forms of Feedback (EF) (i.e., additional explanations, worked-out examples) can easily be provided with KR or KCR effectively to help students gain knowledge from text. As such, once a student engages in feedback processing, its content is expected to challenge his cognitive resources by inducing him to compare his actual performance with a standard level of performance (3a). When formative feedback provides the relevant information a student needs, he can update a response model according to the task requirements (3c).

In light of the literature reviewed in this chapter, it is of interest for us to analyze to what extent the delivery of formative feedback may contribute to students´ concept learning from a text, their reading performance and the students´ monitoring (i.e., predictions of certainty) when they read a text in Sciences and answer questions. However, models about feedback effects are still quite general, especially with regards to how certain contextual features (i.e., feedback timing) and level of integration of the elaborative feedback message might affect concept learning from text. Additionally, those individual variables of the students (i.e., previous background knowledge or certitude of response) may also play an important role in these types of settings.

It seems clear that the content of the feedback message plays a relevant role for both, concept learning (Butler, et al., 2013; Vidal-Abarca et al., 2017) and the self-regulation of the learning process (Kapp et al., 2015). However, there is still a disparity of conclusions in regards to what timing (i.e., immediate versus delayed) and what level of integration of the feedback message (i.e., ranging from the more simple forms of verification and corrections components such as KR and KCR, to more integrative elaborative feedback messages) could benefit students the most at both, the cognitive (i.e., performance and learning outcomes) and metacognitive level (i.e., reading times) (Butler & Roediger, 2008; Butler, Karpicke, & Roediger, 2007; Clariana & Koul, 2006; Golke et al, 2015; Newman, William & Hiller, 1974; Shute, 2008., Shuber & Anderson, 1975; Van Der Kleij, 2011; 2015). The present work attempts to clarify these aspects. To this, in subchapter two and three we will analyze more in depth the role of feedback timing and its content for the teaching of concept knowledge during test-like events.

2. The Role of Timing and Content of Formative Feedback For Concept Learning.

2.1. Effects of Timing of Formative Feedback on Concept Learning.

According to Skinner (1954; 1957), feedback is the immediate post-response provided to increase the likelihood of a particular response or behavior to be performed again under the same circumstances. However, feedback can be distributed under different timing schedules; immediately item by item, or in a delayed form (i.e., at the end of the completion of a set of items, in seconds, minutes or even days) (Butler & Roediger, 2008; Metcalfe, Kornell & Finn, 2009; Narciss, 2014; Smith et al 2008; Kapp et al., 2015; Kulik & Kulik, 1988).

However, conclusions on what time of feedback is best for concept learning are mixed. On the one hand, defendants of immediate feedback and academic testing claim that providing corrective feedback right after mistakes enhances error correction and recall (i.e., word pairing, arithmetic facts, or trivia facts) (Brosvic, Dihoff et al., 2006; Brosvic, Epstein et al., 2006; Dihoff et al., 2003; Dihoff, Brosvic & Epstein, 2004). For example, within the context of problem-solving, Corbett and Anderson (2001) found that students who acknowledged and corrected their errors immediately reported more efficient problem solving strategies than those students who applied their own criteria or requested feedback at their will.

On the other hand, those who advocate for the efficiency of a delayed feedback report that it is the most adequate approach to correct errors and to maintain the correct answers over time (Brackbill, Bravos, & Starr, 1962; Brackbill & Kappy, 1962; Kang, McDermott & Roediger, 2007; Sassenrath & Yonge 1969). The superiority of a delayed feedback may be in part due to two compatible mechanisms. First, providing a delayed presentation of Formative Feedback (FF) induces incorrect responses in memory to dissipate, while allowing students to retrieve the correct ones more easily; an effect that has been classically defined as *The Interference Perseveration Hypothesis* (Kulhavy & Anderson, 1972). This Hypothesis states that the initial learned responses interfere with the students´ ability to recall the correct responses. According to Smith (2007) an alternative explanation for this effect was already given by Bjork & Bjork (1992, 2006) in their *Theory of Disuse.* In his work, Smith (2007) concludes that when learning is understood under a memorization structure of retention (i.e., recall of words), the quality and quantity of the information storage in memory strongly depends on the strength of the connections between those nodes (i.e., a particular node and its associated nodes).

A second explanation refers to the spacing effects of the delayed distribution of the learning material. In this regard, spaced study episodes that imply testing often lead to robust verbal learning compared to massed study episodes (Carpenter & DeLosh, 2005; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Cepeda et al., 2009; Dempster, 1988; 1986; Janiszewski, Noel, & Swayer, 2003; Pyle, 1913). This theory states that spacing testing episodes act as a new restudy event and allow students to review and reinforce the correct answers more efficiently.

In addition to this, Hattie and Timperley (2007) suggested that the type of task is also a relevant factor for feedback efficiency. While immediate feedback works best for easy or low-level items, setting a delay benefits learning of difficult or high-level items. Shute (2008) argues that when feedback is intended to facilitate low-order learning such as the recall and retention of verbal items (i.e., words and names), immediate feedback is efficient; however, when students face high-order learning tasks, they are more likely to benefit from a delayed feedback (Van der Kleij, et al., 2015). Up to date the vast majority of research on feedback timing is still applying low-level learning tasks such as multiple-choice factual tests, vocabulary word-pairing tasks or repeated testing (Mory, 2004; Sinha & Lewis, 2015).

In this line, Dihoff, Brosvic & Epstein (2004) analyzed the effects of different types of delays on the retention of verbal material in a set of low-level retention tasks which implied the mere association of items. Students completed five quizzes using different response formats as follows: no feedback, immediate feedback item by item, delayed feedback at the end of the test, delayed feedback after 24 hours. Twelve weeks later, students were required to complete a posttest test consisting of fifty items that included ten repeated items. Results showed that students in the immediate condition reported a more accurate identification of initial responses and a higher level of response confidence estimates. Students in the delayed condition reported less recall, less identification accuracy of initial responses, lower confidence of response rates and a higher level of incorrect responses. In another study, Clariana and Koul (2006) analyzed the extent to which a different set of immediate trials following students´ errors benefited the application of concept knowledge on a posttest that assessed the relational encoding from a text. Students read four science texts and responded to high-level learning questions that covered a variety of science principles. High-level questions were designed so they tackled non-contiguous textual idea-units in order to help students construct inferences. Students received delayed feedback that allowed them to see the correct answer after the lesson, single-try immediate feedback, multiple-try immediate feedback, no feedback or no feedback and the text only.

Interestingly, experimenters observed that one trial immediate feedback facilitated the encoding of verbatim, while feedback that involved the learner in a multiple-trial procedure seemed to facilitate the relational encoding of ideas from the text.

In summary, and based on the studies we have mentioned so far, an immediate KCR feedback seems to be highly effective for the correction of errors and the retention of correct responses when academic testing implies the completion of low-level learning tasks (i.e., the recall, repeated testing) (Akensson, 2012; Brosvic et al, 2005; Dihoff, Brosvic & Epstein, 2004; Eipstein, Dihoff & Cook, 2006; Metcalfe, Kornell & Finn, 2009). Contrarily, providing students with a delayed feedback message has reported essential benefits for the long term learning of concepts, in particular when students complete high-level learning tasks. Firstly, a delayed feedback provides students with a spaced opportunity to process learning material. And this spacing effect may even be particularly beneficial for learning when experimenters provide not only KR and KCR, but also more specific conceptual elaborations that are relevant in the acquisition of conceptual knowledge. However, a traditional delayed distribution of Formative Feedback (FF) has classically been designed as a summative set of individual messages delivered after the completion of a number of tasks, seconds, minutes or days (Peverly & Wood, 2001; Smith & Kimball, 2010; Kulik & Kulik, 1988). Up to date, there is no scientific literature analyzing how the most beneficial aspects of a delayed elaborative feedback can be implemented during question-answering in such a way that assures the comprehension of the relevant content from text. Analyzing this point is the main objective of the following subchapter.

2.2. Effects of Content of Formative Feedback on Concept Learning

The relational quality of a Formative Feedback (FF) message can influence concept learning in a decisive way (Butler, et al., 2013; Kluger & DeNisi, 1996; Narciss, 2014; Shute, 2008). In fact, several authors have questioned those studies where the effectiveness of simple forms of feedback (i.e., KR and KCR) is analyzed using low-level tasks (i.e., multiple choice question-answering for the retention of words) (van der Kleij, et al., 2015). For example, Andre and Thieman (1988) found that correct answer (KCR) and self-correction feedback benefited performance on repeated questions that required the remembrance of concepts, but did not enhance learning on those questions requiring a more complex type of constructive activity (i.e., the application of the conceptual knowledge to new learning tasks). Since the effects of Formative Feedback (FF) on concept learning may occur at different levels, for example at the task level (Hattie & Timperley, 2007; van der Kleij, et al., 2016), the content of an elaborative feedback message may be designed to address two main types of learning outcomes, low or high-level (*for a systematic review see* Jaehnig & Miller; 2007). In their review, Van der Kleij et al. (2015) describe this basic level of learning as the one induced by those tasks that only require students to recognize and recall words or facts without the need to elaborate upon this knowledge. In fact, these authors found that EF was similarly effective on low-order outcomes (at least 0.37) and positively effective on higher order learning outcomes (at least 0.49). According to these authors, when high-level learning processes are taken into account, elaborated types of Formative Feedback (FF) seem to be commonly preferred by researchers over the other types (Bangert-Drowns et al., 1991; Hattie & Timperley, 2007; Shute, 2008; Van der Kleij et al., 2015).

In another study, Butler, Mash and Godbole (2013) demonstrated how elaborated feedback was able to prompt high-level learning when students dealt with reading tasks that required complex elaboration of knowledge from the text. Authors presented students with ten reading passages and required them to answer a set of definition and inference open-ended questions. Feedback was provided in the form of KCR (a statement of the correct answer) and explanation feedback (KCR accompanied by two sentences elaborating on the correct answer). A group without feedback was included as a control. Learning measures included both, an initial test and a final test containing repeated and new inference open-ended questions. Experimenters found that both types of feedback (KCR and KCR+EF) led in a similar manner to a great proportion of retention on the posttest questions relative to not receiving feedback. However, only the combination of KCR+EF resulted in greater higher-level learning at the final test with new inferential questions.

Similarly, Kapp et al (2015) provided concept-elaborative feedback in an immediate or delayed manner in order to analyze the type of processing activity students engaged in during question-answering. Students read a set of prose excerpts and received elaborated feedback on a question per question basis (after each one of the sixteen subchapters), after eight questions and eight subchapters or, after a block of four learning questions and four subchapters. Feedback consisted of KCR feedback and a detailed conceptual message regarding why a particular answer was correct or incorrect. Results suggested that students receiving elaborated feedback on a question per question basis processed feedback messages quicker compared to those receiving feedback every four ($p < .01$) or eight questions ($p < .001$). Moreover, a delayed presentation of the learning questions and feedback induced students into further reading compared to their counterparts in the other experimental conditions.

It is of importance to mention that, within the context of question-answering for the acquisition of concept knowledge, the ultimate goal of Formative Feedback (FF) is to help readers overcome comprehension difficulties by helping them draw inferences between ideas and repair false-links in the mental representation (Golke, et al., 2015). However, the argument in regards to what content of feedback is more effective for low and high level learning still needs further consideration. According to the results we have reviewed in the present chapter delayed elaborative feedback messages are more likely to benefit learning, especially when students deal with high-level learning tasks.

Nevertheless, when students answer questions from an available text Elaborative Feedback (EF) should be developed in close relationship to the learning content (i.e., relevant learning ideas present in the text) and the type of learning task (low-level versus high level).

Indeed, Formative Feedback (FF) can include messages that combine KR or KCR with the more relevant elaborative content aimed at helping students understand specific cues between concepts (Van der Kleij, et al., 2015). In the present work, we refer to these types of feedback combinations of components as *the level of integration of the elaborative message* within a feedback message. From this perspective, the level of integration of the elaborative Formative Feedback (FF) is understood as an elaborative message that integrates the relevant learning ideas and that helps students build a well-inter related and coherent mental model from the text. Its main effect is expected at the individual level of the student, by helping him integrate and create meaningful responses to both, low and high-level learning questions.

To this, and as proposed in the present work, an efficient Formative Feedback (FF) message that provides a coherent explanation and that helps students learn and understand a set of integrative learning ideas in the text, might facilitate the connection between relevant ideas in a more constructive manner. Up to date, there is no research that has focused on these aspects within the elaborative feedback message during question-answering activities that aim to teach concept knowledge.

From this standpoint, a delayed feedback may be delivered in both, as a summative set of individual messages right after the completion of a number of tasks (i.e., learning questions), or in the form of a unique elaborative message that integrates the relevant learning ideas from the text in a coherent manner so it helps students to understand and build a more coherent and elaborative mental model from the learning content. For this, providing an Immediate Elaborative feedback (EF) message (i.e., *unfortunately, your answer is not correct; Atmospheric pressure can be measured with a mercury barometer*) might have a different effect on learning than, a Delayed Summative set of individual feedback messages to each of the questions delivered after the completion of four questions. In this case, this procedure provides students with a spaced presentation of Formative Feedback (FF) which could help students identify and discard the non-relevant information to answer more effectively (i.e., incorrect answers). Nevertheless, an integrative elaborative message may act as a new restudy opportunity by facilitating the comprehension of the relevant content from text while providing students to set coherence between cross paragraphs.

Additionally, the present work also focuses on those metacognitive aspects described by the literature in feedback effectiveness (Bangert-Down et al., 1991; Butler & Winne, 1995; Narciss, 2013; Vidal-Abarca et al., 2017). For example, knowing in advance that one will receive feedback and will have a second attempt in case of failure could produce different effects on the time and effort students devote to searching the text. These situational variables may induce readers to read the text initially less carefully than knowing that no feedback would be provided, since receiving Formative Feedback (FF) on a particular question may be indicative of when careful reading may be necessary. In addition to this, the level of integration of the elaborative feedback message may also be a key factor. For example, after receiving a message that contains Knowledge of Correct Feedback (KCR) and an integrative elaborated message right after the completion of four questions, students in this delayed feedback condition may decide to search the text to a larger extent compared to students in the immediate feedback condition. However, when receiving a summative set of feedback messages after the completion of four questions, Formative Feedback (FF) could be a proper substitute for the text as the only source of information available, so students could choose to reread the cumulative set of feedback messages more carefully instead.

Based on the theoretical models we previously reviewed, in the third subchapter of the present thesis we provide a theoretical overview of two of the most studied variables within the context of Formative Feedback (FF) instruction as well as their relevance for the processing of the feedback message. For this, we first present response certitude as a metacognitive variable which is implied in the monitoring of the students´ performance. For example, we could assume that a student might feel more or less certain of his ability and resources to answer correctly after a mistake when a second attempt to answer a question is provided, right after receiving corrective feedback on a question per question basis. As a result, he may devote more time and extra effort to reread the available text and might be more willing to closely analyze the relevant ideas from text. Second, we dive into previous background knowledge, which is described as an important cognitive variable for the processing of new information within the context of instruction and learning (Banger-Downs et al., 1991; Kulhavy, Stock & Hancock, 1989; Kulhavy, Stock, Hancock, Eswindell, & Hammrich, 1990).

3. The role of individual variables and Formative Feedback: Response Certitude and Previous Background Knowledge.

3.1. Response Certitude

Response certitude is by far one of the most studied individual variables within the framework of feedback research (Kulhavy & Stock, 1989). Kulhavy et al., (1990) defined this variable as the subjective judgment of certainty that concerns the degree to which a learner perceives a given piece of information as easy to be processed. As such, response certitude is part of the monitoring process that results from the students´ evaluation in regard to both, the task (i.e., content, availability and difficulty) and the perceived ability to perform it correctly (Kulhavy & Stock, 1989). Past research in self-regulated learning in question-answering has considered response certitude as a type of *Judgment of Learning (JOL)* that has been widely measured utilizing scales similar to those used by Kulhavy and Stock (1989). In fact, in question-answering tasks with an available text, the monitoring decisions to reread a text are often dependent on JOLs of certainty. For example, Vidal-Abarca et al. (2010, experiment 2) tested this hypothesis. These authors asked secondary-school students to read two texts and answer twenty questions on a computer-based system. After reading each question, students were asked to make a sort of Judgement of Learning (JOL) decision. After reading each question, students were asked, "*How well will you be able to answer this question correctly without rereading the text?*" Students could answer on a scale ranging from 0 (I will definitely not be able to) to 100 (I will definitely be able to), with 20, 40, 60, and 80 as the intervals and that was recorded by the system. Data reported that students tended to decide not to search the text when they were certain to be able to answer the questions correctly.

Nevertheless, within question-answering feedback contexts, JOL scales are aimed at gathering the students´ certitude of response with both, a first and second attempt to answer a given question. Examples of scales´ wording include statements such as "*How* *confident you are that you answered to this question correctly".* Right below the statement, scale´ rating ranges from a 0% to a 100%. Interestingly, past literature has found that certitude of response strongly determines the students´ ability to correct initial wrong responses (Buterfield & Metcalfe, 2006; Butterfield & Metcalfe, 2001; Kulhavy & Stock, 1989). In fact, after receiving feedback, errors endorsed with high confidence, are more likely to be corrected than errors endorsed with low confidence (Fazio, 2009). The corrective function of feedback would induce the student into a *Hypercorrection effect* determined by his efforts to calibrate his internal standards of reference with the expected level of competence (Kulhavy & Stock, 1989; Bangert-Downs et al., 1991). For example, Butler et al., (2008a) found that when feedback was provided, a high proportion of high-confidence incorrect responses were corrected on a second attempt. In addition, when feedback was provided, a great proportion of initially correct responses were maintained in a posttest.

According to Kulhavy and Stock´ (1989) Model of metacognitive monitoring, when a student is required to complete question-answering tasks, a set of different stages takes place. In a first stage, the student responds to the item following the experimental instruction. In a second stage the student receives formative feedback which confirms or disregards the provided response. Consequently, feedback effectiveness strongly depends on the impact of the verification feedback statement of correctness on the student´s subjective beliefs of performance. For example, when a student receives a verification of a correct answer (e.g., *your response is right*) and he is certain that his response was correct; there is no discrepancy between the student´s confidence of correctness and the feedback message. However, when a feedback message disregards a given response but the student was confident that his response was correct, there will be a high level of discrepancy between his belief of correctness and the feedback message. Desirably, this discrepancy would induce the student to unfold the learning strategies he considers he needs to adjust his performance and reach the standard goal of learning set by the feedback provider (i.e., teacher, peer or computer system).

For example, strategies may include dedicating more time and careful effort to search a text to find the right answer (Mory, 2004; Narciss, 2013; Kapp, et al., 2015; Kulhavy & Stock, 1989). In a third stage, after a feedback message is provided, the student has a second chance to respond. This question-answering cycle allows students to become more aware of the learning process by requiring them to monitor their performance, correct wrong answers and adjust their certainty degree of certitude to their responses (Bangert-Down, et al., 1991; Vidal-Abarca et al., 2017). According to this, students are not only capable of assigning some degree of certainty to a given response (Kulhavy & Stock, 1989), but this monitoring process is intrinsically related to the strategic decisions (i.e., when and what to search) students perform during question-answering and feedback processing (Vidal-Abarca et al., 2017; Vidal-Abarca et al., 2010).

In addition to this, it has also been repeatedly found that students tend to overestimate their knowledge or understanding, and that is, they normally report a high level of certitude in their answers regardless of the actual answer correctness (see Dunlosky & Metcalfe, 2009; Kealy & Ritzhaupt, 2010; Ramos & Vidal-Abarca, 2013). Butler, Karpicke and Roediger (2008, experiment 2) examined this specific issue. They presented undergraduate students general knowledge multiple-choice questions about facts taken out of the World Book Encyclopedia (e.g., *what is the longest river in the world? the answer is the Nile river*). Questions were presented to students in two groups of twenty questions each (there was a third group that is irrelevant for the purpose of the present work): test without feedback, and test with feedback. Questions were presented first as a test, and then as a posttest. After answering each question, either on the test or the posttest, students were prompted to rate their confidence of having responded right on a scale 0-100%, from guessing to totally certain. Students in the two conditions differed in that those in the feedback condition were informed about the correct response (KCR feedback) after rating their confidence of response, whereas those in the other condition were not.

These authors found that feedback after the test made students increase their correct responses on a posttest, which sounds logical since it reinforces the association between the question and the response. However, the most interesting result was that when feedback was provided on the initial test, students produced more accurate predictions (i.e., confidence of response judgments) on the posttest. In other words, feedback made students better able to discriminate between correct and incorrect responses on the posttest.

Thus, the Butler et al. (2008) study suggests that feedback can help to decrease the discrepancy between perceived and actual correctness of response. Therefore, providing feedback after a first attempt on a question-answering situation may make students more accurate on their judgments. In other words, when they claim that they are very confident of giving a right answer to a question, they will be more likely to give a right answer whereas when they are less confident, the likelihood of giving a right answer will be lower. However, it is very common to assume that when students provide lower certitude estimates it is because they are unable to acknowledge task referents appropriately. Nevertheless, quite often students are free to independently reread some sort of source information (i.e., a text excerpt) while responding questions and receiving formative feedback (Máñez et al., 2017; Martínez et al., 2017; Kapp et al., 2015; Llorens et al., 2014).

These sorts of learning tasks quite often require from students not only the comprehension of questions, but more strategic processing and decisions such as when and how to search a text more efficiently. In fact, learning questions have shown to influence students´ monitoring decisions, searching processes and learning outcomes (Cerdán et al., 2009; Goldman & Durán ,1988) in particular when they help students identifying the relevant information (McCrudden & Schraw, 2007). Consequently, the extent to which a student is able to assess both task demands and his own cognitive memory resources to answer correctly is key in the process. For this, response certitude scales are sensible measures to gather the students´ self-perception of performance while answering questions and receive Formative Feedback (FF) on their performance (Vidal-Abarca et al., 2017). Since the information presented in the feedback message has an important instructional function for the correction of incorrect responses and the maintenance of correct responses over time (Kulhavy & Stock, 1989), this monitoring process might be influenced not only by the processing demands of a particular reading task, but also by the timing and content of the feedback message. At the same time, corrective feedback presented on a question by question and immediate basis may induce a type of processing that is quite different from that one imposed by a delayed presentation of feedback. As Kapp and colleagues (2015) described, these instructional conditions might influence the way students unfold question-answering strategies (i.e., time and effort reading a text), organize information from memory in order to answer, and how students pay attention to and process Formative Feedback (FF). And that is the main objective of the present work. For this reason, in *Study 1* a main objective is to analyze this specific assumption by requiring students to complete a response certitude scale right before providing a response to a question-answering task that delivers corrective formative feedback (KR or KCR) on the students´ performance while allowing them to independently search a text after a mistake is committed. In *Study 2*, this monitoring process follows a perspective similar to that one by Gil, Martínez and Vidal-Abarca (2015), where the monitoring accuracy of the students´ search decisions was described as a significant predictor of performance. In our case, right after receiving feedback, students will be free to monitor when and for how long to refer back to an available text for further search.

3.2. Previous Background Knowledge

Previous literature has already demonstrated the crucial role of previous background knowledge for concept learning from academic texts (Gilabert, Martínez, & Vidal-Abarca, 2005; Le Bigot & Rouet, 2007; McNamara & Kintsch, 1996; Priebe, Keenan, & Miller, 2011; O´ Really & Sabatini, 2013; Ozuru et al., 2007; Salmerón, Kintsch & Cañas, 2006; Vidal-Abarca & SanJosé, 1998; Clemens et al., 2018). Previous knowledge has reported being highly related to information processing since it allows readers to integrate new ideas into a coherent mental model from the text (McNamara et al., 1996). PBK acts then as a pre-existing net which helps students establish explicit, relational and elaborative relationships between relevant learning ideas (i.e., connections between referents and referrals). Additionally, it also allows readers to discard irrelevant ideas from the relevant ones (Le Bigot & Rouet, 2007; Vidal-Abarca & SanJosé, 1998). Learners with higher levels of prior knowledge are more likely to understand intricate knowledge from the text (i.e., of the semantic relations) (Salmerón, Kintsch & Cañas, 2006).

Within the context of feedback instruction, previous knowledge has been defined as a crucial component and moderator for feedback effectiveness by several authors (Fyfe, Rittle-Johnson & DeCaro, 2012; Fyfe, 2016; Hannafin, Hannafin, & Dalton, 1993; Narciss, 2011; Smits, Boon, Sluijsmans, & Van Gog, 2008). In fact, within the field of Sciences previous knowledge has reported to moderate the effects of different types of feedback on elementary-school and university students, in particular with a low level of prior knowledge (i.e., Fyfe & Rittle Johnson; Fyfe, 2016; Fyfe, Rittle-Johnson, & DeCaro, 2012; Krause, Stark, & Mandl, 2009). This statement is consistent with previous theoretical models describing how individual variables play a decisive role in feedback instruction and its efficiency (Bangert-Downs et al., 1991; ITF, Narciss, 2013; Narciss, 2014; Kapp et al., 2015). As Hattie and Timperley (2007) state, in order to be genuinely useful, adaptive feedback must align with the student´s prior knowledge.

In a study, Fyfe, Rittle-Johnson and DeCaro (2012) required school students to complete twelve problems in Mathematics. Students performed a set of tests in a pretest, one-on-one computer-based tutoring intervention and an immediate posttest. Right after each problem, students received strategic formative feedback, outcome formative feedback or no feedback. On the one hand, strategic feedback consisted of a message irrespective to the correctness or incorrectness of the students´ answer, which focused on how they had solved each problem independently (e.g., *Right! That is one correct way to solve that problem*). On the other hand, outcome feedback contained a KCR answer feedback on their responses to the problem. Results reported that for those children with low-prior knowledge, feedback led to higher procedural knowledge of the correct strategies than no feedback. However, for children with a moderate prior level of knowledge of the correct strategies, outcome-feedback promoted more excellent concept knowledge than strategic feedback. In another study, Krause, Stark and Mandl (2009) compared the effects of a cooperative learning strategy with a feedback procedure on performance in problem-solving in statistics. After completing each problem, students received strategic feedback in the form of a worked example that they could compare with their solution. Also, those students in the feedback groups performed six additional multiple-choice tests while receiving adaptive elaborated feedback.

Results reported that feedback was especially beneficial for low prior knowledge students who scored higher when the feedback was available. Authors concluded that feedback effectiveness was strongly supported by the multiple-choice testing, which allowed students in the feedback conditions to engage in the additional practice and therefore learning.

With regards to feedback timing, Fyfe and Rittle-Johnson (2016) demonstrated the effectiveness of a computer-generated feedback procedure that provided feedback in an immediate or summative manner compared to a non-feedback condition during a problem-solving activity. A sample of second-grade children received instructions on a correct problem-solving strategy and then solved a set of twelve equivalence problems. Children received one of three options: no feedback, correct-answer feedback in an immediate manner problem after problem, or at the end of the intervention in the manner of summative feedback. During the procedure, children were required to answer a set of self-assessment scores on their performance. The problems and correct answers remained on the screen setting, which allowed students to compare answers across problems while they received feedback. Results reported that during the training session, immediate trial-by-trial feedback was particularly beneficial for problem-solving performance. It seemed to induce students to deploy the correct solving strategy, especially for children with low prior knowledge.

Nevertheless, this had minimal impact on children with higher prior knowledge. Similarly, posttest data reported that both immediate feedback and summative feedback boosted students´ scores relative to no feedback, especially for students with low prior knowledge. Finally, these authors also observed that only immediate feedback improved the mastery of the material for both high and low prior knowledge students.

Altogether, the studies mentioned above clearly confirm the role of previous background knowledge (PBK) within the context of feedback instruction in the field of Mathematics and Statistics. Nevertheless, apart from a recent study within the frame of question-answering with an available text (Máñez et al., 2016) focusing on the mediation effects of PBK on learning, up to date there has been no extensive research analyzing the effects of timing and content of formative feedback on the acquisition of concept knowledge from the text during question-answering. Based on this, *Study 2* in the present thesis analyzes the role of previous background knowledge in the effectiveness of a computer-based question-answering procedure that provides immediate or delayed elaborated feedback which varies in the level of integration of its elaborative message (i.e., summative or integrative).

It seems logical to consider that prior knowledge may moderate the degree to which students fulfill adequately when reading a text and answer performance and learning questions while receiving feedback that differs in its timing of distribution (Immediate, Delayed Summative and Delayed Integrative). Moreover, low and high knowledge students may acquire conceptual knowledge in a different way depending on the timing of feedback that students receive (i.e., question by question versus every four questions).

First, students with lower levels of prior knowledge could benefit from elaborative feedback provided right after answering each question, over other forms of delayed feedback (summative every four questions or integrative elaborative provided after the last question, in the form of a unique message). Immediate elaborative feedback would allow students to detect and correct wrong inferences in a more sequential manner (i.e., by systematically distributing the content on a question by question basis). An elaborative feedback message consisting of an explanation of why a correct answer is correct could compensate for the lack of previous knowledge.

On the contrary, those students with higher levels of prior knowledge could benefit the most from delayed integrative feedback, simply because these types of messages consist of integrative elaborations that provide a relation of learning ideas more coherently. A higher level of prior knowledge would allow students to integrate the new incoming information and feedback content into their mental model from the text in a more practical way.

4. Timing and Level of Integration of Formative Feedback for Question-Answering Performance and Concept Learning

4.1. General and Specific Objectives

In light of the previous literature reviewed in the present work (Bangert-Downs et al., 2013; Butler & Roediger, 2008; Butler & Winne, 1995; Hattie & Timperley, 2007; Kapp et al., 2015; Karpicke, & Roediger, 2007, 2008; Llorens, et al., 2014; Máñez, Vidal-Abarca & Martínez, 2016; Máñez, Vidal-Abarca, Martínez & Kendeu, 2017; Newman, Williams & Hiller, 1974; Narciss, 2013; Narciss et al., 2014; van der Kleij, Eggen, Timmers, & Veldkamp, 2012; van der Kleij, et al., 2015; Vidal-Abarca et al., 2017; Vidal-Abarca, Martínez, Ferrer & García, 2017), the general objective of the present thesis is the evaluation of the effectiveness of a formative feedback question-answering procedure that provides feedback in an immediate and delayed manner, while students answer questions with an available text. Also, the level of integration of the elaborative feedback message is specifically designed to provide students with a relation of the most relevant ideas from the text, in an attempt to help them comprehend learning content in a more coherent manner.

Study 1 analyzes the impact of a question-answering procedure that provides feedback in an immediate (question by question), or delayed manner (after the completion of four questions), compared to a no-feedback condition where a sample of undergraduate university students was forced to re-read a text after committing a mistake. Measurements included the student´s probability to correct errors, response certitude estimates, question-answering performance and online behaviour. Delivering immediate corrective feedback (KR, KCR) while the text is available right after committing a mistake, may establish the referents for students to unfold valuable error correction skills at their own pace (i.e., further reading, location and the relation of ideas situated within distant pieces in the text). Thus, a feedback procedure that provides them with a two-trial attempt to answer after a mistake is committed might allow students to be more aware of errors since they are allowed to monitor their understanding more efficiently and provide absolute certitude of responses concerning their performance. Additionally, *Study 2* dives into the issue of what type of level of integration of the elaborative feedback message is more beneficial for concept learning when provided in an immediate, delayed summative or delayed integrative form.

Following the most common in-classroom practices within the framework of computer-assisted learning, in this study, delayed feedback is provided as both delayed summative or delayed integrative. While delayed summative is delivered in the form of four individual messages to each of the questions, delayed integrative feedback is provided as a unique message, after the completion of four learning questions conceptually related. In this study, the moderation effects of Timing of Feedback (Immediate, Delayed Summative and Delayed Integrative) and Previous Background Knowledge on performance and learning are explored. Additionally, task level (i.e., low versus high-level) and individual differences in performance and learning based on the level of previous background knowledge of students are also considered discussed under the light of the recent literature in reading comprehension and learning from text (Gilabert, Martínez, & Vidal-Abarca, 2005; Le Bigot & Rouet, 2007; McNamara & Kintsch, 1996; Priebe, Keenan, & Miller, 2011; O´ Really & Sabatini, 2013; Ozuru et al., 2007; Salmerón, Kintsch & Cañas, 2006; Vidal-Abarca & SanJosé, 1998; Clemens et al., 2018).

Compared to the processing students engage in when immediately receiving feedback, delayed integrative feedback (i.e., every four questions) would induce in students those cognitive processes necessary for a deep and more coherent understanding of the learning material. This process could allow students to process the relevant ideas from the text as a whole while being able to relate that knowledge with a set of questions conceptually related.

In the following chapters, the introduction, hypothesis, methodology, results and conclusions for both, *Study 1* and *Study 2*, are presented consecutively. Finally, the thesis discloses the general conclusions and limitations of the present work, which mainly focus on some of the most critical points in regards to how computer-based systems and feedback affect concept learning and the type of processing students engage in (low and high) when answering questions with an available text. Educational implications are further discussed.

CHAPTER TWO

STUDY 1: EFFECTS OF TIMING OF CORRECTIVE FORMATIVE FEEDBACK COMPARED TO RESTUDY ON QUESTION-ANSWERING PROCESSES

Adapted from: Candel., C., Vidal-Abarca, E., Cerdán, R., Lippmann, M., & Narciss, S. (2019). Effects of Timing of Corrective Formative Feedback compared to Re-study on Question-Answering Processes, (*manuscript in preparation*).

2.1. ABSTRACT

This study examines the effects of the timing of corrective formative feedback on processing text information on question-answering. Undergraduate students read an expository text and answered questions in two attempts. Students were randomly assigned to a no feedback, immediate feedback and delayed feedback conditions. Students in the feedback conditions received feedback on the correctness of their answer after the first attempt and were informed about the right answer after the second attempt. Students were prompted to restudy the text after failing in their first attempt. However, students in the no feedback condition were prompted to search the text. All students were tested on question-answering, reading performance and a posttest cued-recall test. Results showed that: (a) feedback reduced the initial time reading the text; (b) feedback increased performance on question answering and cued- recall; (d) delayed feedback produced no advantages over immediate feedback. Theoretical and practical implications of these results are discussed.

Keywords: formative feedback, corrective feedback, question-answering, feedback timing, reading comprehension.

2.2. INTRODUCTION

Answering questions from an available text supports learning associated to reading (e.g., Anmarkrud, McCrudden, Bråten & Strømsø, 2013; Cerdán, Gilabert, & Vidal-Abarca, 2011; Graesser & Murachver, 1985; Roelle & Berthold, 2017; Vidal-Abarca, Mañá, & Gil, 2010). In fact, instructional questions are often used to guide the students´ learning of relevant content from the text (McCrudden, Magliano, & Schraw, 2010). As such, questions may be distributed in different ways across a lesson, especially when the main objective of the questions is the teaching of verbal material (i.e., pair-association or factual). For example, after an initial reading of a text, teachers may require students to answer a set of questions either one by one or right after the completion of specific components of a lesson (Felker & Dapra, 1975; Hamaker, 1986; Rothkopf & Bisbicos, 1967).

In this type of question-answering activities, effective teachers also provide formative feedback in regard to specific aspects of the students´ performance, such as the correctness or incorrectness of a response to a particular question (Ness, 2011; Pressley, Wharton-McDonald, Mistretta-Hampston, & Echevarria, 1998). However, this can be a very complex task within the current educational contexts since there are several variables that can interfere with the learning process, as the increasing number of students per class (Van der Kleij, et al., 2015). In light of this, Computer-Based Systems (CBS) allow educators to effectively record the students´ question-answering activity (i.e., correct or incorrect answers and reading times) and provide formative feedback to students' responses (e.g., Kapp, Proske, Narciss & Körndle, 2015; Llorens, Vidal-Abarca & Cerdán, 2016; Maier, Wolf, & Randler, 2016; Narciss et al., 2014; Shute & Rahimi, 2017; van der Kleij, Eggen, Timmers, & Veldkamp, 2012).

Computer-Based Systems (CBS) quite often also include learning questions that provide students with a formative feedback strategy that, after the first attempt with a question, informs readers about the correctness of their responses (i.e., correct vs. incorrect) (Narciss et al., 2014; Shute, 2008). As such, after a first try, a computer system provides Knowledge of Results (KR) feedback, which lets students engage in a second attempt with responding to a question, and Knowledge of the Correct Response (KCR) after the second attempt (e.g., Attali, 2015; Conole, 2016; Narciss & Huth, 2006). In spite of the general benefits reported by formative feedback (e.g., for an example see the reviews by Shute, 2008; van der Kleij, Feskens, & Eggen, 2015), some studies in the field of text-based question answering research have found that providing formative feedback does not benefit learning over letting students search the text on question-answering (Moos, 2011).

Few studies have examined question-answering processes such as the time students invested in text reading, responding to questions or attending to the feedback, to try to understand the effectiveness of formative feedback (Butler, Karpicke & Roediger, 2008; Kapp et al., 2015). Further, studies examining how feedback affects the students' certitude of having responded right to the questions are scarce. One study conducted by Butler, et al. (2008) revealed that formative feedback was beneficial (a) for correcting erroneous responses, (b) for confirming correct responses with low response confidence, and (c) for increasing monitoring accuracy. However, this study did not investigate to what extent feedback effects also influence the students' effort and time spent in restudying and in error-correction. In addition to this, as previous research in the field indicates (Butler, Karpicke, & Roediger, 2007), feedback timing (i.e., providing feedback immediately after a response, or alternatively after some delay) may affect performance and question-answering processes. Therefore, this study has been designed to examine the effects of providing immediate or delayed feedback over providing no feedback on question-answering processes and performance, as well as on the students' certitude of response.

2.2.1. Question-Answering Processes and Formative Feedback

Research on formative feedback shares a number of assumptions (Hattie & Timperley, 2007; Mory, 2004; Narciss & Huth, 2004; Narciss, 2008; Shute, 2008; van der Kleij, et. al., 2015). First, formative feedback is conceptualized as information provided by an agent or a Computer-Based System (CBS) that aims at promoting the subsequent performance or understanding of a learner or more generally spoken a person. To be effective, feedback needs to provide information specifically related to the task or process of learning (see the seminal meta-analysis by Kluger & DeNisi, 1996), including (a) information about the current level of performance or understanding, and (b) information on how to fill the gap between what is understood and what is aimed to be understood. This is why feedback is considered to be particularly useful for correcting the wrong answers, but also for confirming and thus consolidating correct answers with low response-certitude (Butler & Roediger, 2008). Second, formative feedback may include different types of information. The simplest information is knowledge of results (KR) feedback verifying if the response to a question is correct or incorrect. Knowledge of Correct Response (KCR) provides the correct answer (e.g., the right answer is X). Finally, one of the most common types of feedback is elaborated feedback (EF). Elaborated feedback (EF) can take many forms, such as hints, or explanations about the correct answer, to name a few subtypes of EF (see Narciss, 2008 or Shute 2008 for further subtypes). Third, models of formative feedback consider effects of feedback on performance, as well as on cognitive and metacognitive aspects related to the task, although they differ in the specific aspects they consider. Some models consider very specific aspects. For instance, Kulhavy and Stock (1989) emphasized the role played by the students' confidence in their response.

Other models include many cognitive and metacognitive aspects. For instance, Bangert-Drowns, Kulik, Kulik and Morgan (1991) consider different stages, beginning with the initial state of the learner, search and retrieval strategies activated by the task, the response, and how feedback activates evaluation and adjustment processes. Prior knowledge, goals, memory processes, or the degree of certainty on the responses are cognitive and metacognitive aspects considered in this model. Narciss (2008; 2013) developed a conceptual framework (Interactive Tutoring Feedback Model) that considers the interaction between two loops, one pertaining to the learner, whereas the other corresponds to the external feedback source. Cognitive (e.g., internal representations of task requirements and standards), metacognitive (e.g., monitoring and self-regulation), and motivational (e.g., perceived task values and competencies) factors are relevant in this interaction. However, models of feedback are quite general. Therefore, specific cognitive and metacognitive processes for the task at hand, i.e., question-answering in our case, need to be considered when analyzing the effects of feedback.

Question-answering with an available text is an activity that combines comprehension and problem-solving processes (Rouet, Britt, & Durik, 2017). Comprehension processes include processing text information and forming an appropriate task or question model (Vidal-Abarca, et al., 2010). Problem solving processes involves taking strategic actions to achieve the reader's main goal (i.e., answering the questions), which involves searching either on the texts or on the reader's mind for an appropriate answer, as well as monitoring and self-regulating the goal's fulfillment (Farr, Pritchard, & Smitten, 1990; Rupp, Ferne, & Choi, 2006). Question-answering activities usually begin with reading the text. When reading the text initially, there are situational variables (e.g., knowing in advance the availability of the text when answering the questions) that affect the initial processing of the text. Thus, some authors have found that readers spent less time reading a text when they know that the text will be available than when they know it would be unavailable (Ferrer, Vidal-Abarca, Serrano & Gilabert, 2017; Higgs, Magliano, Vidal-Abarca, Martínez & McNamara, 2017). Ferrer et al. (2017) interpret this finding by claiming that the available situation induced readers to decrease their standards of coherence (van der Broek & Helder, 2017), which caused them to detect fewer lacks in the coherence building process while reading the text initially.

In other words, readers in the text available situation read the text less carefully due to their awareness that they could reread the text when answering the questions. Knowing that one will receive feedback after answering questions and that one will have a second attempt to give an answer may produce a similar effect. This situation may induce readers to read the text initially less carefully than knowing that no feedback would be provided, since feedback may be indicative of when careful reading will be necessary.

After the initial reading of the text and reading the questions, readers have to make some decisions, one of which is whether to reread the text or not. When answering questions, readers vary in their judgment about how accurate they are able to answer the questions without rereading the text (Vidal-Abarca, et al., 2010). That is, sometimes they are pretty sure to be able to give the right answer without rereading, whereas other times they consider very unlikely to give the right answer without rereading. These judgments involve monitoring processes, which affect rereading decisions. In fact, Vidal-Abarca et al. (2010, Experiment 2) found that when students are pretty sure to be able to give the right answer without rereading they tend to give an answer without re-reading the text, and vice versa. Getting KR feedback after failing at a question may also affect monitoring processes on question-answering, such as the level of the students’ certitude of response.

2.2.2. Effects of Feedback on the Students' Certitude of Response

According to Kulhavy and Stock´ (1989) model, the discrepancy between the students' confidence of response and feedback is a major factor to explain the time and effort students devote to correct wrong responses when they have a second attempt. When students receive verification of a correct answer (e.g., your response is right) and they are certain that their response is correct, there is no discrepancy between the students' confidence of response and feedback. However, when they are informed that their answer is wrong but they are confident that the response is correct, there will be a high level of discrepancy between their belief and feedback. This discrepancy will induce students to search the text to find the right answer. The model has been tested extensively (see the review by Mory, 2004).

It has been repeatedly found that students overestimate their knowledge or understanding, that is, they often show a high level of certitude of their answer regardless of the actual answer correctness (see Dunlosky & Metcalfe, 2009 for an extensive review of this issue). It is interesting to know to what extent feedback may contribute to making students more accurate on their predictions. Butler, Karpicke and Roediger (2008, experiment 2) examined this specific issue. They presented undergraduate students general knowledge multiple-choice questions about facts taken out of the World Book Encyclopedia (e.g., *What is the longest river in the world?* the answer is *the Nile river*). Questions were presented to students in two groups of 20 questions each (there was a third group that is irrelevant for our purpose): test without feedback, and test with feedback. Questions were presented first as a test, and then as a posttest. After answering each question, either on the test or the posttest, students were prompted to rate their confidence of having responded right on a scale 0-100%, from guessing to totally certain. Students in the two conditions differed on that those in the feedback condition were informed about the correct response (KCR feedback) after rating their confidence of response, whereas those in the other condition were not. The authors found that feedback after the test made students increase their correct responses on a posttest, which sounds logical since it reinforces the association between the question and the response. However, the most interesting result was that when feedback was provided on the initial test, students produced more accurate predictions (i.e., the confidence of response judgments) on the posttest. In other words, feedback made students better able to discriminate between correct and incorrect responses on the posttest. Thus, the Butler et al. (2008) study suggests that feedback can help to decrease the discrepancy between perceived and actual correctness of the response. Therefore, providing feedback after the first attempt on a question-answering situation may make students more accurate on their judgments.

2.2.3. Effects of Feedback Timing on Question-Answering Processes and Reading Performance.

Effects of feedback timing on performance are mixed. Thus, Mory (2004) concludes that studies supporting the superiority of delayed over immediate feedback were conducted using list learning or similar. Hattie and Timperley (2007) suggest that delayed feedback is better for difficult items, although immediate feedback works best for easy items. The reason is that difficult items are more likely to involve greater degrees of processing about the task than easy items, and that delayed feedback provides the opportunity to do this. In a similar vein, Shute (2008) argues that when feedback is intended to facilitate lower-order learning outcomes, immediate feedback works best, but that when higher-order learning outcomes are at stake, it is better to provide feedback with a delay.

In a very careful study, Butler and Roediger (2008) tested the effect of immediate vs. delayed feedback effect over factual learning. Undergraduate students read prose passages covering a variety of historical topics and answered four fact multiple-choice questions per text. Students received feedback on some passages (and questions), but not on others. Feedback was provided after each question (immediate condition) or after answering four questions (delayed condition). The main interest of the authors was on how the two feedback modalities affected the increase of correct responses and the reduction in the proportion of intrusions (i.e., lure responses from the initial multiple-choice test) on a delayed cued recall test. Delayed feedback led to a higher proportion of correct responses than did immediate feedback, but both feedback schedules were equally effective at reducing the amount of misinformation acquired.

The authors explain the superiority of delayed feedback with two compatible mechanisms. Delayed feedback could make incorrect responses to dissipate, making the correct responses provided with feedback easier to learn. Additionally, delayed feedback could also contribute to a spaced presentation of the materials, which is more effective than massed study for enhancing the retention of verbal material, characteristic of immediate feedback (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006).

While the effects of feedback timing on outcome variables have been extensively investigated, the issue of how feedback timing affects processing variables such as time-on-task has received less attention. Regarding the question-answering process, which is the main focus of the present paper, there is no reason to assume any difference between immediate and delayed feedback for the initial reading of the text, previous to receiving either immediate or delayed feedback. Regardless of the feedback schedule, students may read the text initially less carefully when they know they will receive feedback than when they know that no feedback would be provided. However, feedback timing may affect text search behavior. For example, studies conducted on the basis of the response certitude model reveal that post-feedback study times are longer when students encounter discrepancies between their response confidence and the feedback (e.g., Webb, Stock, &

McCarthy, 1994). As this may well be the case, students in the delayed feedback condition will likely search the text longer before a second attempt than students in the immediate feedback condition, especially when receiving corrective KR feedback after a mistake has been committed. As such, it is possible that the relative monitoring accuracy of their judgments be higher when feedback is provided immediately than after a delay since the association between the students' certitude of response and the information students consider it is correct to provide a response, is performed immediately compared to the delayed feedback, which would require more effortful processing on behalf of the students.

2.2.4. The current study

The main goal of this study was to examine the effects of feedback on question answering processes and performance. A subsidiary goal was to compare the relative effect of immediate versus delayed feedback on the same variables. A sample of undergraduate students read a text and answered twelve multiple-choice comprehension questions with the text available. Students had two attempts to answer each question. After each attempt of answering a question, students rated their certitude of response. Processing data such as time on reading the text and answering the questions was traced through log-files. After a distractor task, students were tested on part of the same question, and they performed a cued-recall task.

Students were randomly assigned to three conditions: no-feedback, immediate feedback and delayed feedback. Students in the immediate feedback condition received also KR feedback after answering each question on the first attempt; then, they had a second attempt, after which they were informed about the right answer (KCR feedback). Students in the delayed condition received also KR-feedback after a first attempt and KCR-feedback after the second attempt, but feedback was provided after a block of four questions. For the no-feedback conditions, students were prompted to search the text, but no corrective feedback was provided. According to the theoretical framework provided earlier, we formulated hypotheses regarding question-answering processes and performance.

First, we predicted that students in the two feedback conditions would read the text initially quicker than those in the no-feedback conditions. Our second hypothesis was that feedback would enhance the students' accuracy of their response certitude judgments on the second attempt in comparison to no-feedback students. We third predicted that students in the delayed feedback condition would search the text longer than those in the immediate condition. Our fourth hypothesis was that students from the two feedback groups would improve their performance for the second attempt and the posttest measures in comparison to the no-feedback group of students. Finally, we predicted that students on the delayed feedback condition would outperform students in the immediate condition on the cued-recall task.

2.3. METHOD

2.3.1. Participants

A sample of 107 senior undergraduate students in English Philology of the University of Valencia (Spain), with a mean age of (*M* = 20. 56; SD = 2.7) years, participated in the experiment in exchange for some extra credit. The data of three students were excluded from specific data analysis due to software failures.

2.3.2. Study Design

For this study we utilized a simple between-subjects factor experimental design in which we manipulated the variable Timing of feedback (immediate, delayed). Therefore, we created two feedback conditions, *Immediate* (feedback delivered on a question per question basis), *Delayed* (feedback delivered after a block of four questions), and a control non-feedback condition. Students were randomly distributed to one of the three experimental conditions as follows, immediate feedback (35 students), delayed feedback (37 students) and non-feedback condition (35 students). The study was developed on two phases, the testing phase and the posttest phase. On the testing phase, participants read a text and then they answered 12 multiple-choice questions. Students had two attempts when answering the questions. Those in the two feedback conditions got feedback, either immediate or delayed, depending on their conditions, after each attempt, whereas those in the control condition received no feedback, although they were asked to answer the questions twice to equate the procedure to students in the two feedback conditions (see details in the Procedure section). In the posttest phase students were tested on a Cued Recall Multiple-Choice Test.

2.3.3. Materials

Text

Reading material consisted of a continuous expository text on Scientific Exploration on Mars. The final version contained a shortened-8-paragraph version of 804 words adapted from its original version as found in *Fostering links between environmental, and space exploration. The Earth and Space Foundation, Cockell, C., White, D., Messier, D. and Dale Stokes, M., Elsevier Science Ltd, 2002*) from IELTS students´ book by Cambridge University Press 2013).

Questions

We constructed twelve inferential multiple choice questions aimed at promoting the activation of high inferential processes (Kintsch, 1998). Following Cerdán´s et al. (2009) systematic construction process, we developed questions that required a deep understanding of the relationships between ideas located within a maximum of three different paragraphs. We developed four answer options for each multiple-choice question similar to those utilized by Ozuru et al., (2007). To this, options included one target correct option and three types of distracters: idea - overlap distracters, error inducing distracters and unrelated distracters. Idea-overlap distracters consisted in surface features present in the text but located in an inappropriate context inducing the reader into false matching of ideas (Cerdán et al., 2010). Error inducing distracters contained erroneous information regarding the ideas present in the text inducing the reader into confusion, and unrelated distracters contained information that was completely irrelevant to the statement of the question per se. In all of the cases the answer to the question was not explicitly stated in the text and it required the student to engage in active inferential activity.

2.3.4. Apparatus

A study-desk interface was developed specifically for this experiment in order to collect students´ data (Narciss, Proske & Körndle, 2013). The interface presented readers with a full screen containing the text. By scrolling up and down students could read the text at their own pace (see Figure 1). A button below the text (i. e., *Next*) gave access to the questions screen. Once students had read the text, students accessed each question individually. Questions were presented in a randomized order. For the Immediate and Delayed conditions (Figure 2 and 3) feedback was designed as a two- loop algorithm that delivered a message after attempts one and two to every comprehension question. In the immediate feedback condition the system provided feedback on a question per question basis while in the delayed feedback condition, the system provided delayed feedback question per question after every block of four questions. Students in both feedback conditions were provided with a second attempt to each question in case they had previously failed in their first attempt.

After a first attempt to a question, in case they answered correctly, confirmatory feedback was provided consisting of the four lure multiple-choice question accompanied by Knowledge of Correct Response Feedback (KCR) and students were redirected onto the next question. However, in case they failed, students received Knowledge of Response feedback (KR) explicitly signaling the incorrect response (Figure 1b). Then, the learner was prompted to re-read the full version of the text by clicking on the *Next* button. Subsequently, the student was required to complete again both, first the comprehension question and second, the response certitude scale. After a second attempt and irrespective of the correctness of their response, students received the comprehension question along with KCR (Figure 3) and were redirected onto the next question.

Finally, in the no feedback condition, following the same order as in the rest of the conditions, students completed each comprehension question twice. After each attempt to each question, students were prompted to re-read the text (irrespective to the correctness of their given answer). In the no feedback condition students were not informed about the correctness of their response.

After students completed the experimental task containing all the twelve multiple-choice questions, students were directed to a screen where the message "*Thanks for your collaboration*" indicated the end of the interactive testing session.

Figure 1

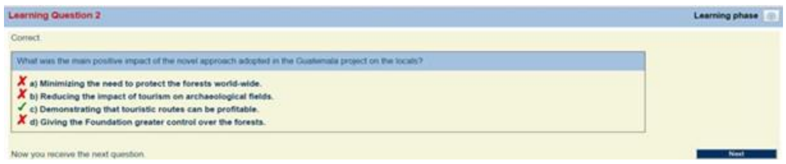
Captions of the text passage during initial reading (students accessed question number one by clicking on the Next button)

Figure 2

Caption of a feedback message after a first incorrect response to a question in the Immediate and Delayed feedback conditions

Figure 3

Caption of a feedback message after a correct second attempt to a question in the Immediate and Delayed feedback conditions



2.3.5. Procedure

Participants were tested in two consecutive sessions (Testing and Posttest phase) that took place during the period of a scheduled university class. Prior to the start of the experiment, the researcher provided instructions on how to perform the experiment supported by a power point presentation and a projector. Additionally, online instructions on how to further proceed were also embedded in written form for each student prior to the start of the testing phase. First, all participants accessed each interactive learning session utilizing a user name and a code that was provided by the researchers. Once students had logged in their online sessions, they were required to complete a set of tasks consisting of the following elements; (1) Demographic questionnaire in which participants reported their gender, age, mother tongue and current university grade, (2) reading specific instructions for each experimental group and (3) completing the online learning session where learning practice took place. The testing phase followed the specific requirements of each experimental condition (immediate, delayed and no feedback).

The testing phase took place for 45 minutes. After a 15 minutes-break, the Posttest learning phase took place for 40 minutes approximately. Students completed a Cued Recall Test consisting of six repeated multiple choice comprehension questions in paper and pencil format.

2.3.6. Measurement

We took three types of measures; online processing (e.g., reading times) during testing, as well as testing and posttest measures, (e.g., comprehension and recall).

Online Processing Measures: Reading times

We computed a number of global and specific time measures in seconds as follows: (a) total time on task for each condition, (b) time reading the text initially, (c) total time re-reading the text and finally, only for those conditions where students received formative feedback, we calculated c) the mean time reading feedback in attempts one and two. In the case of our no feedback condition, we were unable to extract reliable re-reading times due to a system error therefore they are not being reported in the present paper.

Response Certitude Estimates

During the testing phase, response certitude was assessed utilizing a response certitude scale similar to that one used by Kealy and Ritzhaupt (2010). This variable aimed at gathering students´ certitude of response for both, first and second attempt to answer a given question. Students´ response certitude ratings were collected by a horizontal scale graded into 6 percentage tick-marks. The statement, *How confident are you that you answered to this question correctly*, appeared in the upper part of the screen. Right below the statement, the following options were labeled 0%, 25%, 50%, 75% and 100%.

For the first attempt, we computed the means (total score divided by the number of questions) of the response certitude estimates for all of the 12 multiple- choice questions in all of the conditions. For the second attempt, we selected and computed the means by dividing student´s total scores by the number of wrong answers in the first attempt in all of the conditions.

Response Certitude Accuracy

We computed the Goodman and Kruskal´s gamma correlations between the students´ performance in the multiple-choice questions and the students´ certitude of response to each of the questions as a measure of metacognitive accuracy. We expected to analyze the direction of the potential association between our dichotomous variable (success vs fail) when students answered each of our comprehension questions (attempt 1 and 2) and the students´ estimation of certitude in each of the questions. To this, we computed the number of right (1) and wrongs (2) for each question as well as the number of response certitude estimations students selected (0% to 100%). Please note that those results referring to attempt 1 corresponded to the complete sample of students responding to the 12 comprehension questions in all of our experimental conditions, whereas those of attempt 2 corresponded to the questions answered after having failed in attempt 1 (more or less a 50% out of the total).

Testing Phase

Corrective Probability

We computed the conditional probability that an incorrect response would be corrected in the second attempt. We computed a formula as follows, (a) number of right answers on the second attempt and (b) number of errors on the first attempt for each of the conditions: *RA 2nd attempt x 100 /1st attempt errors*. In the case of the no feedback condition, a detailed analysis of the data helped us discern and compute only those scores corresponding to a second attempt to a question completed only after a previous mistake in a first attempt to answer the twelve comprehension questions.

Posttest phase

Multiple-Choice Cued-Recall Test

Students completed a cued-recall multiple choice test consisting of 6 repeated comprehension questions students had been previously exposed during the testing phase. We defined this score as the percentage of rightness in students´ responses as follows: *Number of right answers * 100/number of questions.*

2.4 Results

We performed a set of analyses of variance (ANOVAs) for the independent variable Timing of feedback (immediate vs. delayed feedback) compared to *no feedback.* Additionally, we performed a set of t-test analyses focusing on the specific effects of timing of feedback (immediate vs delayed). Analyses were conducted for our dependent measures: (a) online question-answering and feedback processing, (b) a metacognitive measure of certitude of response and c) comprehension and cued-recall. Three subjects were extracted from the sample due to a system error when analyzing data referring to the metacognitive measure of certitude of response. Post-hoc analyses were performed using Bonferroni´s statistical correction. We explain our results in accordance to those capturing the online and feedback processing (e.g., reading times), students´ corrective probability and cued-recall.

2.4.1. Online question-answering and feedback processing measures

With respect to *total time on task*, an ANOVA reported that those students in the no feedback condition (M = 2019.57; SD = 537.31) spent significantly more time completing the experimental task compared to their counterparts in both, immediate (M = 1603.31; SD

= 468.47) and delayed feedback conditions (*M* = 1677.00; *SD* = 510.57), F (2,104) = 6. 76, $p < .00$, *partial* $\eta^2 = .11$. Independent sample post-hoc *t-test* analysis revealed no significant differences in-between the immediate and delayed conditions, t (70) = - 0.63, $p > .05$. However, *t-test* analysis confirmed significant differences between the no feedback and both, immediate, t (68) = 3.45, $p < .00$ and delayed feedback conditions, t (70) = 2.77, $p < .00$, respectively (see Table 1). An ANOVA analysis on the *initial reading time* of the text reported that students in the no feedback condition read significantly more (*M* = 1269.00; *SD* = 416.74) compared to those students receiving immediate (*M* = 888.91; *SD* = 306.47) or delayed feedback (*M* = 886.03; *SD* = 346.16), F (2,104) = 13.29, $p < .00$, *partial* $\eta^2 = .20$. Regarding both feedback groups, independent sample *t-test* analysis revealed no significant differences in-between the immediate and delayed conditions, t (70) = 0.3, $p > .05$. However, further *t-test* analysis confirmed significant differences between the no feedback group and both, the immediate, t (68) = 4.34, $p < .00$, and the delayed feedback conditions, t (70) = 4.25, $p < .00$, respectively (see Table 1).

Analysis regarding our third no feedback condition are not reported. For our no feedback condition, we were unable to extract reliable re-reading times due to a system error therefore they are not being reported in the present result section. For the mean time reading feedback, an independent sample t-test performed for the feedback conditions, revealed that those students receiving delayed feedback spent significantly more time reading feedback (*M* = 11.71; *SD* = 6.74) compared to those students who received feedback on a question per questions basis (*M* = 5.46; *SD* = 3.23), t (70) = - 4.9, $p < .00$. Similar results were found on the second attempt, right after committing a mistake. Findings revealed that students in the delayed feedback condition (*M* = 18.02; *SD* = 9.79) spent remarkably more time reading feedback compared to their counterparts in the immediate feedback condition, (*M* = 8.70; *SD* = 5.28), t (70) = - 4.98, $p < .00$ (see Table 1).

Table 1.

Means and standard deviations (in parenthesis) of the online-processing measures in the testing phase by condition†

	No feedback	Immediate Feedback	Delayed Feedback
Total time on-task	2019.57(537.31)	1603.31(468.47)	1677.00(510.57)
Time on initial reading	1269.00(416.74)	888.91(306.47)	886.03(346.16)
Total time re-reading the text		168.05(53.47)	187.81(119.02)
Mean time reading feedback			
1st attempt		5.46(3.23)	11.71(6.74)
2nd attempt		8.70(5.28)	18.02(9.79)

Note. † Time in seconds per condition; No feedback, immediate feedback and delayed feedback.

2.4.2. Students´ Certitude Estimates

In order to explore the differences in the students´ certitude estimates between attempt one and attempt two across conditions, a two-way ANOVA was conducted with Condition as a between-groups variable (no feedback, immediate and delayed) and attempt (one and two) as a within-subjects variable. Data reported that there was a significant interaction effect in-between attempts one and two and condition, F (2, 101) = 12.63, p < .05 *partial* η^2 = .05. Overall, students in the no feedback condition (M = 3.37; SD = .09) reported a trend towards significance with higher RC estimates compared to students in the delayed condition (M = 3.05; SD = .09) (F (2,101) = 2.96 $p \leq .05$, *partial* η^2 = *.05*). No further significant differences were found in between conditions. Post-hoc comparisons reported significant differences in-between attempt one and two for the no feedback (p <.00) and delayed (p <.05) conditions (F (2,101) = 8.38 $p < .00$, *partial* η^2 = .14). Significant differences found on the second attempt indicated that students in the no feedback condition (M = 3.66; SD = .62) reported higher RC estimates over both the delayed (M = 3.15; SD = .63) and the immediate condition (M = 3.09; SD = .63). No significant differences were found in attempt one (see **Table 2**). Further intra *t-test* for the no feedback condition confirmed significant differences in-between attempt one and two, t

(33) = - 6.99, $p < .00$. Similarly, significant differences were found for the delayed condition in-between attempt one ($M = 2.95$; $SD = 0.59$) and two ($M = 3.15$; $SD = 0.63$), t (34) = -2.219, $p < .03$. Analysis for the immediate condition reported no significant differences in-between attempt one ($M = 3.20$; $SD = 0, 60$) and two ($M = 3.09$; $SD = 0.63$), t (34) = .91, $p > .05$.

In order to determine the association between success answering the comprehension questions and students´ certitude of response, we run a Goodman and Kruskal´s gamma coefficient. For the first attempt, in all of the experimental conditions the gamma value was positive and significant; no feedback ($G = .338$, $p < .0005$), immediate ($G = .307$, $p < .0005$) and delayed ($G = .344$, $p < .0005$). Please note that the results referring to attempt one correspond to those from all the sample of students responding to the twelve comprehension questions in all of our experimental conditions whereas those of attempt two refer to those questions answered after having failed in attempt one (50% average). For those questions answered in the second attempt, the gamma value was positive and significant only for the immediate condition ($G = .275$, $p < .012$) whereas the no feedback ($G = .127$, $p > .293$) and delayed condition ($G = - .136$, $p > .213$) reported no significant association.

Table 2

Means and standard deviations (in parenthesis) of students´ certitude estimates in the posttest phase by condition†

	No feedback	Immediate Feedback	Delayed Feedback
Attempt 1	3.07 (0.65)	3.20 (0.60)	2.95 (0.59)
Attempt 2	3.66 (0.62)	3.09 (0.63)	3.15 (0.63)

Note. † Certitude of response score max =5 per condition; No feedback, immediate feedback and delayed feedback

2.4.3. Corrective probability during question-answering on the Testing Phase

Regarding our measure Corrective Probability, an ANOVA reported a significant effect of feedback over no feedback, F (2,101) = 10.58, $p < .05$, *partial* $\eta^2 = 0.17$. Post-hoc comparisons showed that receiving both, immediate (M = 61.96; SD = 20.50) and delayed feedback (M = 60.60; SD = 22.48) overcame not receiving feedback (M = 41. 16; SD = 19.75). In addition, results of an independent sample *t-test* revealed no significant differences in-between the immediate (M = 61.96; SD = 20.50) and delayed conditions (M = 60.60; SD = 22.48), t (68) = 0.26, $p > .05$. Further *t-test* analyses confirmed the significant differences between the no feedback and both immediate, t (67) = - 4.20, $p < .05$, and delayed feedback conditions, t (67) = - 3.81, $p < .05$.

2.2.4. Cued-recall measure on the Posttest Phase

Our Cued-Recall Multiple Choice Test measure reported a rate of item difficulty ranging from (M = 0.66; SD = 0.47) to (M = 0.82; SD = 0.38). To this, an ANOVA analysis of the percentage of correct answers showed that there was a significant effect of Timing of feedback over no feedback, F (2,101) = 4.47, $p < .05$ *partial* $\eta^2 = 0.08$. Post-hoc comparisons reported that those students receiving immediate feedback (M = 80. 95; SD = 16. 24) overcame students in the no feedback condition (M = 65. 68; SD = 26. 57). When comparing immediate and delayed feedback effects, an independent sample *t-test* confirmed no significant differences in-between both conditions, immediate (M = 80.95; SD = 16.24) and delayed feedback (M = 74.76; SD =19.95), t (68) = 1.42, $p > .05$. Similarly results confirmed no significant differences between the no feedback and delayed feedback conditions, t (67) = - 1.60, $p > .05$. However, further *t-test* analyses confirmed the significant differences between the no feedback and the immediate condition, t (67) = - 2.88, $p < .05$ (see **Table 3**).

Table 3

Means and standard deviations (in parenthesis) of the cued-recall scores in the posttest phase by condition†

	No feedback	**Immediate Feedback**	**Delayed Feedback**
Multiple-Choice Cued Recall	65.68 (26.57)	80.95 (16.24)	74.76 (19.95)

Note. † Cued-Recall score max. = 100 per condition; No feedback, immediate, delayed feedback

2.5. DISCUSSION

The study aims at examining the effects of immediate and delayed feedback on question answering processes and performance. Therefore, we compared students receiving feedback with the two schedules, with students who did not receive feedback, while they performed a two-attempt question-answering task. Our first hypothesis was that students in the two feedback conditions would read the text initially quicker than those in the no-feedback conditions. Our results confirm the hypothesis. Knowing that feedback would be provided after the first attempt, and that students would have a second attempt to correct their mistakes, made them spend less time than knowing that no feedback would be provided. Vidal-Abarca and colleagues (Ferrer et al., 2017; Higgs et al., 2017) report a result that helps us to interpret our findings. They found that knowing that a text would be available when answering question made students spend less time when reading the text initially than knowing that it would be unavailable. Ferrer et al., (2017) argue that this effect can be interpreted as a decrease in the readers' standards of coherence when reading on the available condition. Students in the available condition likely thought that after reading the questions they would have more information on what and how to read, whereas this possibility would not exist for students in the unavailable condition. Feedback may have a similar effect. Students in the two feedback conditions might think that feedback would provide them with information to make the decision to either reread the text or not.

Therefore, students would adapt their processing to this situational variable.

The second hypothesis predicted that feedback would enhance the students' accuracy of their response certitude judgments on the second attempt in comparison to no-feedback students. Our results only support this prediction for the immediate feedback condition. Our results are similar to those found by Butler et al. (2008). Students in the immediate condition were better able to discriminate between correct and incorrect answers. Students in this group decreased their response certitude judgments from the first to the second attempt, at the same time that they increased the proportion of correct responses. The close association between their response certitude judgment and the feedback due to the immediacy of feedback made them adjust their judgments to their responses. However, this association was not so close for the feedback delayed condition, quite likely because the time elapsed between their judgment and the feedback. In fact, these students increased their performance on the second attempt over the first attempt, but their judgments also increased significantly over the first attempt. It may explain the negative correlation between response certitude and performance on the second attempt. Therefore, it seems that the immediacy of feedback after response certitude judgment is needed for the improvement of this metacognitive process. It is not surprising the lack of improvement of this metacognitive variable for the no-feedback students, due to the absence of external feedback. These students decreased their performance on the second attempt regarding the first attempt, at the same time that they significantly increase their response certitude judgments. Therefore, the absence of feedback may have a negative impact not only for performance but also for the improvement of monitoring accuracy.

The third hypothesis predicted that students in the delayed feedback condition would search the text longer than those in the immediate condition. We reasoned that the connection between the question and the text information relevant for an answer would be weaker in the delayed feedback condition in comparison to the immediate condition, which would increase searching. However, our results do not support the hypothesis since no differences between the two conditions were apparent. A possible explanation for these results lies in the format of the questions. Multiple-choice questions provide readers with more retrieval cues than open-ended questions (Little, Bjork, Bjork & Angello, 2012). These cues may discourage readers to search the text. Further, sometimes students have difficulties choosing between two equally attractive choices. So, when feedback indicates that their decision was wrong, they just select the alternative option, instead of searching the text. In fact, Ferrer et al. (2017) found that students decided to search the text to answer multiple-choice questions less often than they did to answer open-ended questions. Unfortunately, the software used in this experiment did not record rereading times question by question, which prevents us from having evidence of search decisions. In spite of that, it sounds reasonable to think that the use of multiple-choice questions may explain why students did not search the text longer under the delayed feedback condition in comparison to the immediate condition.

The fourth hypothesis predicted that students from the two feedback groups would improve their performance for the second attempt and the posttest measures in comparison to no-feedback students. We partially confirmed our hypothesis. Students who received feedback outperformed no feedback student on the second attempt as well as on the repeated multiple-choice questions on the posttest. These results are not surprising since feedback simply informed students about the correctness of their answers on the first attempt (KR feedback) and about the correct response after the second attempt (KCR feedback). In the first case, KR feedback reduced the options among the choices, whereas in the second feedback informed students about the correct response. Both results are consistent with those reported in the literature (van der Kleij et al., 2015). However, no significant differences among groups were apparent for the cued-recall measures on the posttest. We found that total time on task was significantly higher for students who did not receive feedback in comparison to participants in the two feedback groups. It sounds logical that this extra-time would be employed for reading the text, either initially or when searching the text, which may have compensated for the lack of feedback on a cued-recall measurement. A study by Moos (2011) on question-answering from a text provides additional data to explain our results. This author found that undergraduate students who received KR feedback scored lower on a posttest that another group who answered questions without feedback. Moos recorded verbal protocols while students read and answered the questions. He attributed the higher scores of students in the no-feedback group to the fact that they used more strategies that students in the feedback condition. For example, after being informed that their response was correct students stopped interacting with the text. In other words, students in the feedback condition may have decreased, or even discontinued, their use of some strategies useful for learning text information. Therefore, KR feedback utility limits to knowing whether the response is correct or incorrect, but it does not help to understand the right response (van der Kleij, et al. 2015). Furthermore, it may inhibit learning by reducing the used of strategic behavior.

The fifth hypothesis predicted that students in the delayed feedback would outperform students in the immediate condition on the cued-recall test. We disconfirmed this hypothesis as no differences were found among the two groups. We based our prediction on the assumption that students in the delayed feedback group would search the text longer than those in the immediate feedback group. As we said above, this assumption was wrong, maybe because the format of the questions discouraged student to search the text.

The findings from this study taken together have important implications for the theory and practice of corrective formative feedback for question-answering activities. First, feedback reduces the time of text processing for both, the initial reading of the text and the text search for answering the questions, a result not reported in the literature.

Second, formative feedback may contribute to increasing the students' metacognitive accuracy of their responses, although this effect limits to immediate feedback, a result reported for facts learning, but not for question-answering activities. Third, corrective delayed feedback provides no advantages over immediate corrective feedback and some disadvantages, at least with multiple-choice questions. Fourth, the use of corrective formative feedback for question-answering has some limitations. Since it focuses on informing students about the correctness of their response (KR) or, additionally, about the correct response (KCR), it does not directly contribute to increasing the students' strategic behaviour and deep understanding. These findings clarify the effects of corrective feedback on question-answering settings.

The present study also has a number of limitations that present opportunities for future research. First, we used a single text and a limited rank of question. Using only one text limits the possibility of generalizing our results. Using a limited rank of questions affects the students' processing. Actually, Ferrer et al. (2017) found that students tended to search more often when answering text-based questions, than when answering inferential questions, as we did in our study. Thus, including text-based questions would allow us to examine the effects of corrective feedback on the students' strategic processing after failure and success.

Second, in this study, we controlled to some extent prior knowledge by using non-academic expository texts. Individual differences in prior knowledge, however, are an important factor to consider (Kendeou & O'Brien, 2016). For example, empirical evidence in the area of mathematics suggests that primary-school and university students with little prior knowledge benefit from feedback interventions more than participants with higher prior knowledge (e.g., Fyfe, Rittle-Johnson, & DeCaro, 2012; Krause, Stark & Mandl, 2009). Thus, more research is needed to understand how prior knowledge influences the students' processes of formative feedback. Third, the system we used for presenting materials did not record some important variables that are informative about the students' strategic behaviour (e.g., search decisions or time spent on searching the text question by question). Fortunately, now we have such a system (Vidal-Abarca et al., 2018). Some of the issues unanswered in this study could now be answered with this new tool.

In conclusion, the findings from the present study shed light on how corrective formative feedback affects the students‟ processing and performance on a question-answering environment. The study documents how feedback impacts the students‟ reading and the metacognitive accuracy of their responses, as well as the differences between immediate and delayed feedback on these effects. It also shows the limitations of formative feedback in terms of performance. These findings provide useful information for the design of computer-based systems aimed at teaching conceptual information.

2.6. REFERENCES.

Anderson, M., & Thiede, K. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica*, *128*, 110–118.

Anmarkrud, Ø., McCrudden, M. T., Bråten, I., & Strømsø, H. I. (2013). Task-oriented reading of multiple documents: Online comprehension processes and offline products. *Instructional Science, 41*, 873-894.

Attali, Y. (2015). Effects of multiple-try feedback and question type during mathematics problem solving on performance in similar problems. *Computers & Education*, *86*, 260-267.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991).The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*, 213–238.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low confidence correct responses. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*, 918-928.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*(4), 273.

Cepeda, N.J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis. *Psychological Bulletin*, *132*(3), 354–380. doi: 10.1037/0033-2909.132.3.354

Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences, 21*(2), 201-205.

Conole, G. (2016). MOOCs as disruptive technologies: strategies for enhancing the learner experience and quality of MOOCs. *RED- Revista de Educación a Distancia, 50.* Retrieved from http://www.um.es/ead/red/50/conole.pdf

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition: A textbook of cognition, educational, life span, and applied psychology*. Thousand Oaks, CA: Sage.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, *27*(3), 209-226. doi.org/10.1111/j.1745-3984.1990.tb00744.x

Felker, D. B., & Dapra, R. A. (1975). Effects of question type and question placement on problem-solving ability from prose material. *Journal of Educational Psychology, 67*(3), 380-384.

Ferrer, A., Vidal-Abarca, E., Serrano, M. Á., & Gilabert, R. (2017). Impact of text availability and question format on reading comprehension processes. *Contemporary Educational Psychology, 51,* 404-415.

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, *104*(4), 1094-1108.

Grasser, A. C., & Murachver, T. (1985). Symbolic procedures of question answering. In A.C.Graesser & J.B. Black (Eds). *The psychology of questions* (pp. 15-88). Hillsdale, NJ: Erlbaum.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, *56*(2), 212-242.

Hattie, J., Gan, M. (2011).Instruction based on feedback. In Alexander, P., Mayer, R. E. (Eds.), *Handbook of research on learning and instruction* (pp.249–271). New York, NY: Routledge.

Hattie, J., Timperley, H. (2007).The power of feedback. *Review of Educational Research, 77*, 81–112. doi:10.3102/003465430298487

Higgs, K., Magliano, J. P., Vidal-Abarca, E., Martínez, T., & McNamara, D. S. (2017). Bridging Skill and Task-Oriented Reading. *Discourse Processes, 54*(1), 19-39.

Kapp, F., Proske, A., Narciss, S., & Körndle, H. (2015). Distributing vs. Blocking Learning Questions in a Web-Based Learning Environment. *Journal of Educational Computing Research*, *51*(4), 397-416.

Keith W. Thiede, Mary C. M. Anderson, and David Therriault. (2003). Accuracy of Metacognitive Monitoring Affects Learning of Texts. *Journal of Educational Psychology*, *95*(1), 66–73.

Kealy, W. A. & Ritzhaupt, A. D. (2010). Assessment certitude as a feedback strategy for learners‟ constructed responses. *Journal of Educational Computing Research*, *43*(1), 25-45. doi:10.2190/EC.43.1.c

Krause, U.-M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction, 19*(2), 158 –170. doi:10.1016/j.learninstruc.2008.03.003

Kulhavy, R. W., Stock, W. A. (1989).Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*,279–308. doi:10.1007/BF01320096

Little, J.L., Bjork E. L., Bjork, R.A., & Angello, G (2012). Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychological Sciences*, *23*(11), 1337-44. doi: 10.1177/0956797612443370.

Llorens, A.C., Vidal-Abarca, E., & Cerdán, R. (2016). Formative feedback to transfer self-regulation of task-oriented reading strategies. *Journal of Computer Assisted Learning*, *32*(4), 314-331.

Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education*, *95*, 85-98.

Moos, D.C. (2011). Self-regulated learning and externally generated feedback with hypermedia. *Journal of Educational Computing Research, 44*(3), 265-297.

Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Erlbaum Associates.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J.M. Spector, M.D. Merrill, J.J.G. van Merrienboer, & M.P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology (3rd ed.). pp. 125-144)*. Mahaw, NJ: Lawrence Erlbaum Associates.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review*, *23*, 7-26. Retrieved from https://files.eric.ed.gov/fulltext/EJ1013726.pdf

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Ed.), *Instructional design for multimedia learning* (pp. 181–195). Munster, New York: Waxmann.

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, *16*(4), 310-322.

Narciss, S., Koerndle, H. & Proske, A. (2013). Challenges of investigating metacognitive tool use and effects in (rich) web-based learning environments. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 243-260). New York: Springer Science & Business Media, LLC. DOI: 10.1007/978-1-4419-5546-3_17

Narciss, S., Sosnovsky, S., Schnaubert, L., Andres, E., Eichelmann, A., Goguadze, G., E., Melis (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, 71, 56–76. doi:10.1016/j.compedu.2013.09.011

Ness, M. (2011). Explicit reading comprehension instruction in elementary classrooms: Teacher use of reading comprehension strategies, *Journal of Research in Childhood Education*, *25*(1), 98-117. doi: 10.1080/02568543.2010.531076

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, *25*(4), 399–438.

Pressley, M., Wharton-McDonald, R., Mistretta, J., & Echevarria, M. (1998). Literacy instruction in ten fourth- and fifth-grade classrooms in upstate New York. *Scientific Studies of Reading*, *2*, 159-194.

Roelle, J., & Berthold, K. (2017). Effects of incorporating retrieval into learning tasks: The complexity of the tasks matters. *Learning and Instruction, 49*, 142–156. doi:10.1016/j.learninstruc.2017.01.008

Rouet, J.F., Britt M.A., & Durik A. M. (2017). RESOLV: Readers' Representation of Reading Contexts and Tasks, *Educational Psychologist*, *52*(3), 200-215. doi:10.1080/00461520.2017.1329015

Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, *58*, 56-61.

Rupp, A.A., Ferne. T., & Choi. H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Language Testing*, 23, 441-474.

Schooler, L. J., & Anderson, J. R. The disruptive potential of immediate feedback. In *Proceedings of the twelfth annual conference of the Cognitive Science Society* (pp. 702-708), (1990, July).

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189. doi:10.3102/0034654307313795

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, *33*(1), 1-19.

Van den Broek P.W. & Helder A. (2017). Cognitive Processes in Discourse Comprehension: Passive Processes, Reader-Initiated Processes, and Evolving Mental Representations, *Discourse Processes 54*(5-6), 360-372.

Van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, *58*, 263–272.doi:10.1016/j.compedu.2011.07.020

Van der Kleij, F. M., Feskens, R.C.W., Eggen, T.J.H.M. (2015) Effects of Feedback in a Computer-Based Learning Environment on Students‟ Learning Outcomes. *Review of Educational Research*, *85*(4), 475-511.

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

Vidal-Abarca, E., Martinez, T., Serrano, M-A., Gil, L., Maña, A., Mañez, I., García, A., Gilabert, R., Cerdan, R., Rubio, A., Candel, C., Ferrer, A., Avila, V., Grimaldo, F., & Garcia, D. Read&Learn: A Research Tool to Record Online Processing While Learning. *Paper presented at the 28th Annual Meeting of the Society for Text & Discourse. Brighton (UK),* (July, 2018).

Webb, J. M., Stock, W. A., & McCarthy, M. T. (1994). The effects of feedback timing on learning facts: The role of response confidence. *Contemporary Educational Psychology*, *19*(3), 251-265.

CHAPTER THREE

STUDY 2: EFFECTS OF TIMING AND LEVEL OF INTEGRATION OF ELABORATIVE FEEDBACK MESSAGE FOR CONCEPT LEARNING IN QUESTION-ANSWERING SETTINGS

3.1. ABSTRACT

The study analyzes the effects of a reading procedure that delivered immediate, delayed summative or delayed integrative elaborative feedback on question-answering performance and concept learning from text. The level of integration of the elaborative feedback message was manipulated across conditions. One hundred thirty-three secondary-school students read a scientific text and answered twelve high and low-level multiple-choice questions on a computer-based system. A day later, students completed a learning test consisting of twenty paraphrased low and high-level questions. Neither the immediate nor the delayed feedback conditions influenced students' overall learning. However, individual differences were identified. While students with higher levels of previous background knowledge in the delayed integrative feedback condition enhanced their performance on both low and high-level questions, those students with lower levels of prior background knowledge in the immediate condition outperformed their counterparts only on the low-level posttest questions. Implications for learning of such a procedure are further discussed.

Keywords: formative feedback, elaborative feedback, concept learning, feedback timing, question-answering, reading performance

3.2. INTRODUCTION

Secondary school students are often required to read a scientific text to answer learning questions on a computer. In this type of learning scenario, Computer-Based Systems (CBS) bring significant advantages to the learning experience, like the possibility to deliver Formative Feedback (FF) while students interact with the text and questions (Vidal-Abarca et al., 2017; Shute, 2008). Not only this, but feedback can be delivered under different timings; for example, a corrective message can be presented in an immediate (i.e., question per question) or delayed fashion (i.e., after the completion of a certain number of questions). Candel et al., (2017) found that providing immediate feedback was the most effective procedure for both the correction of inferential errors and final concept learning.

To a larger extent, Formative Feedback (FF) may not only be aimed at correcting errors, but can also contain the elaborative material students need to build a coherent mental model from text, which may help students to either recall terms in a text, or elaborate on more complex forms of knowledge like inference-making (Vidal-Abarca et al., 2017). Although some authors seem to disagree on what type of feedback timing is more efficient for concept learning (Butler, Karpicke & Roediger, 2007; Newman, Williams & Hiller, 1974; Shute, 2008), delayed forms of feedback have been pointed out as the most beneficial types for the learning of verbal material, particularly when tasks require low processing demands such as pair-association or word-retention (McDaniel, Thomas, Agarwal, McDermott & Roediger, 2013; McDaniel, Fadler & Pashler, 2013; Butler, Karpicke & Roediger, 2007; More, 1969). The argument is that providing delayed feedback messages (i.e., after the completion of a set of questions) allows students to revise and elaborate on material in the form of a new re-study opportunity (Butler, Godbole & Marsh, 2013; Butler, Marsch & Godbole, 2013; Kapp, Proske, Narciss &

Körndle, 2015; van der Kleij, et al., 2015).

In the present work, we define delayed feedback like the one that is provided in a summative manner. That is, in the form of independent messages delivered right after responding to a certain number of tasks, seconds, minutes or even days (for a review see Kulik & Kulik, 1988). Nevertheless, different types of delayed elaborative feedback are still very possible; for example, delivering a feedback message after the completion of a certain number of questions in the form of an integrative elaborative message, which provides students with the relation of ideas they need to build a coherent mental model from the text, maybe decisive for high-level concept learning. In fact, as van der Klij et al., (2015) claim in their review, complex forms of elaborated feedback messages seem to be efficient for high-level learning tasks. Such high level learning tasks certainly require students to integrate complex and distant forms of information as well as elaborations and inferences, which have shown to be strong precursors for deep learning (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, 2009; Vidal-Abarca, Gilabert, & Rouet, 1998; Vidal-Abarca et al., 1996). Unfortunately, the current body of research lacks enough literature analyzing how the level of elaboration and integration of a delayed feedback message might support concept learning when students answer questions with an available text. Since a delayed feedback message leaves enough room for experimenters to include different content relations (i.e., prompts, definitions) (Shute, 2008), it is of importance to mention that its potential effects can vary depending on not only on its content but on the type of learning questions students are required to complete (i.e., low or high level) (Butler & Roediger, 2008; Sinha & Lewis, 2015; Van der Kleij, 2015).

As such, an elaborative feedback message that provides specific explanations on the correct answers could be given on a question per question immediate basis; as in the form of delayed summative feedback or in a more practical delayed manner; in the form of a unique elaborative integrative message aiming at facilitating the integration of the relevant ideas from a text for a group of four learning questions conceptually related. In this context, the present study aims at analyzing how feedback timing (i.e., either question by question or delayed after the completion of four questions) and the level of integration of the elaborative feedback message (i.e., provided on an immediate question per question compared to a delayed summative or delayed integrative basis) might impact question-answering performance and content learning from text. A second objective is to deepen the extent to which type of question (i.e., low or high level), and previous background knowledge (PBK) might influence the potential relationship between students´ performance and concept learning. Previous research shows that the type of question is directly related to different effects on learning, for example, while low-level questions are more likely to ease the recall of literal information from text amongst students with low-levels of prior background knowledge, high-level questions tend to induce students with higher level of prior knowledge into a more relational and elaborated pattern of constructive learning from text (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, 2009; McNamara & Kintch, 1996).In order to provide the reader with a complete view of these effects, in the next section, we firstly provide a synthesis of some of the most outstanding findings within the frame of feedback timing and question-answering.

Secondly, relevant literature within comprehension and learning from text elaborates on the important relationship between the level of integration of the elaborative content of the feedback message and the important role of the type of question (low and high level) for concept learning form a text. Finally, the paper explores and discusses the potential impact of the students´ individual differences on previous background knowledge on reading performance and concept learning, within the frame of Formative Feedback (FF) question-answering procedures.

3.2.1. Feedback Timing in Question-Answering with an Available Text

Computer-based feedback procedures provide educators with a wide range of interactive tools that enrich the students´ learning experience (Dunlosky & Rawson, 2015). Within the context of concept learning, question-answering seems to be one of the most effective ways to tackle concept learning (i.e., factual or inferential); however, how feedback is distributed throughout the learning environment may strongly determine how students structure and process learning content (Kapp et al., 2015; Van der Kleij, Feskens, & Eggen, 2015). For example, Knowledge of Response (KR) and Knowledge of Correct Response (KCR) provide information on errors and correct answers, while Elaborative Feedback (EF) focuses on teaching specific knowledge by providing additional explanations that may be useful for learning (Shute, 2008). Classically, Formative Feedback (FF) has been delivered by following an immediate (i.e., question per question) or delayed distribution (i.e., after the completion of a set of questions) (Butler & Roediger, 2008; Kulik & Kulik, 1988; Shute, 2008; Mory, 2004; Van der Kleij, et al., 2015). In this regard, while Formative Feedback (FF) provided in an immediate manner has proven to reinforce the correction of erroneous responses and the retention of verbal material (Brosvic, Dihoff, & Epstein, 2003; Brosvic, Epstein, Cook & Dihoff, 2005; Butler & Roediger, 2008; Candel, et al., 2017; Eipstein, Dihoff & Cook, 2006), delayed feedback seems to lead to better posttest learning, as mostly reflected by learning tasks implying the cued-recall of the correct or incorrect answers (Butler, et al., 2007; Metcalfe, et al, 2009; Shute, 2008).

In the majority of these studies, immediate feedback is delivered question after question, while delayed feedback not only contains similar information as in its immediate form but it is provided in a summative manner; that is, independent messages to each question are delivered after responding to a certain number of questions (for a review see

Kulik & Kulik, 1988). (Bangert-Drowns, Kulik, Kulik & Morgan, 1991; Narciss, 2013; McDaniel, et al., 2013; Butler, et al., 2007; More, 1969; Kulik & Kulik, 1988). Several theoretical models have highlighted the importance of this sequential interactivity between questions and feedback for self-regulation and learning (Bangert-Drowns, Kulik, Kulick & Morgan, 1991; Kapp & Körndle, 2011; Kapp et al. 2015; Narciss, 2013). In fact, researchers state that delaying a feedback message permits students to engage in a new re-study opportunity that allows for more conscious monitoring of learning, an effect commonly referred to as *the Interference Perseveration Hypothesis* (Kulhavy & Anderson 1972).

According to these authors, when feedback is provided after some delay, students are more likely to cognitively dismiss the wrong answers from memory, while the access to the correct ones seems to ease. This type of spaced presentation of the learning material seems to be highly beneficial for learning; especially when students deal with question-answering tasks that require a high-level type of processing (Kapp et al., 2015; Van der Kleij et al., 2015). In this line, Candel et al., (2017) (*manuscript in preparation*) tested the effectiveness of a feedback procedure that required students to read an expository text and answer a set of high-level learning questions. A Computer-Based System provided immediate or delayed knowledge of response feedback (KR) right after the completion of four questions, compared to a non-feedback procedure in which students were prompted to re-read the text. Access to the text was provided only after failing. Knowledge of the correct response feedback (KCR) was provided only when students answered correctly. Results reported that receiving either immediate or delayed formative feedback benefited students´ corrective efficiency compared to non-feedback.

However, students in the immediate feedback condition were more certain in their responses during question-answering and obtained higher scores in a final repeated multiple-choice test, compared to their counterparts in the rest of the conditions. Authors also tested students on an open-ended recall task; nevertheless, no significant learning effects were apparent. The authors concluded that both immediate and delayed feedback might have provided students with valuable clues for error-correction during question-answering with the high-level learning questions, but might have not been enough to induce students into a deeper processing of the learning content, which would have resulted in a more relational mental model from text; a gap that aims to be filled in the current study.

In this regard, some studies have shown how the delivery of complex feedback messages facilitate text-based learning when focusing on specific instruction to improve students´ identification of relevant information (Máñez, Vidal-Abarca, Martínez & Kendeou, 2017; Vidal-Abarca, Martínez, Ferrer & García, 2017). Even though some research on content feedback has already analyzed its effects on learning according to its distribution and the inferential nature of the items being used (i.e., low versus high order) (Butler et al., 2012; Golke et al., 2015; Shute 2008; McDaniel & Fisher, 1991; Van Der Kleij, 2012), no study has yet delved into how a different level of integration of the elaborative content feedback message may impact concept learning when provided right immediately (i.e., question by question) or after some delay (i.e., after the completion of four related questions). In this light, as the present study intends, providing an immediate feedback message facilitating a set of cues to infer the correct answer for a group of low and high-level inferential questions might influence learning positively. A similar effect may be expected from a set of delayed elaborative feedback messages provided in a sequential manner.

However, a remarkable benefit could be expected by providing a unique delayed integrated message as in the form of interconnected information that would facilitate the learning of conceptual knowledge for the same group of questions. This procedure could be particularly beneficial since our questions require different levels of inferential activity from students. As such, the next section deepens into this issue. Firstly, we will review the relevant literature in regard to the importance of the level of integration and elaboration of delayed content feedback when students are required to deal with low and high-level learning questions with an available text. Secondly, we will provide the first step into the role of individual differences in previous knowledge, and its potential interaction with the type of question, within text-based question-answering feedback procedures.

3.2.2. Level of Integration of the Elaborative Feedback for Concept Learning

Elaborative Feedback (EF) has been reported to be the most effective type for concept learning when provided after some delay (Van der Kleij et al., 2015). Still, the level of integration and elaboration of an individual feedback message may vary considerably (Shute, 2008). While Knowledge of Response (KR) and Knowledge of Correct Response (KCR) have proven to facilitate error correction and the learning of the right answers (Candel at al., 2017; Butler et al., 2008; Moreno & Valdez, 2005; More, 1969), Elaborative Feedback (EF) has shown to be particularly beneficial in question-answering at both, the conceptual and procedural level of learning (Butler, Godbole & Marsh, 2012; Jaehnig & Miller, 2007; Kapp et al., 2015; Llorens et al., 2014; Martínez, et al, 2017; Mañez et al., 2016; Van der Kleij, 2015). In fact, plenty of literature highlights the benefits of those feedback explanations that provide elaborative information on the correct and incorrect answers, particularly when students respond to new or repeated posttest learning questions (Butler et al., 2013; Golke et al., 2015; Martínez, García & Vidal-Abarca, 2017; Kapp et al., 2015; Vidal-Abarca, Martínez, Ferrer & García, 2017). Furthermore, in question-answering with a text available scenarios, EF on the error can, in some cases, be preferred by secondary school students as an alternative to rereading the text (Martínez, García & Vidal-Abarca, 2017).

A potential explanation for this effect may rely on the fact that certain forms of elaborative feedback (EF) focusing on explaining the inferential error allow students to repair inconsistencies by helping them acknowledge and discard the irrelevant information to answer. However, this is a complex process that is mostly based on the correct comprehension and understanding of the individualized feedback message per se. Therefore, an efficient EF needs to be effective to trigger those important comprehension aspects of the learner´s constructive activity, resulting in the efficient acquisition of complex concept knowledge from the text (Vidal-Abarca & SanJosé, 1998; McNamara et al. 1996, McNamara, Jacovina & Varner, *in press*).

Nevertheless, no study has yet analyzed how different levels of elaboration and integration of the EF could influence concept learning during a text-based question-answering procedure, in which delayed elaborative feedback is provided in a summative or in a more practical integrative way, right after the completion of a set of four related questions. As such, an individual immediate and a delayed summative feedback may both contain Knowledge of Correct Responding (KCR) and an elaboration of the correct answer (i.e., Atmospheric pressure can be measured with a mercury barometer); while a delayed integrative feedback message, provided after the completion of four questions, may contain an elaborative message that provides students with the relation of the relevant ideas from the text they need to build a coherent and more inter-related mental model from the learning material (i.e., *Atmospheric pressure can be measured with a mercury barometer (hence the commonly used synonym is barometric pressure), which indicates the height of a column of mercury that exactly balances the weight of the column of atmosphere over the barometer*, example extracted from britannica.com, 2019). To this end, the present work takes a well-known model of the situation perspective which is based on the cognitive Kintsch´ theory of comprehension and learning from text (Kintsch, 1998; McNamara et al., 1996; Cerdán & Vidal-Abarca, 2008; Cerdán et al., 2009). This approach also highlights the important role of the type of task (i.e., questions) as a valuable facilitator for concept learning from the text, since they can be designed to induce a more or less relational type of processing of the content within a text (Cerdán et al., 2009; McNamara et al. 1996; LeBigot & Rouet, 2007). For example, low-level questions requiring the location, recognition and recall of explicit text statements have shown to induce the superficial memorization of the information, while high-level questions, that require the relational processing of distant idea-units within a text, allow students to process content in a more integrative network of ideas (Ozuru et al., 2007).

In this fashion, a delayed integrative feedback message that specifically elaborates on the relevant textual ideas for answering a block of four questions that are conceptually related could facilitate the student´s inferential activity compared to receiving an immediate elaborative feedback message question per question. Similarly, this type of delayed integrative feedback message could specifically be effective when students answer high-level learning questions; which are specifically designed to relate the targeted relevant ideas within the text. As a result, and compared to an immediate or delayed summative manner (i.e., a delayed feedback provided in the form of a set of summative messages for each of the questions), a delayed integrated elaborative feedback would facilitate the relational processing of the information targeted by a group of related questions to a greater extent. On the contrary, an immediate feedback would induce a more isolated processing of the learning material, primarily focused on the processing of the correct responses for each single question and would, in consequence, decrease the processing of the learning material in a more coherent and interconnected manner (Cerdán et al., 2011; Cerdán et al., 2009; McNamara, et al., 1996; Ozuru, Dempsey, & McNamara, 2009; SanJosé, Portolés &Vida-Abarca, 1993). This would be particularly observable for low-level questions since these questions are designed to help the student locate and recall explicit information within a text. Therefore, as we propose in the present study, an effective delayed integrative elaborative feedback is expected to facilitate students´ inferential processing from a text, especially when answering high-level questions.

Thus, following the existing literature on the positive effects of delayed feedback in concept learning (Butler & Roediger, 2008; Dempsey & Wager, 1988; Kulik & Kulik, 1988; Shute, 2008; Surber & Anderson, 1975; Mory, 2004), providing this sort of message in a delayed integrative manner could also serve the function of a re-study episode not only by facilitating the extraction of the relevant ideas from text to answer, but by fulfilling any possible gap that may arise in the students´ level of previous background knowledge.

3.2.3. Previous Background Knowledge and Feedback Instruction

Previous background knowledge (PK) seems to play an important role in feedback effectiveness and students´ learning (Fyfe, Rittle-Johnson & DeCaro, 2012; Narciss, 2008). Nonetheless, apart from one study analyzing the effect of PK and feedback in the context of question-answering (Máñez, Vidal-Abarca, Martínez & Kendeou, 2017), the majority of literature refers to the field of mathematics, physics or genetics (i.e., Fyfe, 2016; Fyfe, Rittle-Johnson, & DeCaro, 2012; Krause, Stark, & Mandl, 2009; Smits, Boon, Sluijsmans, & Van Gog, 2008). In this type of learning scenarios recent literature reports that while an immediate trial-by-trial feedback seems to boost mathematical problem-solving skills over a delayed end-of-test procedure (Fyfe & Rittle-Johnson, 2016), a global procedural feedback is more likely to benefit performance of those students with high-previous knowledge students in genetics (Smits et al., 2008).

Within the context of question-answering scenarios, previous literature in comprehension has highlighted the moderating effects of PK in learning, especially when students answer tasks implying different levels of processing (i.e., low and high-level questions) (McNamara et al., 1996). In this line, Máñez, Vidal-Abarca, Martínez &

Kendeou (2017) found that the relationship between students' prior knowledge and posttest learning was significantly moderated by the number of times students accessed content elaborated feedback during a previous testing phase. In fact, these authors found that those students with low-levels of prior background knowledge, who decided to consult feedback more often, benefited significantly more in comparison to the ones that consulted feedback to a lesser extent. In the light of these results, it seems logical to consider that in addition to Previous Background Knowledge, Timing of Feedback (Immediate, Delayed Summative or Delayed Integrative) may also have a moderation effect on performance and learning. In fact, according to the literature mentioned above on the powerful effects of Delayed form of elaborated feedback on learning (Shute, 2008; Van der Kleij, 2015), a Delayed Integrative Feedback may benefit performance and learning to a larger extent compared to a Delayed Summative or Immediate Feedback. A delayed integrative elaborative feedback would facilitate students´ inferential processing from the text. Additionally, an immediate EF provided right immediately after the completion of each question would be particularly beneficial for those students with lower levels of prior knowledge, since it would allow them to detect and correct wrong inferences in a more sequential manner (i.e., by systematically distributing the content on a question by question basis). This effect would be particularly observable, during both training and posttest phases, in our low-level questions, and that is, questions that require the mere recognition and recall of concepts with minimal inference to identify the correct answer and that have a literal relation to the text (Andre, 1979; Cerdán et al., 2009; Goldman & Durán, 1988; Vidal-Abarca, Gilabert, & Rouet, 1998; Wiley, & Voss, 1999).

On the contrary, students with a higher level of prior knowledge would benefit more from delayed integrative feedback, since it provides a unique elaborative message with the inferential relation of the correct ideas in a more coherent and inter-related manner. Students with higher levels of prior-knowledge would be able to integrate the new incoming information into a mental model from the text in a more coherent and practical way, which would positively influence concept learning, especially in the high-level learning questions at posttest.

3.2.4. The current study

The main goal of the present study was to analyze the effects of a computer-based question-answering procedure that provides elaborative content feedback messages which vary in their level of integration of the elaborative feedback message according to different timings as follows: immediate, delayed summative and delayed integrative. Students read a text in the science topic *Atmospheric Pressure* and answered twelve multiple-choice questions with the text available. Six questions were low-level questions while the other six were high-level questions. Students in the immediate condition received elaborative feedback explaining the correct answer after the completion of each question. Students in the delayed feedback conditions received delayed feedback either in a summative manner (i.e., one elaborative feedback message per question) or in an integrative manner (i.e., a unique elaborative message that explains the correct answers and relation of ideas in regard to every block of four questions conceptually related). Students had one attempt to answer each question. A day later, students completed a learning phase containing twenty high and low-level open-ended learning questions.

A subsidiary goal was to compare the potential effect of the variable previous background knowledge on reading performance at training and posttest learning, with particular attention to the moderation effects of the variable Timing of Feedback (Immediate, Delayed Summative and Delayed Integrative). Additionally, we analyze individual differences on performance and learning based on the level of Previous Background Knowledge of students and the type of question (low and high-level).

Online data such as time on reading the text and reading feedback was recorded by the computer-based software *Read&Learn* (Vidal-Abarca et al., 2017). Following previous literature (Butler et al., 2013), in regard to the question-answering performance in the training phase, our first hypothesis predicted that students in the immediate feedback condition would benefit the most from receiving a unique feedback message explaining the correct answer on a question per question basis during training.

With regards to the learning phase, our second hypothesis states that, in general, integrative delayed feedback would facilitate the inferring of the correct answers while allowing students to build a more coherent mental model from the text (McNamara et al., 2009), compared to delayed summative feedback. Such type of processing would lead to greater learning gains in the learning phase, since providing a unique integrative elaborative message would provide a clear organizational structure allowing students to relate questions´ clues with the information from the text. This type of constructive processing has been reported as being involved in the construction of a mental representation that contributes to a coherent comprehension of content (Broek & Helder, 2017; van den Broek, Beker, & Oudega, 2015; van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011).

Accounting for the effect of previous background knowledge on question-answering performance in the Training phase, our third hypothesis predicts that students with higher levels of previous background knowledge will benefit the most from a delayed integrative elaborative feedback than those students with a low level prior knowledge. The timing and spaced effect of the delayed integrative feedback messages will facilitate how students with a high level of previous background knowledge process incoming information. As such, these students will be more likely to report the highest scores in both training and learning phases, especially on the high-level questions. However, this might not be the case for those students with lower levels of prior background knowledge who instead might benefit the most from receiving immediate elaborative feedback on a question by question basis.

In fact, we expect this effect to be particularly relevant on the low-level questions at both, Training and Posttest phases. Regarding behavioural online measures, our hypothesis four predicts that elaborative immediate feedback provided question by question will be processed faster as reflected by the online reading times of the feedback messages. In addition, we expected the delayed integrative elaborative feedback message to provide students with a more stable, inferential and precise mental model from the text in regards to each of the four sets of questions answered. This will have a clear impact on how students approached text re-visits compared to students in the immediate and delayed summative conditions, making students in our third condition less dependent on text re-reading.

In light of this, we have a fifth prediction that students in the integrative delayed feedback condition will search the text to a lesser extent compared to their counterparts in the delayed summative feedback condition. On the contrary, receiving delayed summative feedback in the form of isolated messages every four sequential questions undoubtedly would induce students to re-visit feedback messages to a greater extent. Such an interaction would reflect the students´ effort to relate each of the questions with the elaborative content of feedback provided in a delayed but individual manner.

3.3. METHOD

3.3.1. Participants and Design

A sample of 133 high school Spanish students completing the first and second cycles of secondary education with a mean age of (*M* = 12.72; *SD* = .67) years, participated in the study. Students were randomly distributed into three groups; forty- five students in the *immediate feedback (question per question)*, forty-five in the *delayed summative feedback* (*after the completion of four questions on a question per question basis*) and forty-three in the *delayed integrative feedback condition (a unique elaborative feedback message delivered after the completion of four questions)*. In a first previous session, students completed a *Previous Background Knowledge Questionnaire*. Out of the total sample (*M* =.50; *SD* =.18), and following the mean split method, 65 students were labeled as low-prior background knowledge students (*M* =.46; *SD* =.16) while 68 were labeled as high-previous background knowledge (*M* =.55; *SD* =.20).

3.3.2. Materials

Text. The continuous expository text we used dealt with *Atmospheric Pressure and The Wind*. We used a version of a text used in a previous study (Vidal-Abarca et al., 2017). It contained 20 paragraphs and 1.126 words. It was divided into 20 paragraphs, 17 of which contained information relevant for answering the questions. Any student wishing to construct an integrated mental model from text would need to inspect the whole text, by processing and integrating relevant units of information in order to answer our comprehension questions.

Previous Background Knowledge Questionnaire. We administered a Previous Background Knowledge Questionnaire consisting of 30 true/false items on the general domain of sciences relevant to the information presented in the text.

17 test items were true while 13 were false, and the maximum test scored up to 30 points. The Previous Background Questionnaire had previously reported a moderate reliability index of .74 as reported by Rubio (2018). An example of a true general domain item was the following: "*Compressing a gas increases its density*."

Training Questions. Questions in this study were the same as in Vidal-Abarca et al., (2017). It consisted of six low-level and six high-level multiple-choice questions aimed at helping readers to comprehend and learn the text better. Learning questions contained a statement of the questions plus four lures for each question. Similar to those questions in Cerdán et al (2009), high-level questions were aimed at relating different pieces of information by interconnecting ideas present in the 17 paragraphs in the text. Low-level questions referred to verbatim units of information located within one paragraph or a maximum of two, with no or minimal inferences (*see Appendix A*). Similar to Ozuru et al., (2007), lures included one target correct option and three types of distracters: idea – overlap (surface text features located in an inappropriate context), error inducing (erroneous information from text) and unrelated distracters (irrelevant information to the statement of the question per se).

Posttest Questions. A final test was used to assess learning which consisted of twenty open-ended questions, ten of which were low-level and the other ten were high-level questions. These questions were constructed so that they covered the same ideas as those manipulated in the training phase to produce both, the low-level and high-level questions (*see Appendix A*). We counted whether the fourteen learning ideas from the text were or not present in students´ responses, according to a coding scheme previously established by the experimenters. Scoring was set as one point was given for each idea included in the answer. Half point was given when experimenters considered a given response (in terms of number of ideas) was not entirely complete. The scoring procedure for the post-test questions was conducted by two raters.

To guarantee inter-rater reliability, one third of the responses were first scored simultaneously by the two raters and a correlation index was calculated; it was r ¼ 0.70. The rest of the question sets were then scored separately by each rater. Further discrepancies were discussed and resolved.

Experimental Conditions; Level of Integration of the Elaborative Feedback Message:

Immediate Elaborative Feedback Condition: Right after answering to each question students received KCR feedback with information on the rightness or wrongness of the student´ response visually depicted on the screen and an elaborative feedback message regarding to correct answer (Figure 1c). For example, for the question:

Why does a storm originate?

a) Because the surface air is denser than the surrounding air, causing it to descend.

b) Because the air of the surface is as dense as the one that surrounds it but it is colder.

c) Because the air of the surface is less dense than the surrounding air, causing it to rise.

d) Because the air of the surface is as dense as the one that surrounds it but it is hotter.

Elaborative Feedback Message: Right at the surface of a storm, there is a mass of air which contains fewer particles compared to the air masses that surround it; and therefore, this central air mass is surrounded by air with a larger amount of particles that pressures it.

Delayed Summative-Elaborative Feedback Condition: Right after answering to every group of four questions, students received KCR feedback with information on the rightness or wrongness of the student´ response visually depicted on the screen and an individual elaborative feedback message to each of the four questions (Figures 1d). For example, for the next group of four questions:

Why does a storm originate?

a) Because the surface air is denser than the surrounding air, causing it to descend.

b) Because the air of the surface is as dense as the one that surrounds it but it is colder.

c) Because the air of the surface is less dense than the surrounding air, causing it to rise.

d) Because the air of the surface is as dense as the one that surrounds it but it is hotter.

Elaborative Feedback Message: Right at the surface of a storm, there is a mass of air which contains fewer particles compared to the air masses that surround it; and therefore, this central air mass is surrounded by air with a larger amount of particles that pressures it […]

On the seashore, during the summer days at noon, sometimes we can notice a bit of breeze. What is it due to?

a) The inlet of fresh air that comes from the sea refreshing the seashore.

b) The exit of air from the interior towards the sea refreshing the seashore.

c) The movements of large masses of air due to atmospheric pressure.

d) The earth after noon, which is colder than the sea.

Elaborative Feedback Message: During the day, the interior air becomes less dense due to changes in its temperature, so the interior air becomes less dense as its temperature varies, so it tends to ascend and its space will be occupied by the cooler air that comes from the sea.

Integrative Elaborative Feedback Condition: Students in this condition received KCR to each of the questions and a unique elaborative feedback message that appeared right after answering the fourth and last question of each group of four questions. For example, for the next group of four questions:

Why in the atmosphere the equivalent of the volume of one liter of cold air weighs more than that of a liter of hot air? […]

Why does a storm originate? […]

What would happen in Spain if there was a storm in the Mediterranean and an anticyclone in Portugal? […]

On the seashore, during the summer days at noon, sometimes we can notice a bit of breeze. What is it due to? […]

Integrative Elaborative Feedback Message: "…The temperature influences the speed at which the particles of a gas move and therefore, how close they may be. Thus, two equal volumes of air at different temperatures may have different densities. The density of a gas is related to its movement. As such, right at the surface of a storm, there is a mass of air which contains fewer particles compared to the air masses that surround it; and therefore, this central air mass is surrounded by air with a larger amount of particles that pressures it. Since in the low pressure areas the air is less dense, it rises and that space tends to be occupied by dense air coming from the high pressure areas. That is why the wind always moves from the high pressure areas to the low pressure areas…Likewise, during the day on the seashore, the interior air becomes less dense as its temperature varies, so it tends to ascend and its space will be occupied by colder air…"

3.3.3. Apparatus

The Text and learning questions were presented on a computer screen using the software *Read&Learn* (Vidal-Abarca et al, 2017). The software presents a full screen of the text completely visible to the student (see Figure 1a). By scrolling up and down students could read the text at their own pace. A button located up at the left side of the text (i. e., Questions) gives access to the questions´ screen. Each question´s statement and its four respective lures appear masked on the screen. Only the statement of the question and questions lures are visible at a time (see Figure 1b). Students need to unmask a segment (i.e., question statement or question lures) by clicking on it; question segments and feedback messages unmask alternatively. *Read&Learn* (Vidal-Abarca, et al. 2017) allows readers to reread the segments in any order they choose By clicking on button located below the question (i.e., *Back to the text* or *Next*), students are able to go back and forth from one question to another, back to the text (prior to provide a response) and vice versa. After students selected an option to answer, a feedback message appeared on the screen according to each feedback condition. Students had one attempt to answer each question and feedback and questions appeared masked on the screen. Students could unmask a segment (i.e., question statement, lure or feedback message) but were not allowed to refer back and forth from the feedback screen to the text after answering. All students proceed to complete the experimental task according to their corresponding feedback conditions; *immediate elaborative feedback (question per question)*, *delayed summative elaborative feedback* (*after the completion of four questions on a question per question basis*) and *delayed integrative elaborative feedback condition (a unique elaborative feedback message delivered after the completion of four questions)*.

Succeeding at completing the Testing Phase, students were directed to a screen where the message "*Thanks for your collaboration*" indicated the end of the interactive learning session.

Figure. 1

Screenshots of the text and questions while receiving feedback

Figure. 1a.

Screenshot of the text passage while reading initially on Read&Learn (students accessed the question by clicking on the Questions button)

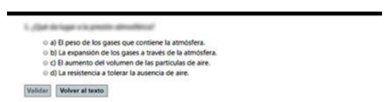


Figure. 1b.

Screenshot of a question.

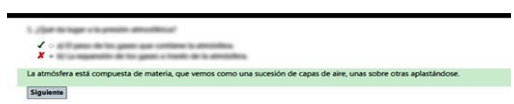


Figure. 1c.

Screenshot of a question and feedback for the Immediate Elaborative Feedback Condition.

Figure. 1d.

Screenshot of the feedback messages for the Delayed Summative Elaborative Condition.

Figure. 1c.

Screenshot of the feedback messages for the Delayed Integrative Elaborative Condition

3.3.4. Procedure

The study took place in two consecutive sessions of 50 minutes, Training and Posttest Phases both took place during the period of a scheduled secondary education class. Prior to the start of the experiment, researchers provided instructions on how to perform the experiment supported by a projector. First, students completed a paper and pencil Previous Background Knowledge Questionnaire. This phase took place for 15 minutes, approximately. In a second day, students joined the Training Phase where all participants accessed each interactive training session utilizing a user name and a code that was provided by the experimenters. A day later, in the Posttest phase students completed the learning questionnaire consisting of twenty open-ended low and high level learning questions.

3.3.5. Measurement

Off-line Performance Measures

Training Phase:

Question-Answering Performance. We computed students‟ scores for the twelve multiple-choice training questions. This score was computed as the number of right answers divided by the maximum number of points that each student could obtain in the experimental task (twelve points).

Previous Background Knowledge: Additional analyses to the Training phase were performed by computing the score on PBK as the total of the number of the right responses to the 30 items included in the questionnaire on the general domain of sciences. The maximum test score was 30 points.

Posttest Phase:

Learning. Students answered twenty low and high level open-ended learning questions. The maximum score in these questions was 20 points.

On-line Behavioral Measures

Time recordings. To register the students´ behavior during the Training Phase, a number of online measures were taken using the *Read&Learn* software. Based on the global time measures (in seconds), we obtained: (a) the total reading time on the experimental task (i.e., *time students spent reading the text and answering the twelve learning questions on Read&Learn*), (b) the initial reading time of the text (i.e., *time students spent reading the text initially, prior to answer the questions*),(c) text re- visits (i.e., *the number of times students accessed the text while answering the questions and receiving feedback*), (d) the time spent re-reading the text (i.e., *time students spent re-reading the text during question-answering*), (e) feedback re-visits (i.e., *the number of times a student unmasked feedback messages after responding*), (f) time reading feedback messages (i.e., the *time students spent reading the feedback messages after responding*)**,**

(g) and time reading questions (i.e., *the time students spent reading each question prior to answer*).

3.4. RESULTS

Relevant data of our dependent variables are listed below. A preliminary analysis was performed using a set of factorial ANOVAs. As such, results for the comparison between feedback conditions (Immediate, Delayed Summative and Delayed Integrative) for the *Training* (question-answering performance) and *Posttest* (Learning) *Phases* are presented (hypothesis 1 and 2). Next, results report a moderation analysis which reports the moderation effects of Timing of Feedback (Immediate, Delayed Summative and Delayed Integrative) and Previous Background Knowledge on Performance (Training) and Learning (Posttest). Further, an additional set of 3x2 factorial analyses accounting for the individual differences across conditions are presented, based on the level of previous background knowledge of students (high and low) (PBK). Finally, data on the behavioral online question-answering measures is presented (hypothesis 5).

Off-line Performance Measures

For the variable *Performance* on the *Training Phase,* contrary to our first and second hypothesis, a one-factor ANOVA reported no significant differences in-between our Immediate, Delayed Summative and Delayed Integrative feedback conditions, (F (2,130) = 1.94, p = .14, η^2 =.02) (Table 1). Similarly, the comparison between conditions regarding *the type of question* on the *Training Phase* reported no differences for our *low-level questions* (F (2,130) = 0.93, p = .39, η2 =.01) or *high-level questions (F* (2,130) = 1.93, p > .14, η^2 =.02). For the *Learning* variable on the *Posttest Phase*, no significant differences were apparent between our Immediate, Delayed Summative and Delayed Integrative feedback conditions, (F (2,130) = 1.42, p = 0.24, η^2 = .02) (Table 1). Alike, the comparison between conditions regarding the type of *question* in the *Posttest Phase* reported no differences for our *low-level questions* (*F* (2,130) = 0.64, p = .52, η2 =.01) or *high-level questions (F* (2,130) = 1.67, *p* = .19, η^2 =.02).

Table 1.

Mean and standard deviations of feedback timing on Question-Answering Performance and Learning in the Training and Posttest Phases.

	Immediate		***Delayed Summative***		***Delayed Integrative***	
	Mean	***SD***	***Mean***	***SD***	***Mean***	***SD***
Training Phase						
Question-Answering Performance	*.46*	*.16*	*.53*	*.19*	*.52*	*.20*
Low-level learning questions	*.54*	*.19*	*.59*	*.22*	*.60*	*.20*
High-level Learning Questions	*.37*	*.21*	*.47*	*.25*	*.43*	*.25*
Posttest Phase						
Learning	*.28*	*.13*	*.34*	*.17*	*.28*	*.17*
Low-level Learning Questions	*.30*	*.17*	*.33*	*.21*	*.28*	*.21*
High-level Learning Questions	*.27*	*.17*	*.33*	*.17*	*.28*	*.16*

Note: SD, standard deviation

Moderation Analyses for Previous Background Knowledge (PBK)

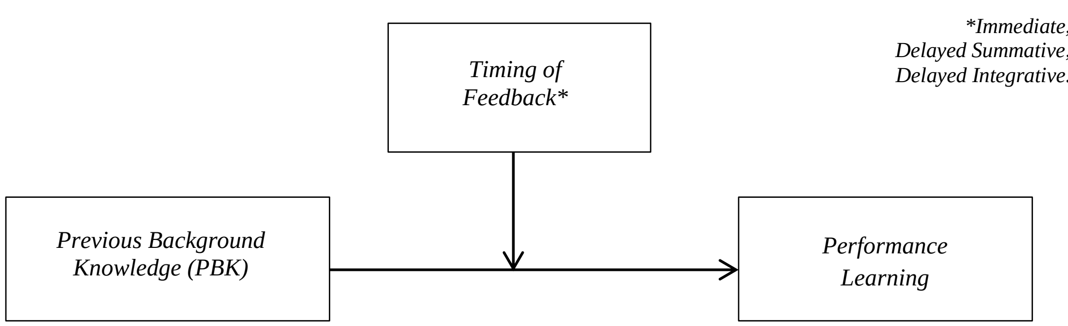


Figure 1.

Moderation model with a multi-categorical moderator (Timing of Feedback) influencing the size of Previous Background Knowledge on Performance (Training) and Learning (Posttest)

Moderation analyses for the Performance outcomes in the Training Phase showed interesting results. The overall model reported that the effect of the variable Previous Background Knowledge on Performance was moderated by the variable Timing of

Feedback (Immediate, Delayed Summative and Delayed Integrative), *F* (5,127) = 3.99, *p* = .001, R^2 = 0.13. Further, this interaction was significant, *ß* = .01, *t* (127) = 2.02, *p* = .00, CI95% (0, 0004 - 0, 0320). The interaction is illustrated in Figure 1. Conditional effects of the variable Previous Background Knowledge at values of the moderator Timing of Feedback reported that Delayed Integrative Feedback seemed to produce an increment in Performance over Immediate Feedback, ΔR2=.04, F(2, 127) = 3.19, *p* = .04. This interaction was significant, *ß* = .02, *t* (127) = 2.02, *p* = .00, CI95% (0, 0107 - 0, 0353). When analyzing the moderation effects of Timing of Feedback (Delayed Feedback Conditions) for the variable Previous Background Knowledge (PBK) on Performance with a sample of 88 students, the overall model reported that the effect of Previous Background Knowledge (PBK) was not moderated by Timing of Feedback (Delayed Summative or Delayed Integrative), *F* (3, 84) = 4.91, *p* = .00, R^2= 0.14. The interaction was not significant, *ß* = .01, *t* (84) = 1.17, *p* = .24, CI95% (- 0, 0084 - 0, 0326).

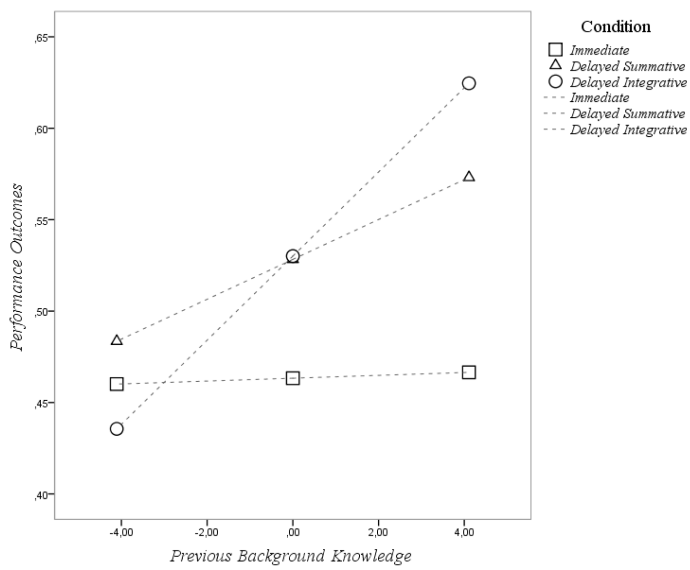


Figure 2.

Moderating Effects of Timing of Feedback (Immediate, Delayed Summative, Delayed Integrative) for Previous Background Knowledge on Performance (Training).

Moderation analyses for the Learning outcomes in the Post-test showed different results. The overall model reported that the effect of the variable Previous Background Knowledge on Learning outcomes was not moderated by the variable Timing of Feedback (Immediate, Delayed Summative and Delayed Integrative), $F(5,127) = 11.24$, $p = .00$, $R^2 = 0.30$. Interactions in this model were not significant for both the effects of Immediate versus Delayed Integrative Feedback, *ß* = .00, $t\,(127) = .92$, $p = .35$, CI95% (- 0, 0065 - 0, 0180), and the effects of Immediate versus Delayed Summative Feedback, *ß* = .01, $t\,(127) = 1.80$, $p = .07$, CI95% (- 0, 0013 - 0, 0293) (Figure 2). When analyzing the moderation effects of Timing of Feedback (Delayed Feedback Conditions) for the variable Previous Background Knowledge (PBK) on Learning with a sample of 88 students, the overall model reported that the effect of Previous Background Knowledge (PBK) was not moderated by Timing of Feedback (Delayed Summative or Delayed Integrative), $F\,(3, 84) = 13.12$, $p = .00$, $R^2 = 0.31$, since this interaction was not significant, *ß* = .01, $t\,(84) = 1.70$, $p = .09$, CI95% (- 0, 0023 - 0, 0303).

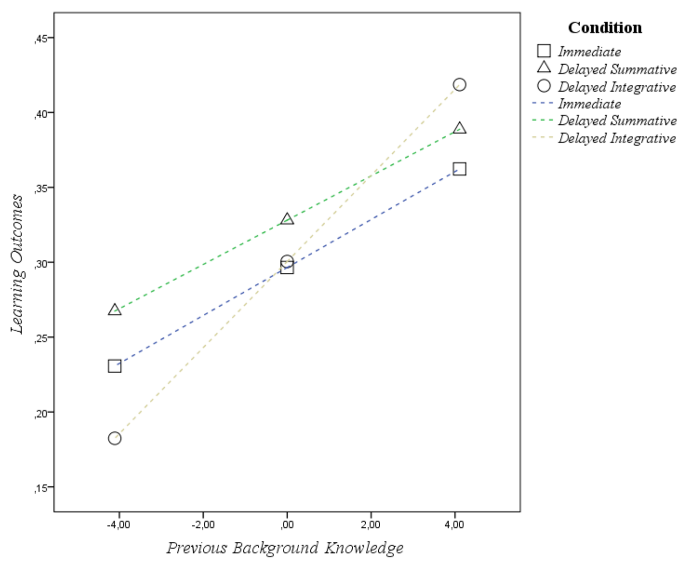


Figure 3.

Moderating Effects of Timing of Feedback (Immediate, Delayed Summative, Delayed Integrative) for Previous Background Knowledge on Learning (Posttest).

For the additional analyses on the differences according to the level of *Previous Background Knowledge*, we performed a set of factorial 3x2 ANOVAs where condition (Immediate, Delayed Summative and Delayed Integrative) was computed as the inter-variable while the level of previous background knowledge (high and low) was computed as the intra-variable. For our dependent variable *Question-Answering Performance* on the *Training Phase*, the overall model reported a significant interaction effect between condition and the level of previous background knowledge (F (2,127) = 3.92, p =.02, η^2 = .05). Students with higher levels of previous background knowledge (M =.55; SD = .02) performed significantly better than those with low prior knowledge (M = .46; SD = .02) (F (1.127) = 8.69, p = .00, η^2 = .06). In fact, those students with a higher level of previous knowledge in the Delayed Integrative feedback condition outperformed their counterparts in the Immediate condition (F (2, 127) = 4.62, p = .01, η^2 = .06). Moreover, comparison analyses within conditions partly confirmed our third hypothesis, those students with a high level of previous background knowledge in the Delayed Integrative feedback condition scored significantly higher than those with a low level of previous background knowledge (F (1,127) =15.10, p = .00, η^2 = .10). Unfortunately, no further significant differences were found (See Table 2).

For our *Learning* variable on the Posttest phase, and in line with our third hypothesis, a significant interaction between condition and the level of prior knowledge was found (F (2,127) = 6.2, p = .00, η^2 = .09); students with a high level of prior knowledge (M = .38; SD = .01) reported higher scores than those with low previous background knowledge (M = .24; SD = .01) (F (1,127) = 31.39, p = .02, η^2 = .19). In this case, this effect was particularly significant for both Delayed feedback conditions; Summative (F (1, 127) = 5.01, p = .02, η^2 = .03) and Integrative (F (1,127) = 36.20, p = .00, η^2 = .22). In addition, students with higher levels of prior knowledge in the Delayed Integrative condition not only outperformed students with low levels of prior knowledge in the same condition, but also in the Immediate condition (F (2,127) = 3.12, p. = 00, η^2 = .04). Interestingly, comparisons between conditions analyses reported that those students with a lower level of previous background knowledge in the Delayed Summative feedback condition outperformed their counterparts in the Delayed Integrative condition (F (2, 127) = 3.86, p = .02, η^2 = .05) (p=.03) (See Table 2).

Table 2.

Means and standard deviations for High and Low Previous Background Knowledge students in Question-Answering Performance (Training) and Learning (Posttest) per condition.

		Q-A Performance		*Learning*	
		Mean	*SD*	*Mean*	*SD*
Immediate					
	Low	*.46*	*.14*	*.26*	*.11*
	High	*.46*	*.18*	*.32*	*.15*
Delayed Summative					
	Low	*.50*	*.16*	*.28*	*.16*
	High	*.56*	*.21*	*.38*	*.17*
Delayed Integrative					
	Low	*.42*	*.17*	*.23*	*.13*
	High	*.64*	*.17*	*.37*	*.15*

Note: SD, standard deviation

A factorial 3x2 ANOVA Analyses on the differences in *the type of question* (*low and high level*) according to the level of *Previous Background Knowledge* on the *Training Phase*, reported a similar trend. For the *low-level questions*, a significant interaction was found between condition and the level of previous background knowledge (F (2,127) = 3.19, p = .04, η^2 = .04) where those students with a high level of previous background knowledge (M = .63; SD = .02) scored significantly higher than those with a low level of previous background knowledge (M = .53; SD = .02) (F (1,127) = 7.20, p = .00, η^2 = .05). In fact, those students with a higher level of previous knowledge in the Delayed Integrative feedback condition outperformed those students with a high-level of prior background knowledge in the Immediate condition (F (2, 127) = 3.53, p = .03, η^2 = .05). Additionally, students with higher levels of prior knowledge in the Delayed Integrative feedback condition outperformed those with a low level of prior background knowledge in the same condition (F (1.127) = 12.70, p = .00, η^2 = .09). For the *high-level questions*, even though students with higher levels of previous background knowledge scored higher than students with a low level of previous background knowledge (F (1,127) = 4.62, p = .03, η^2 = .03), the interaction between condition and level of prior background knowledge was not significant (F (2, 127) = 2.18, p = .11, η^2 = .03) (Table 3).

Table 3.

Means and standard deviations for High and Low previous background knowledge students for the low-and high-level questions in the Training Phase.

		Low-level questions		*High-level questions*	
		Mean	*SD*	*Mean*	*SD*
Immediate					
	Low	*.25*	*.16*	*.26*	*.13*
	High	*.35*	*.18*	*.29*	*.16*
Delayed					
Summative	*Low*	*.26*	*.19*	*.27*	*.17*
	High	*.38*	*.21*	*.37*	*.16*
Delayed					
Integrative	*Low*	*.13*	*.12*	*.20*	*.11*
	High	*.47*	*.13*	*.40*	*.14*

Note: SD, standard deviation

For *the type of question* on the *Posttest Phase*, a 3x2 factorial ANOVA on the *low-level learning questions* reported a significant interaction between condition and the level of prior knowledge (F (2,127) = 6.24, p = .00, η^2 = .09), where students with a high level of previous background knowledge (M = .40; SD = .02) obtained higher scores than those with low previous knowledge (M = .22; SD = .02), (F (2,127) = 36.64, p = .00, η^2 = .22). Comparisons within each condition showed that this effect was clearly significant in both the Delayed Summative (F (1,127) = 5.36, p = .02, η^2 =.04) and the Delayed Integrative feedback conditions (F (1,127) = 39.67, p = .00, η^2 = .23) being the latter mean difference remarkably higher (.34) (See Table 4). Additional data also supports part of our third hypothesis on the performance of students with low-levels of prior background knowledge;

students in the Immediate condition outperformed their counterparts in the Delayed Integrative condition (F (2,127) = 3.87, p = .02, η^2 = .05). In the case of our *high-level learning questions,* a significant interaction between condition and the level of prior knowledge was also found (F (2,127) = 3.58, p = .03, η^2 = .05) where students with a higher level of previous knowledge (M = .35; SD = .01) scored higher than students with a low level of prior knowledge (M = .24; SD = .01), (F (1, 127) = 17.23, p = .00, η^2 = .12). No further significant effects between conditions were found. Analyses within each of the conditions showed that those students with a high level of previous knowledge receiving Delayed Summative feedback outperformed those with a low level of previous background knowledge (F (1, 127) = 4.69, p < .03, η^2 = .03). A similar but a slightly higher effect was found within the Delayed Integrative feedback condition (F (1, 127) = 18.89, p = .00, η^2 = .12), where students with a high level of previous background knowledge outperformed those with a low level with a significant mean difference of .20 (see Table 4).

Table 4.

Means and standard deviations for High and Low previous background knowledge students for the low-and high-level questions in the Posttest Phase.

		Low-level questions		*High-level questions*	
		Mean	*SD*	*Mean*	*SD*
Immediate					
	Low	*.54*	*.18*	*.38*	*.21*
	High	*.55*	*.21*	*.37*	*.22*
Delayed					
Summative	*Low*	*.57*	*.18*	*.43*	*.25*
	High	*.61*	*.25*	*.50*	*.25*
Delayed					
Integrative	*Low*	*.50*	*.18*	*.34*	*.24*
	High	*.72*	*.15*	*.55*	*.22*

Note: SD, standard deviation

Online Behavioral Measures

Analyses reported very similar results to our previous data in regards to the general comparison of our three conditions (Immediate, Delayed Summative and Delayed

Integrative). A set of one-factor ANOVA´s were performed and contrary to our fourth hypothesis for the immediate condition, no significant differences were found on *the time students spent reading feedback messages* (*F* (2,130) = .22, p = .80, $\eta2$ = .00). Alike, data refute our fourth hypothesis since no differences were found in the time students spent re-reading the text (*F* (2,130) = 0.56, p = .57, $\eta2$ = .00) or the *number of text re-visits* (*F* (2,130) = 0.63, p = .53, $\eta2$ = .01). Similar results were found for the majority of our online behavioral measures such as *total time on the experimental task*, (*F* (2,130) = 0.40, p = .66, $\eta2$ = .00), *initial reading time of the text* (*F* (2,130) = 0.16, p = .85, $\eta2$ = .00), *time reading questions* (*F* (2,130) =1.29, p = .27 $\eta2$ = .02), (see Table 5). However, data partly confirm our fifth hypothesis. In fact, we found significant differences for the number of times students *re-visited the feedback messages*, with students receiving summative delayed feedback reporting higher numbers of re-visits to feedback messages corresponding to each block of four questions, compared to their counterparts in the immediate and delayed integrated feedback conditions, (*F* (2,130) = 16.58, p = .00, η^2 = .20).

Table 5.

Means and standard deviations of the effects of timing on the online measures (reading times and number of clicks) in the Training Phase.

	Immediate		*Delayed Summative*		*Delayed Integrated*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Total Time on Task	*1684.98*	*620.88*	*1711.12*	*605.91*	*1603.90*	*510.28*
Initial Reading of the Text	*469.31*	*216.25*	*453.09*	*142.05*	*449.38*	*158.05*
Number of Text re-visits	*7.13*	*5.13*	*6.00*	*5.33*	*6.95*	*4.81*
Time Re-reading the Text	*214.00*	*181.47*	*183.58*	*174.85*	*224.97*	*213.32*
Time Reading Feedback	*8.92*	*5.65*	*9.23*	*5.67*	*8.39*	*6.60*
Number of Feedback re-visits	*1.04*	*1.53*	*4.33*	*4.74*	*1.04*	*1.99*
Time Reading Questions	*47.91*	*16.37*	*43.81*	*17.42*	*42.85*	*12.92*

Note: SD, standard deviation

3.5. DISCUSSION

In the present study, a computer-based question-answering procedure (Read&Learn, Vidal-Abarca et al., 2017) delivered Elaborative Feedback (EF) that varied in the level of integration of the elaborative feedback message according to different timing schedules; Immediate (question per question), Delayed Summative (after the completion of four questions on a question per question basis) and Delayed Integrative (a unique elaborative feedback message delivered after the completion of four questions). Students read a text in the Atmospheric Pressure and answered twelve multiple-choice low and high-level learning questions with the text available. A day later, students were required to provide their responses on pencil and paper on a Posttest Phase containing twenty open-ended low and high learning questions. Our first goal was to analyze the effects of such a procedure on question-answering performance, online behavioural measures in the Training phase, and concept learning on the Posttest Phase. Similarly, a subsidiary goal was to explore the moderation effect of Timing of Feedback (Immediate, Delayed Summative and Delayed Integrative) and Previous Background Knowledge on Performance and Learning. Additionally, we explored the individual differences based on previous background knowledge on question-answering performance in the Training phase and Learning in the Posttest Phase, with attention to the type of question (low and high-level).

Contrary to our first and second hypothesis, first general analyses on the effects of Formative Feedback (FF) on both, question-answering performance on the Training Phase and learning on a Posttest Phase, did not report significant differences between conditions. However, we found that the effect of the variable Previous Background Knowledge on Performance was only moderated by Timing of Feedback (Immediate, Delayed Summative and Delayed Integrative) in the Training Phase. More specifically, data reported that

Delayed Integrative Feedback seemed to benefit Performance over Immediate Feedback. No more moderation effects were found. Additional analyses based on the level of prior background knowledge and the type of question showed that for our question-answering performance measure on the Training Phase, students with higher levels of prior background knowledge not only benefitted the most from receiving a Delayed Integrative feedback every four questions than those students with a lower level of prior background knowledge, but also outperformed those students with higher levels of prior knowledge in the Immediate condition. This effect was especially observable in the low-level learning questions. Unfortunately, we did not find significant effects on the high-level learning questions. Our Learning measure on the Posttest Phase reported rather different results; in fact, students with a high level of prior knowledge in both of the Delayed feedback conditions (Summative and Integrative) performed significantly better than those students with a low level of previous background knowledge. This effect was particularly observable when analyzing data by type of question (low or high level). In fact, for our low-level questions, students with a high level of prior background knowledge in the Delayed Integrative delayed feedback condition reported slightly better scores than those receiving Delayed Summative feedback.

For the high-level questions, high previous background knowledge students in the Delayed Integrative Feedback condition outperformed those with a low level, with a significant mean difference of .20. In addition, differences between students with either high or low levels of prior knowledge in both Delayed feedback conditions (Summative and Integrative) were significant. Those students with higher levels of prior knowledge who received a Delayed Integrative feedback message every four questions also scored significantly better than those students with higher levels that instead, received Immediate feedback question by question. In the case of our students with low levels of prior background knowledge, data also reported some interesting results. While those low-knowledge students in the Delayed Summative feedback condition outperformed their counterparts in the Delayed Integrative condition, when looking at the specific learning effects by type of question (low and high level), we found that for the low-level questions, those students with a low level of prior knowledge in the Immediate condition seemed to outperformed those students in the Delayed Integrative condition.

Although our results should be interpreted with caution, Prior Background Knowledge (PBK) seems to play an important role in the effectiveness of Formative Feedback for learning within the context of Sciences (Fyfe, 2016; Fyfe, Rittle-Johnson, & DeCaro, 2012; Krause, Stark, & Mandl, 2009; Smits, Boon, Sluijsmans, & Van Gog, 2008). Additionally, PBK is usually related to learning within the context of question answering and reading comprehension. In our case, we found that the Timing of Feedback (Immediate, Delayed Summative & Delayed Integrative) moderated the effect of Previous Background Knowledge on Performance. More specifically, Delayed Integrative Feedback seemed to benefit Performance over Immediate Feedback.

Quite commonly, within the context of computer-based feedback procedures, the powerful effects of Delayed forms of feedback are the result of the well-known spacing effect, by which a spaced presentation of feedback messages, i.e., in the form of single messages, has a beneficial effect on learning especially when students deal with question-answering retention tasks (Pashler, Rohrer, Cepeda & Carpenter, 2007, Peverly & Wood, 2001, Whittam, Dwyer & Leeming, 2004). This might be the case in the present study since, following those findings by Smith et al., (2008), we could conclude that students seem to benefit more from more complex and elaborative forms of Formative Feedback (FF) that aim to deliver global and structured information in relation to the task at hand.

In relation to those additional analyses on the type of question, we must clarify that our low-level learning questions were designed to induce students into input and retrieval processes based on explicit information from the text (i.e., names); a phenomenon that has classically being defined as The Testing Effect (Kulhavy & Stock, 1989; Butler & Roediger, 2008; Khang, McDermott & Roediger, 2007). According to the theory of disuse by Bjork and Bjork (2006), this type of memory retention process result from the association of items and are frequently based on the mere repetition and association to similar items in memory. Consequently, receiving an integrative elaborated feedback message after four questions may have eased this type of processing for high level background knowledge students. On the contrary, deep learning results from more complex cognitive processing that implies the relation and elaboration on the relevant concept information from the text (Cerdán & Vidal-Abarca, 2008; Cerdán et al., 2009). As previous research in reading comprehension and learning clearly states, the presence of high-inferential tasks might have a decisive impact on how students process information, especially for low-knowledge learners (Hamaker, 1986; Goldman & Durán, 1988; Cerdán & Vidal-Abarca, 2008; Máñez et al., 2017).

Results may have been the effect of the combination of our questions and the type of integration of the elaborative feedback in the present study. In this light, our high-level questions were designed to trigger complex inferences by prompting students to increase their efforts to relate and process complex ideas from both, the feedback messages and the text. Quite commonly, in these sorts of question-answering settings, students also need to deploy some important decisions such as when to revisit the text and what information is relevant to answer, like regulatory decisions that will allow them to update their mental model in a coherent manner while receiving complex Formative Feedback (FF) (Vidal-Abarca et al., 2017). As such, this might have been the type of processing of those students with high levels of prior knowledge in both delayed conditions (summative and integrative), who in fact, outperformed those students with lower levels of prior background knowledge in both types of questions, low and high-level.

The powerful effects of the delayed presentation of Formative Feedback (FF) compared to an immediate presentation may be the result of first, how feedback messages were distributed (every four questions) and second, how the level of integration of the elaborative feedback influenced the relation of ideas from the text inducing a more coherent understanding. While a spaced or Summative presentation of the elaborative feedback seems to benefit concept learning, a Delayed Integrativo feedback message aiming at providing a global and related structure of ideas from the text may have had a remarkable impact on how good comprehenders formed a coherent mental model from the learning information.

Likewise, students with a low level of prior background knowledge receiving Delayed Summative feedback outperformed those receiving Delayed Integrative feedback. In fact, analyses in regards to the type of question reported that students with lower levels of prior background knowledge in the Immediate condition outperformed those in the Delayed integrative condition. These results support the powerful benefits of the sequential presentation of Formative Feedback according to the traditional type of studies which have provided feedback on a question by question basis, either immediately or in its delayed form. As Kulik & Kulik (1977) defined, a traditional question by question sequence facilitates the interrelation of ideas that are presented in the feedback messages. This may have been beneficial for our low-knowledge students since it could have allowed them to process content in a more practical and relational way.

In a similar fashion to our first general analysis on question-answering performance and Learning on the Training Posttest Phases, we did not find significant differences in the majority of our online question-answering behavioural measures such as the time students spent reading feedback messages. A potential explanation for this effect may rely on how text availability was set up by our software Read&Answer (Vidal-Abarca et al., 2017). Students accessed the text right before answering each of the questions but did not once they had provided an answer to each of the questions and received elaborative feedback.

Quite often, research in question-answering tasks highlights the importance of receiving Formative Feedback (FF) right after answering a given question (Butler & Roediger, 2008; Butler, Roediger y Karpicke 2007; Eipstein, Dihoff, Cook, 2006; Llorens et al., 2016; Kapp et al., 2015), nevertheless, receiving Formative Feedback (FF) that varies in its timing (immediate or delayed at the end of four questions) during question-answering tasks with an available text requires students to engage in effortful cognitive and self-regulatory processing, such as how to discern the relevant information from the irrelevant information to answer (Rouet, 2006; Rouet, Britt & Durik, 2017; Vidal-Abarca et al., 2010).

This might be the case, especially when students receive complex elaborative Formative Feedback (FF). In this line, how often a student decides to search the text after receiving FF, strongly depends on several self-regulatory factors such as the accuracy to which students update their mental model while receiving FF that differs in its timing, and how students evaluate their own competence to answer appropriately (Candel, Vidal-Abarca, Cerdán, Lippman & Narciss, manuscript in preparation; Vidal-Abarca et al., 2017). In this light, other authors have also highlighted that any sort of inconvenient exposure to information that is intrinsically related to the elaborative content of a feedback message may be, in some cases, detrimental for learning (Kulhavy, 1977; Jeahnig & Miller, 2007). For example, some studies analyzing the effect of question-answering procedural Formative Feedback on reading behaviour have found that delivering certain types of task-specific feedback induced students into deeper but faster processing of the feedback message when the text was not available (Máñez et al., 2018). Other studies report rather different conclusions such as Llorens et al., (2016), who found that allowing students to access the text after receiving procedural feedback positively influenced the transfer of reading skills. In our study, access to the text after receiving Formative Feedback was not provided in an attempt to induce students to interact to a larger extent with the elaborative feedback messages, which in turn could facilitate feedback processing.

In our experiment, feedback messages did not contain explicit information from the text; however the level of integration of the elaborative feedback message allowed them to maintain the most relevant content-specific elaborations between the text and the feedback messages needed to help students elaborate on a mental model of the information coherently, especially for students in our Delayed Integrative feedback condition. Our experience suggests that text availability, questions and the level of integration of the elaborative information within each feedback message needs to be distributed very carefully across the learning environment especially when question-answering implies the teaching of complex concept knowledge.

Another factor may have been the format of our learning questions. While previous research has stated that multiple-choice questions may promote error correction and performance (Butler & Roediger, 2008), this may not be the best option when experimenters aim at enhancing deep processing from the text. In this regard, answering open-ended questions with an available text has shown to be a better promoter for students´ performance and for the active on-text searching (Ferrer et al., 2017). As a result, answering multiple-choice questions while receiving elaborative content feedback might not have been enough to tackle the deep processing of the feedback messages in order to produce significant gains in posttest learning.

Regarding the online behavior of students, results in part confirmed our fifth hypothesis. In fact, even though differences were not significant, contrary to our expectations, students receiving Delayed Integrative feedback reported higher rereading times of the text. However, students in the Delayed Summative feedback condition were more likely to interact with the feedback messages more often. This is an interesting result since in this condition FF was delivered in a sequential manner. Such a sequence might have eased the students´ interaction by inducing them to read back and forth each one of the feedback messages in a more repetitive manner. That type of reading activity would have allowed students to process the relevant content information from the feedback messages more easily. However, this may be only an assumption since data regarding the time students spent reading feedback messages did not report any significant differences between conditions. As a result, part of our previous third hypothesis is also refuted, since an immediate procedure did not ease the processing of the elaborative feedback messages.

Limitations and further research

Up to date, several studies on the effectiveness of the timing Formative Feedback for learning have focused on the role of the type of task as low or high-level outcomes (Shute, 2008; Van der Kleij et al., 2015). However, our results suggest the potential impact of the variable previous background knowledge in combination with delayed forms of feedback on question-answering performance on computer-based question-answering feedback settings with an available text. Further, task characteristics and individual differences (i.e., PBK) seem to be important factors for learning but should be considered carefully. Our data suggests that question-answering performance and concept learning are to some degree dependent on each other; and in fact, may differ depending on the type of task (i.e., low-level or high-level) and the level of the students´ prior background knowledge. However, future research needs to embrace that significant differences may rely on those variables that pertain to the student (i.e., prior knowledge, reading skills) and that, under certain circumstances they may remain imperceptible to the eye of the researcher.

However, our study presents a number of limitations. First, further research is needed into how individual variables may have a moderation effect on reading performance and learning when students complete question-answering tasks and receive Formative Feedback (FF). While moderation analyses are reliable ways to measure these potential effects, analyses based on individual differences in-between high and low students may suggest a potential trend, but cannot be widely generalized. Manually dividing the sample of students may help us explore the data; however more accurate analysis on the potential relationships between individual and experimental variables should be considered, in particular in the design of Formative Feedback (FF) procedures with multiple-choice questions when the text is available. Second, the timing of Formative Feedback and its distribution needs to be foreseen consciously. Complex concept feedback may be distributed rather immediately or after a delay, and still be presented in a serial or integrative manner. These may have a different impact on how the task design should follow those reliable learning principles that allow experimenters to differ and therefore measure, different types of learning as low or high level learning (Kintch, 1998; Bloom, 1956).

Further, a question pending is how restudying opportunities and an available text may hinder or induce the recognition and processing of the relevant information and how elaborative formative feedback should be interleaved within study trials in such a way that potentiates and supports learning. Undoubtedly, this reminds us that the level of students´ previous background knowledge is strongly related to an adequate understanding of information, especially in the context of sciences (Narciss, 2011; Hancock et al., 1995). Future research should also consider the potential effects of feedback timing immediate or delayed and any other individual variables that have reported to be essential for reading performance and content learning from text such as reading skills (Ramos-Soriano & Vidal-Abarca, 2014; Vidal-Abarca et al., 2010).

3.6. REFERENCES

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research Spring*, *49*(2), 280-318.

Butler, A. C., Godbole, N., & Marsh, E. J. (2012). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*(2), 290-298.

Brosvic, G. M., Dihoff, R. E., & Epstein, M. L. (2003). The role of feedback during academic testing: the delay retention effect revisited. *Psychological Record*, *53*(4), 533-548.

Brosvic, G. M., Epstein, M. L., Cook, M. J., & Dihoff, R. E. (2005).Efficacy of error for the correction of initially incorrect assumptions and of feedback for the affirmation of correct responding: Learning in the classroom. *The Psychological Record*, *55*, 401–418.

Butler, A.C., & Roediger, H.L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604-616.

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100* (1), 209–222.

Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences*, *21*(2), 201-205.

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*, 13-27.

Dempsey, J. V., & Wager, S. U. (1988). A taxonomy for the timing of feedback in computer-based instruction. *Educational Technology*, *28*, 20-25.

Dunlosky, J., & Rawson, K. A. (2015). Do students use testing and feedback while learning? A focus on key concept definitions and learning to criterion. *Learning and Instruction*, *39*, 32-44.

Epstein, M. L., Epstein, B. B., & Brosvic, G. M. (2001). Immediate feedback during academic testing. *Psychological Reports*, *88*, 889–894.

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology*, *108*(1), 82-97.

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, *104*, 1094-1108. DOI:10.1037/a0028389.

Fyfe, E.R. (2016). Providing feedback on computer-based algebra homework in middle-school classrooms. *Computers in Human Behavior*, *63*, 568-574.

Golke, S., Dörfler, T., & Artelt, A. (2015). The impact of elaborated feedback on text comprehension within a computer-based assessment. *Learning and Instruction*, *39*, 123-136.

Jaehnig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *Psychological Record*, *57*, 219–232.

Kapp, F., & Körndle, H. (2011). Was lerne ich aus einer Lernaufgabe? a) gar nichts, b) Faktenwissen, c) etwas über meine Lernstrategien, d) Antwort b und c sind richtig. In T. Köhler & J. Neumann (Eds.), Wissensgemeinschaften. Digitale Medien – Öffnung und Offenheit in Forschung und Lehre (pp. 178-187). Münster: Waxmann Verlag.

Kintsch, W. (1988). *Comprehension: A paradigm for cognition*. Cambridge, Uk; Cambridge University Press.

Krause, U.M., Stark, R. & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction*, *19*(2), 158-170.

Kulhavy, R. W., & Anderson, R. C. (1972). The delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 63*, 505-512.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*, 79–97.

Le Bigot, L., & Rouet, J.F. (2007). The impact of presentation format, task assignment, and prior knowledge on students´ comprehension of multiple online documents. Journal of Literacy Research, *39*(4), 445-470.

Llorens, A., Cerdán, R., & Vidal-Abarca. E. (2014). Adaptive formative feedback to improve strategic search decisions in task-oriented reading. *Journal of Computer Assisted Learning*, *30*(3), 233-251. DOI: 10.1111/jcal.12050.

López, S. J., Portolés, S., & Vidal-Abarca, E. (1993). Mejorando la efectividad instruccional del texto educativo en ciencias: Primeros resultados. *Enseñanza de las Ciencias, 11*(2), 137-148.

Mañá, A., Vidal-Abarca, E., Domínguez, C., Gil, L. & Cerdán, R. (2009). Papel de los procesos metacognitivos en una tarea de pregunta-respuesta con textos escritos. *Infancia y Aprendizaje*, *32*, 553-565.

Máñez, I., Vidal-Abarca, E., Martínez, T., & Kendeou, P. (2017, July). *Does feedback influence learning? The role of text availability and prior knowledge*. Paper presented at the 27th Annual Meeting of the Society for Text & Discourse. Philadelphia, USA.

Martínez, T., García, A., & Vidal-Abarca, E. *Elaboration On The Correct Response Or Elaboration On The Mistake: What Does It Work On Feedback?* 28th Annual meeting of the Society for Text and Discourse (ST&D), Brighton, United Kingdom, (July, 2018).

McDaniel, M. A., Fadler, C. L., & Pashler, H. (2013). Effects of spaced versus massed training in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1417-32.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27*(3), 360-372.

McDaniel, M.A., & Fisher, R.P. (1991). Tests and Test Feedback as Learning Sources. *Contemporary Educational Psychology*, *16*, 192-201.

McNamara, D.S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247-288. DOI:10.1080/01638539609544975.

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children‟s and adults‟ vocabulary learning. *Memory & Cognition, 37*(8), 1077-1087.

More, A. (1969). Delay of feedback and the acquisition and retention of verbal materials in the classroom. *Journal of Educational Psychology, 60*(5), 339-342.

Newman, M., Williams, R. G., & Hiller, J. K. (1974). Delay of information feedback in an applied setting: Effects on initially learned and unlearned items. *The Journal of Experimental Education*, *42*(4), 55-59.

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D.S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, *25*(4), 399-438.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *3*(19), 228-242.

Proske, A., Körndle, H., & Narciss, S. (2012). *Interactive learning tasks*. In N. M. Seel (Ed.), Encyclopedia of the Sciences of Learning (pp. 1606-1610). New York: Springer US. DOI:10.1007/978-1-4419-1428-6_1100.

Proske, A., Körndle, H., & Narciss, S. (2012). Interactive learning tasks. In N. M. Seel (Ed.), Encyclopedia of the Sciences of Learning (pp. 1606-1610). New York: Springer US. doi:10.1007/978-1-4419-1428-6_1100.

Rubio, A. (2018). Técnicas de aprendizaje de conocimiento científico a partir de textos. *Investigaciones Sobre Lectura, 10*, 1-29.

Sinha, N., & Lewis, A. (2015). Delayed, but not immediate, feedback after multiple-choice questions increases performance on a subsequent short-answer, but not multiple-choice, exam: Evidence for the dual-process Theory of Memory. *The Journal of General Psychology*, *142*, 118-134.

Smits, M., Boon, J., Sluijsmans, D., & van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: effects on learning as a function of prior knowledge. *Interactive Learning Environments*, *16*(2), 183-193.

Surber, John R., Anderson, Richard C. (1975). Delay-retention effect in natural classroom settings. *Journal of Educational Psychology, 67*(2), 170-173.

Susan R. Goldman & Richard P. Durán (1988) Answering questions from oceanography texts: Learner, task, and text characteristics. *Discourse Processes*, *11*(4), 373-412. DOI: 10.1080/01638538809544710.

Vidal-Abarca Gámez, E., & Sanjosé López, V. (1998). Levels of Comprehension of Scientific Prose. *Learning and Instruction*, *8*(3), 215–233.

Vidal-Abarca, E., Gilabert, R., & Rouet, J.F. (2005). El papel de las preguntas intercaladas en los textos de ciencias. *Revista de Investigación e Innovación Educativa, 36*, 129-147.

Vidal-Abarca, E., Martínez, T., Ferrer, A., & García, A. *Searching, Elaboration On The Correct Response Or Elaboration On The Mistake: What Does It Work On Feedback?*. 27th Annual meeting of the Society for Text and Discourse (ST&D), Philadelphia, PA, United States, (July, 2017).

Vidal-Abarca, E., Máñez, A., & Gil, A. (2010). Individual Differences for Self-Regulating Task-Oriented Reading Activities. *Journal of Educational Psychology*, 102 (4), 817–826.

Wiley, J., & Voss J.F. (2009). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*(2), 301-311.

CHAPTER 4

GENERAL DISCUSSION

4. Discussion

The general objective of the present thesis has been to analyze the effectiveness of two Formative Feedback question-answering procedures that specifically aimed at enhancing question-answering performance and learning from the text in two critical samples of students within the current education system; university undergraduates (*Study 1*) and secondary school students (*Study 2*). Students read an expository text in sciences and answered a set of learning questions using a Computer-Based System (CBS) named *Read&Answer* (Vidal-Abarca et al., 2017).

The system provided Formative Feedback (FF) that differed in the timing of its distribution (immediate or delayed) and the level of integration of the feedback message. To this, and building on an extensive theoretical synthesis on the role of Formative Feedback in question-answering with an available text (Llorens, Vidal-Abarca, & Cerdán, 2016; Llorens, Cerdán, & Vidal-Abarca, 2014; Máñez, Vidal-Abarca & Martínez, 2016; Máñez, Vidal-Abarca, Martínez & Kendeou, 2017; Martínez, García & Vidal-Abarca, 2018; Vidal-Abarca, Martínez, Ferrer & García, 2017), *Study 1* analyzed the extent to which a corrective feedback message provided in an immediate (question per question) or delayed manner (after the completion of four questions) had an impact on question-answering performance, online behavior, certitude of response and on the probability to correct errors compared to a non-feedback situation where students are required to reread the text, while they performed a two-attempt question-answering task with an available text.

Study 2 analyzed the effectiveness of a Formative-Feedback question-answering procedure that provided feedback in an Immediate, Delayed Summative and Delayed Integrative manner on the students´ question-answering performance and online behaviour. Additionally, *Study 2* explored more in-depth the moderation effect of Timing of Feedback and Previous Background Knowledge on performance and learning. Additional analyses explored how the level of prior background knowledge also had a potential effect on the way students responded to questions of different nature (low and high level).

As such, feedback was distributed in an immediate (question per question), delayed summative (an individual message to each question provided after the completion of four questions) or delayed integrative fashion (a unique integrative elaborative feedback message provided after the completion of four questions). For this study, and based on previous literature in reading and learning from text (Cerdán & Vidal-Abarca, 2008; Cerdán et al., 2009), we designed these elaborative messages as a coherent relation of ideas from the text in which the main function was to help students process and construct on knowledge in a more coherent manner.

From the evidence of the analyses of both experiments (*Study 1* and *Study 2*), we have concluded a variety of assumptions as follows; First, results of *Study 1* helped us clarify the role of an available text in the first place. Knowing that KR feedback would be provided after the first attempt and that students would have a second attempt to correct their mistakes made them spend more time rereading the text compared to when feedback was not provided. Due to the lack of feedback studies analyzing how an available text affects reading behaviour and question-answering performance we strongly consider this is an important finding. Ferrer et al., (2017) found that when students knew a text would be available during-question answering, they were more likely to spend less time reading the text initially compared to when they knew the text would be unavailable. These authors argued that this effect referred to a decrease in students´ standards of coherence when reading on the available condition. As a result students adapted their reading strategy to a situation in which an available text would serve as a source of information when needed. Students in the available condition likely thought that after reading the questions they would have more information on what and how to read, whereas this possibility would not exist for students in the unavailable condition. Feedback may have a similar effect. Students in the two feedback conditions probably thought that receiving corrective feedback while having a second attempt to respond after a mistake had been committed would provide them with the relevant information to make the decision to either reread the text or not. Therefore, students would adapt their processing to this situational variable.

Previous research in reading states that an available text is an important context variable that affects learning by facilitating or hindering comprehension (Ozuru et al., 2007; Schroeder, 2011). This question-answering process, in fact, has been widely highlighted by those theoretical models that focus on the student as the main constructor and self-regulator of knowledge.

For example, in their Model Bangert-Drowns, Kulik, Kulick and Morgan´ (1991), assume that, once students have processed feedback according to their own level of prior knowledge, they need to adjust their searching and retrieval memory strategies which at the same time, are activated by questions´ characteristics. While this process takes place, the student is able to evaluate his performance via monitoring adjustment processes (i. e., *the degree of certainty of the response*) and provide a response. In fact, results from *Study 1* report support this assumption since students who received immediate feedback were better able to discriminate between their correct and incorrect answers. Those students who committed a mistake decreased their response certitude judgments from the first to the second attempt, while they increased the proportion of correct responses.

This relationship between performance and response accuracy was also described in depth in Butler and Winne´ (1995) Feedback Model. According to these authors, after receiving feedback students generate an internal evaluation of their own performance, where somehow they are also able to assess their learning goals and the own strategies applied during the learning activity. While we found interesting results in regards to the monitoring accuracy of those students in the immediate condition, we did not find differences in the time students spent rereading the text, reading question or feedback messages in-between conditions. Still, after a mistake was committed, students in the Immediate condition were able to adjust their performance to a larger extent after being provided with the opportunity to reread the text and answer in a second attempt. In regards to question-answering performance, students who received feedback, regardless of its timing (immediate or delayed), outperformed those students who reread the text instead on the second attempt as well as on the repeated multiple-choice questions on the posttest.

A first conclusion is that the results of *Study 1* clearly confirmed the robust effects of corrective feedback on the probability of error correction and the retention of concept knowledge from Text. However, important consideration needs to be made; in *Study 1* Formative Feedback simply informed students about the correctness of their answers on the first attempt (KR feedback) and about the correct response after the second attempt (KCR feedback).

In the first case, KR feedback reduced the options among the choices, whereas in the second feedback students were informed about the correct response. As such, our results on performance seemed to be the result of the powerful effect of feedback by which correct answers are reinforced by the feedback message while allowing students to successfully retrieve the correct answers from memory (Eipstein, Dihoff & Cook, 2006). Nevertheless, these results are particularly interesting since our students responded to twelve high-inferential learning questions. These questions were specifically designed to induce students in the active inferential processing from the text. In such a case, we could assume that simple forms of Formative Feedback (FF) such as Knowledge of Response and Knowledge of the Correct Response are still powerful tools for the correction of the inferential mistake when question-answering implies an available text. However, the design of multiple-choice questions needs to be performed under considerable caution; for example by following an exhaustive analyses of the learning content that promotes comprehensive relations between complex ideas from an available text, in an attempt to avoid a selective memory benefit only of the previously tested items (Pan, Gopal & Rickard, 2016), an effect that has been commonly referred as the *Testing Effect* (McDermot et al., 2014; Roediger & Karpicke, 2006a).

Indeed, results from *Study 1* align with previous studies that have proved an available text to be an essential tool for the teaching of procedural knowledge when Formative Feedback is provided during question-answering tasks (Llorens, Vidal-Abarca, & Cerdán, 2016; Llorens, Cerdán, & Vidal-Abarca, 2014; Máñez, Vidal-Abarca & Martínez, 2016; Máñez, Vidal-Abarca, Martínez & Kendeou, 2017; Martínez, García & Vidal-Abarca, 2018; Vidal-Abarca, Martínez, Ferrer & García, 2017).

In *Study 2*, a computer-based question-answering procedure named *Read&Learn,* Vidal-Abarca et al., (2017) delivered Elaborative Feedback (EF) that varied in the level of integration of the elaborative information according to different timing schedules; Immediate (question per question), Delayed Summative (after the completion of four questions on a question per question basis) and Delayed Integrative (a unique elaborative feedback message delivered after the completion of four questions). Results from *Study 2* did not reveal significant differences for question-answering performance during the *Training Phase* or concept learning on the *Posttest Phase*.

However, data reported that the effect of the variable Previous Background Knowledge on Performance was moderated by Timing of Feedback (Immediate, Delayed Summative and Delayed Integrative). More specifically, Delayed Integrative Feedback seemed to benefit Performance over Immediate Feedback. The powerful effects of a delayed presentation of Formative Feedback (FF) over Immediate on learning have been addressed in past research (Shute, 2008). In fact, the moderation effect may have been the result of how feedback messages were distributed in the Delayed Integrative Feedback Condition (every four questions) and how the level of integration of the elaborative feedback presented a relation of ideas from the text which aimed at inducing students into a more coherent understanding. A Delayed Integrative feedback message aiming at providing a global and related structure of ideas from the text may have had a remarkable impact on how students formed a coherent mental model from the information in the text.

Additional analyses on the individual differences on performance, which were based on the level of previous background knowledge of students, were included in the statistical analyses. Even though these need to be considered very carefully, data reported some findings that may be worth discussing. In general terms, during the completion of the *Training Phase* students with higher levels of prior background knowledge not only benefitted the most from receiving a Delayed Integrative feedback every four questions but also scored higher than those students with higher levels of prior knowledge in the Immediate condition. This effect was particularly observable in the low-level learning questions. This result is interesting but it is to a certain extent, understandable. In fact, this result is pretty consistent with the literature that claims that delaying Formative Feedback (FF) is beneficial for the learning when students are required to complete low-level learning tasks such as pair-association of vocabulary items or the retention of words (McDaniel, Thomas, Agarwal, McDermott & Roediger, 2013; McDaniel, Fadler & Pashler, 2013; Butler, Karpicke & Roediger, 2007; More, 1969).

In this line, compared to an immediate feedback, providing Delayed elaborative messages after the completion of a set of questions, irrespective to its Summative or Integrative form, seemed to allow students with higher level of prior knowledge to construct and elaborate on the elaborative content from the Text as a whole more efficiently, as in the form of a new re-study episode (Butler, Godbole & Marsh, 2013; Butler, Marsch & Godbole, 2013; Kapp, Proske, Narciss & Körndle, 2015; van der Kleij, et al., 2015). Results from the Posttest also reported findings that align with those studies stating that the inclusion of high-level learning tasks may induce students into a more elaborative learning episode, which may be indicative of long-term learning gains (Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, 2009; Vidal-Abarca, Gilabert, & Rouet, 1998; Vidal-Abarca et al., 1996).

In fact, those students with a high level of prior knowledge in both of the Delayed feedback conditions (Summative and Integrative) performed significantly better than those students with a low level of previous background knowledge. In general terms, students with high levels of prior knowledge benefitted the most from a delayed presentation of Formative Feedback in both types of questions (low and high level). Moreover, those students with higher levels of prior knowledge who received a Delayed Integrative feedback message every four questions also scored slightly better than those students with higher levels of prior knowledge that instead received Immediate feedback question per question.

In the case of our students with low levels of prior background knowledge, we found that those who received Immediate feedback benefitted more on the low-level questions than those students in the Delayed Integrative Feedback condition. These results are interesting since they provide us with a potential view of how prior background knowledge can mediate the learning of concept knowledge from an available text while students receive Elaborative Formative Feedback during question-answering. Prior knowledge has shown to play an extremely important role in the effectiveness of Formative Feedback for learning in *Sciences*, *Biology* and *Maths* (Fyfe, 2016; Fyfe, Rittle-Johnson, & DeCaro, 2012; Krause, Stark, & Mandl, 2009; Smits, Boon, Sluijsmans, & Van Gog, 2008). However, within the context of feedback research, further research is needed to fully explore how continuous individual variables may have a moderation effect on performance. While moderation analyses are reliable ways to measure these potential effects, further individual differences between high and low students cannot be broadly generalized. To proceed with a split-mean by manually dividing the sample of students can certainly help us explore the data and infer interesting theories; however more accurate analysis on the potential relationships between the individual and experimental variables should be considered, in particular in the design of Formative Feedback (FF) procedures with multiple-choice questions when the text is available.

In our study, receiving Immediate elaborative feedback while students completed the low-level questions seemed to enhance the retention of literal information from the text. While the completion of complex tasks during Training can hinder performance while enhancing Posttest outcomes (Smith and Bjork, 1992), we did not observe the same effect for the high-level learning questions at Training or Posttest. Yet, this may be an important point to address in future research.

On the one hand, several results strongly confirm the powerful effect of feedback on the acquisition of concept knowledge when delivered after a certain delay or spaced presentation (Pashler, Rohrer, Cepeda & Carpenter, 2007, Peverly & Wood, 2001, Whittam, Dwyer & Leeming, 2004). On the other hand, our low-level learning questions were particularly designed to induce students into an input and retrieval process based on the processing and recall of explicit information from the text (i.e., names); a phenomenon that has classically being defined as *The Testing Effect* (Kulhavy & Stock, 1989; Butler & Roediger, 2008; Khang, McDermott & Roediger, 2007). As such, our results also alert us of the need to pay close attention to those students with lower levels of prior background knowledge, since these sorts of tasks may not be enough to tackle the constructive and inferential activity that is required in order to really help them acquire relevant concept knowledge from a text.

In light of the theoretical models we reviewed in the introduction of the present thesis, those variables that pertain to the students are of extreme importance. In this vein, the *Interactive Tutoring Feedback Model ITF* by Narciss (2013) emphasizes the type of multidimensional interaction between a feedback message and the student characteristics (e.g. previous knowledge, the certainty of response). In fact, this Model also assumes that the quality of the instructional activity (e.g. the nature, structure and form presentation of the information provided) is vital for the effectiveness of Formative Feedback (FF). The level of the integrative elaborative content of feedback messages in *Study 2* could have set a *standard or reference level* that was really appropriate for those internal standards of students with a high level of prior knowledge. As such, providing students with the challenging opportunity to receive Delayed Feedback messages, especially in their integrative form, may have helped setting the *external representation of competencies* that was required by the task at hand.

In *Study 2*, results did not report any significant differences in the online question-answering behaviour (i.e., time reading the text or time reading feedback messages). We assume that the presence of the elaborative feedback messages sufficed to address the level of comprehension of those students with higher levels of prior knowledge. However, this was not the case for low-level prior knowledge students as those *internal control actions* that are needed to perform successfully in high-level learning tasks might have implied the decision to deploy more information searching from the text. For this type of low-level of previous background knowledge students, this question-answering process may have a great impact on the cognitive, metacognitive and volitional resources of students as Körndle et al. (2004)., Kapp and Körndle (2011) and Kapp et al., (2015) described in their Model. Even though these are interesting results, they need to be considered carefully. Indeed, in order to successfully explore the role of the individual variables within the context of question-answering and feedback, first experimenters need to utilize reliable statistical analyses. Second, they also need to adjust question-answering procedures to the existing theoretical feedback models within sociological and psychological sciences.

In fact, when question-answering implies complex concept learning and an available text, students may need to cope with a set of cognitive and metacognitive processes such as an *orientation* and *planning stage*, where students activate prior-knowledge and select the relevant study strategies they need to follow according to the demands of the learning context. Also, during the *processing* and *evaluation of learning stages*, instructional questions provide students with the clues they need to monitor the searching, select and process relevant information. In fact, these cyclical process have been extensively addressed by recent theoretical models in comprehension and question-answering (Rouet, 2006; Rouet, et al., 2017). But how does this process really occur?

According to Vidal-Abarca et al., (2017) and their most recent theoretical question-answering synthesis, the processing of complex forms of Elaborative Feedback (EF) requires the disclosure of very cyclical strategic monitoring and processing decisions that imply not only the processing of question clues, but self-regulation skills such as *when* to revisit a text and *what* information is relevant to answer according to question demands.

For example, *Study 2* reported that those students in the Delayed Summative feedback condition revisited feedback messages more often compared to the rest of the conditions. Based on the relevant literature in reading comprehension and learning from text (Cerdán and Vidal-Abarca, 2008; Cerdán, Vidal-Abarca, Martínez, Gilabert & Gil, 2009; Goldman & Durán, 1988; Kintsch, 1998; McNamara & Magliano, 2009; McCrudden & Schraw, 2007; Rouet, Britt & Durik, 2017), we strongly consider our results confirm Vidal-Abarca et al., (2017) theoretical synthesis, since we can conclude that the level of integration of the elaborative feedback content in *Study 2* partly moderated performance and potentially determined how those students with higher level of prior knowledge were able to build and adjust a task model that allowed them to successfully complete our low and high-level learning questions on a Posttest.

In one way, these results show that Computer-Based Systems are useful but complex learning environments and there are a few aspects that remind us that the type of task and those variables that pertain to the student may, in some cases, compromise feedback efficiency (Narciss, 2011; Kulik & Kulik, 1988). As such, a first aspect refers to the inferential nature of the task (i.e., the level of integration of the feedback message). When providing formative feedback to students, educators need to be aware of the fact that the learning objectives strongly determines the inferential nature of the learning content within a feedback message as well as the type of question-answering tasks to apply. These are major factors that influence how students will be able to process the relevant elaborative information from a text into a coherent mental model (Vidal-Abarca et al., 2017).

Regardless of the timing at which feedback is provided (question by question or in a delayed cumulative manner after the completion of a set of questions), this effect may also vary according to the type of components a feedback message has. For example, simple forms of feedback (KR or KCR) may be an efficient manner to enhance error correction and question-answering performance of a sample of university undergraduates (*Study 1*); nevertheless, this may not be the case for a sample of secondary school students (*Study 2*). In our second study, we combined the elaborative feedback messages with questions that required two main different types of processing; the location and relation of specific units of information within the text, as demanded by our low-level questions, and the inferring of more complex relationships between ideas from a text, as required by our high-level learning questions. The inferential nature of the feedback messages might have added some extra processing demands on students. In addition, no opportunity for re-reading the text after receiving feedback was provided, which would have inhibited students from performing more searching in the text in case they were not able to understand the feedback messages. This could have resulted in a more demanding integration of the information present in the feedback message into a coherent mental model in memory. Even though we observed a beneficial effect of delayed formative feedback on high-inferential processing items, we did not observe significant effects on learning.

We must conclude that, when providing students with formative feedback, educators need to address both, the cognitive processing associated with the task and the level of integration of the elaborative feedback message. In this regard, learning taxonomies (Bloom, 1950; 1984) and other theoretical categorizations of the cognitive processes implied in reading comprehension and learning from text (Kinstch, 1998) must be followed and applied carefully, especially when educators attempt to design high-order learning question-answering tasks (McNamara, Jacovine & Varner, *in press*). It is well known that younger readers are more likely to struggle to understand academic texts, therefore failing to access deep meaning from the learning material; an essential process for the learning of conceptual knowledge within academic settings (McNamara & Kintsch, 1996). In this regard, educators must develop effective question content-analysis so the formative assessments and the delivery of formative feedback help students reach the current reading evaluation standards (OCDE, 2010).

Within the classroom, teachers may provide one-to-one feedback in regard to specific aspects of the inferential process within question-answering (Bloom; 1984; Golke et al., 2015). Nevertheless, individualized teacher-pupil instruction is highly difficult due to the number of students per class (Van der Kleij, et al., 2012). In this vein, Computer-Based Systems (CBS) are effective tools to support educators since they allow them to provide individualized formative feedback depending on the students´ responses and learning goals. Finally, an effective feedback message may combine different content elements that range from a simple form of correction (KR) or verification (KCR) to more conceptual and more procedural explanations (Bangert-Downs, et al., 1991).

It is of importance to mention that not all sorts of formative feedback may be equally effective for low and high order learning tasks (Butler, Godbole & Marsh, 2013; Marshall, 1991; Shute, 2008). While KR and KCR provided in an immediate manner right after answering to a particular item seem to be highly effective for error correction and the retention of the correct response of factual knowledge, the situational and cognitive demands may vary when educators aim at providing complex conceptual explanations to students. A delayed presentation of conceptual elaborations seems to be the most efficient manner to enhance students´ comprehension when these deal with high-level learning tasks (Kapp et al., 2015). Nevertheless, as results of *Study 1* confirmed, when students answer learning questions from an available text, the role of the text needs to be addressed carefully.

Educators must provide students with prompts, inferential explanations or any conceptual knowledge in a way that does not overlap with the content present in an available text. Educators must not forget that Formative Feedback (FF) needs to complement the information presented in the text; and in some cases, even providing the relevant information needed to fully understand the complex relationships between ideas.

Additionally, when answering interactive questions in Computer-Based Systems (CBS), students need to be able to assess the operational conditions of the learning environment (Kaap et al., 2015). These are specific needs characteristic of the computer-based reading environment; which educators need to be well aware of. Instructional feedback is then provided in an interactive environment that demands students to understand question demands, searching and retrieving the relevant information as well as selecting the right sources of information when available. The initial state of the learner may be then altered by the lack of domain-knowledge (i.e., web-LES environments) and may be strongly dependent on interests, motivation, reading goals or intuitive reading behaviour, amongst others. Students need to be able to manage the interactive aspects of the reading environment while being able to construct, evaluate, adjust and provide a response on the basis of the type and content of the feedback that is provided.

Teaching conceptual knowledge through the delivery of Formative Feedback during question-answering helps students to analyze and repair inferring errors while guiding students on the evaluation of their own self-efficiency.

Limitations

The general objective of the present work has been to analyze the effectiveness of two feedback tutoring strategies on reading performance and conceptual learning from text. Nevertheless, some limitations need to be considered. First, both studies presented above were unable to find significant differences in regard to the online question-answering reading activity students engaged in. This may be due to the lack of mechanical specificity in *Study 1*, in which the software we utilized was unable to record the students´ reading times of the text when prompted to re-read after a mistake was committed. The aforementioned issue limited us to perform very basic statistical analyses in relation to the students´ monitoring of re-reading opportunities. Similarly, in *Study 2*, the lack of differences in monitoring seems to depend on the experimental conditions set up by the methodology we applied itself. We presented students with three different conditions that administrated formative feedback on an immediate (i.e., question per question) or delayed manner (i.e., after the completion of four questions) in a cumulative or integrated manner. However, the presentation of the comprehension questions during the learning activity was roughly the same. In other words, for the delayed conditions even though the feedback was provided after the completion of four questions, these were presented to students in a sequential fashion; and that is, one by one.

As such, the text was available just prior to answering each question and as long as each question was presented to students. This may have equalized the students´ reading monitoring of both, each question and the textual information associated to answer. These effects may have been stronger in the delayed conditions since it may have diminished the need of these students to refer back to the text after receiving feedback. Researchers may account for these types of effects in learning and monitoring when designing formative feedback procedures. Adequate control of the availability of the text would have allowed us to gather the students´ revisiting times and reading times more accurately, which would have allow us to extract more reliable monitoring indexes while analyzing our data. Previous research in reading states that online question-answering data which is based on the direct measurement of students´ actions provide reliable data in regard to the students´ certainty and calibration of their own reading activity (Kapp, et al., 2015; Vidal-Abarca et al., 2010; Llorens et al., 2016). Such a measurement would have allowed us to compare the students´ searching decisions and re-reading times with the students´ correct answers at both performance and learning. We conclude the present study lacks this type of online measuring specificity in regard to the evaluation students make of their own need to search for information. To this, future researchers aiming at implementing formative feedback procedures within task-oriented reading scenarios need to be well aware of the need to design and adapt feedback tutoring accordingly in order to gather and analyze these types of metacognitive indicators.

A second limitation refers to the extension in time sequence of the experimental interventions (i.e., Training versus learning phase). Both studies included a testing and a post-test phase that students performed with a relatively short time difference. In *Study 1*, both phases were separated from one another by a maximum fifteen-minute interval, while in *Study 2* we extended this difference to one day. Previous studies analyzing the effectiveness of formative feedback for conceptual learning tend to apply longer lags of time in both, testing and post-test (Shute, 2008; Newman, et al., 1974). While some studies apply a testing phase and a follow-up post-test phase in a delayed manner (minutes, days and weeks), some others utilize longer and more repeated testing procedures before a post-test is applied (Sinha & Glass, 2015). Longer time lags have been discussed to influence memory and retention of facts in a more positive way (Cepeda et al., 2009; Shute, 2008), nevertheless in the present studies we were specifically interested in analyzing the potential effects of the experimental feedback strategies in the online question-answering processing and posterior comprehension of students, and its consequent learning when simulating a real, within-the-classroom reading situation; and that is, when conceptual learning activities within the classroom imply students to read a text and answer comprehension questions while receiving formative feedback from their teachers.

Finally, the last limitation refers to the type and format of the questions used in this work. In both studies, learners were required to answer multiple-choice questions during the testing and open-ended questions during a post-test phase. However, apart from the fact that we included a cued-recall test in study one and a set of text-based multiple-choice questions in study two, the major objective in both experiments was to measure deep learning at a post-test phase through the use of open-ended learning questions. To this, learning questions were particularly designed to induce high inferential activity in students. Especially in *Study 2*, this methodology differs widely form previous research in formative feedback in which multiple-choice questions are applied at both, a testing and a post-test phase (Van der Kleij, 2012).

Real academic settings quite often require students to construct an answer and elaborate on it (i.e., exams, argumentative essays, reading comprehension and analysis of expository texts) (Anderson & Thiede, 2008). These types of learning tasks are very effective to enhance learning from the text (Braten et al., 2017), nevertheless, the use and application of computer systems and feedback tutoring strategies to enhance conceptual knowledge from text is still a challenge for educators when students face these types of academic tasks. Designing formative feedback strategies that provide adaptive information on the students´ inferential mistakes is a complex matter.

Up to date, some researchers even question the validity and effectiveness of the application of computer-based systems and formative feedback in comparison to human tutoring in these sorts of comprehension activities (Golke, Doerfler and Arterlt, 2015). The study of the application of questions for conceptual teaching while students receive formative feedback has been fully addressed in past decades (Kulhavy, 1977), nevertheless, the vicissitudes and the added complexities of this kind of comprehension tasks demands exhaustive content analysis on the part of the researchers and eventually, on the educators. Future research should contemplate the role of the content of the adaptive feedback message in the inferential processing of students to a further extent.

Future Research

The present work intends to take a step further in the study of feedback tutoring for the teaching of conceptual knowledge from a text. It proposes new lines for the analysis of the feedback message from classical research in reading for learning, which otherwise would not have been properly considered. The present work highlights the importance of current practices in academic reading; and that is, the teaching of conceptual knowledge within the classroom which often implies the use of comprehension questions that require the analysis of specific contents within the educational curricula. In addition, the application of electronic feedback tutoring strategies to this kind of academic activities brings numerous advantages. In fact, computer-based instruction provides educators with a tool that beyond excluding the teacher from the learning experience, and it complements his/her function effectively while providing formative feedback that adapts to each of the student responses. Nevertheless, in order to generalize feedback effects on learning, educators need to be able to apply adaptive formative feedback in real academic contexts. To this, schools and universities need to provide educators with the online resources, training and support to design curricular material on their own.

Within a real academic context, there are several factors at stake and those variables that belong to the learner play a very important role in the analysis of the effectiveness of formative feedback for learning. To this respect, a future research line may contemplate the temporary length at which a feedback tutoring strategy can be adaptable to students´

responses. This would be particularly useful when students read an academic text and answer comprehension questions that require them to build and elaborate on their answers. In this line, the internal variables of the reader (i.e., previous background knowledge or the certainty of response) may also be considered. These variables may pertain to the cognitive and metacognitive states of the learner. For example, differences in learning may be apparent due to individual differences in the level of comprehension, reading skills, motivation or interest for a particular subject.

In addition to this, future research lines should also consider the type of formative feedback provided to students, the timing of distribution as well as the quality of its content.

Literature in the effective principles of guided instruction states that teachers´ feedback must be provided always after the completion of a task. It can be provided in the form of a daily review or a summary of the main ideas. In this line, new research practices may study how formative feedback tutoring strategies may help educators´ oral prompts during instruction. By allowing them to evaluate students´ progress, correct answers and level of confidence.

It is well-known that the more effective teachers ask comprehension questions to determine how efficiently students have acquired, rehearsed and connected knowledge (Rosenshine, 2012). In addition, they may correct errors, provide explanations, examples or conceptual maps, amongst others. Apart from promptings as a cognitive knowledge, educators model a determined use of a prompt and then guide students as they develop independence. In this line, future research should not only consider the evaluation of the students´ independent practice but consider both, the cognitive and metacognitive content of an elaboration feedback message and its impact on the students´ learning and performance.

In addition, the design of adaptive electronic environments allows researchers to analyze the role of scaffolding during learning. Academic question-answering activities for learning normally imply the use of different types of questions; these may require both, from the location of isolated content within a text to the writing of a more inferential summary of the content. Since one of the most important targets of an effective teaching procedure is to support low-skilled comprehenders, individualized feedback may also be provided according to a previous selection of tasks that may sequentially increase in complexity (Narciss, 2013; 2014; Hattie and Timperley, 2007; Shute, 2008).

In the case of a student lacking the previous-knowledge needed to understand a specific concept, a feedback message could provide students with a conceptual map in case they need to review a past lesson. After such an approach, formative feedback may provide specific explanations after students respond to a set of preliminary text-based questions inserted within an expository text. Then, after the students have reached a certain milestone in conceptual comprehension from the text, a computer based system may present them with more complex questions. "Why" and "How" questions are good examples of inferential questions that demand deep processing on the behalf of the student. A feedback message may focus then on providing a more complex and elaborated message in the form of a summary that connects distant parts of a text. Such a procedure may influence in a decisive manner how students interact with questions and feedback messages in order to comprehend the relevant information from a text effectively.

Implications for Educational Practice

The present work emphasizes the advantages of providing students with a computer-based procedure that provides formative feedback for the teaching of specific conceptual knowledge from the text. In this line, the reading of a text and the use of comprehension questions is still one of the most utilized approaches for learning within classroom settings. In this line, after answering questions, providing students with immediate or delayed conceptual explanations is vital for the comprehension of the previous studied material (Rosenshine, 2012). Such feedback should be not uniquely aimed at the correction of errors but at enhancing the deep understanding of the academic content. Indeed, educators need to be aware of the need to evaluate the students understanding of what a question demands as well as those actions students perform in order to succeed at the reading task at hand. These processes have been considered as high-order thinking by the current international assessments in paper-and-pencil format (i. e., PISA) as well as in online informational reading testing (i.e., PIRLS).

From the early years of schooling to primary education, educational practices seem to focus on providing students with the relevant declarative knowledge they need to adapt to national and international educational curricula. Declarative knowledge may vary depending on the subject at hand; however, the ability to comprehend written conceptual information refers to an interdisciplinary skill that is common to a wide range of academic subjects (i.e., *Sciences, Language literacy, Biology*). Students need to be able to comprehend the content and connect isolated pieces of information "as a whole". As students evolve, they are expected to become competent readers, good information searchers and ultimately, effective analyzers of relevant information. Academic settings require students to understand, compare and contrast knowledge in an autonomous manner. High order cognitive and metacognitive processes involve the processing of complex information as well as the required ability to read, locate and interact with comprehension questions in electronic learning environments.

It seems reasonable to advise educators to bring the practice of these high-order processes to the classroom, by utilizing the proper learning tasks. Processes may focus especially on the cognitive abilities that allow students to integrate isolated pieces of information as well as those tasks that require more complex elaborations. To this, beyond practice, formative feedback needs to focus on support students when facing these types of questions. There is a growing need for educators and educational systems to provide the resources students need to adapt to and overcome the information challenges within the current digital world since students will encounter learning situations in which they will be required to generate information-problem solving solutions (Rouet, 2009).

In a similar vein, past literature already defined the frame of successful instruction and undoubtedly resembles the current needs in instruction. As Hayes (1989) highlighted, educators must be aware of the need to consider the scope and sequence and to follow a clear cognitive taxonomy in order to induce students into the ongoing processing skills. Moreover, it needs to provide a solid evaluation procedure of behavioural indicators as well as indicators of the students´ attitudinal change (Hayes, 1989). In this line, the design of formative feedback tutoring strategies needs to be considered not as an exclusive procedure in question-answering but as a cognitive and metacognitive tool that support students through the learning process, and that additionally, it is adaptable to every course level and across a wide range of interdisciplinary course curricula.

4. REFERENCES

Anderson, J. R., Conrad, F. G., y Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science*, *13*(4), 467-505.

Azevedo, R., & Cromley, J. G. (2004). Does Training on Self-Regulated Learning Facilitate Students' Learning With Hypermedia? *Journal of Educational Psychology*, *96*(3), 523-535. Doi: 10.1037/0022-0663.96.3.523.

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L.-G. Nilsson & N. Ohta (Eds.), Memory and society: Psychological perspectives (p. 109-133). New York: Psychology Press.

Bjork, R. A., & Bjork, E. L. (1992). *A new theory of disuse and an old theory of stimulus fluctuation*. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), Essays in honor of William K. Estes, Vol. 2: From learning processes to cognitive processes (p. 35-67). Hillsdale, NJ: Lawrence Erlbaum.

Brackbill, Y., Bravos, A., & Starr, R. H. (1962). Delay-improved retention of a difficult task. *Journal of Comparative and Physiological Psychology*, *55*, 947-952.

Brackbill, Y., & Kappy, M. S. (1962). Delay of reinforcement and retention. *Journal of Comparative and Physiological Psychology, 55*, 14-18.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619-636.

Bråten, I., McCrudden, M. T., Stang, E., Brante, E. W., & Strømsø, H. (2017). Task-Oriented Learning With Multiple Documents: Effects of Topic Familiarity, Author Expertise, and Content Relevance on Document Selection, Processing, and Use. *Reading Research Quarterly*, *53*(3), 345–365. doi:10.1002/rrq.197.

Brosvic, G. M., Epstein, M. L., Cook, M. J., & Dihoff, R. E. (2005). Efficacy of error for the correction of initially incorrect assumptions and of feedback for the affirmation of correct responding: Learning in the classroom. *Psychological Record*, *55*, 401-418.

Brosvic, G. M., Epstein, M. L., Dihoff, R. E., & Cook, M. J. (2006). Acquisition and retention of Esperanto: The case for error correction and immediate feedback. *Psychological Record*, *56*, 205-218.

Butler, D. L., y Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281. doi: 10.3102/00346543065003245.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354-380.

Clariana, R. B., & Lee, D. (2001). The effects of recognition and recall study tasks with feedback in a computer-based vocabulary lesson. *Educational Technology Research and Development*, *49*(3), 23–36.

Clariana, R. B., & Koul, K. (2006). The effects of different forms of feedback on fuzzy and verbatim memory of science principles. *British Journal of Educational Psychology*, *76*, 259–270.

Corbett., A. T., & Anderson, J. R. (2001). *Locus of Feedback Control in Computer-Based Tutoring: Impact on Learning Rate, Achievement and Attitudes*, In Jacko, J., Sears, A., Beaudouin-Lafon, M. and Jacob, R. (*Eds*.) Proceedings of ACM CHI‟2001 Conference on Human Factors in Computing Systems, 245-252. New York: ACM Press.

Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2003). The role of feedback during academic testing: The delay retention effect revisited. Psychological Record, 53, 533-548.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. Psychological Record, 54, 207-231.

Ferster, C. B.; Skinner, B. F. (1957). *Schedules of reinforcement*. East Norwalk, CT, US: Appleton-Century-Crofts.

Freund, L., Kopak, R., & O‟Brien, H. (2016). The effects of textual environment on reading comprehension: Implications for searching as learning. *Journal of Information Science*, *42*(1), 79-93.

Gil, L., Martínez, T., & Vidal-Abarca, E. (2015). Online assessment of strategic reading literacy skills. *Computers & Education*, *82*, 50-59.

Holding, D. H. (1965). *Principles of Training*. Oxford, England: Pergamon Press.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research*, *30*, 138-149.

Juric, M. (2016). *The role of the need for cognition in the university students' reading behavior*. In Proceedings of ISIC, the Information Behaviour Conference, Zadar, Croatia, 20-23 September, 2016: Part 2. Information Research, 22(1), paper isic1620). Retrieved from http://InformationR.net/ir/22-1/isic/isic1620.html

Kapp, F., Narciss, S., Körndle, H., & Proske, A. (2011). Interaktive Lernaufgaben als Erfolgsfaktor für das nachhaltige Lernen und Lehren mit neuen Medien. *Zeitschrift für E-Learning*, *6*(1), 21-32.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, *95*, 163-182.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284. doi:10.1037/0033-2909.119.2.254.

Körndle, H., Narciss, S., & Proske, A. (2004). *Konstruktion interaktiver Lernaufgaben für die universitäre Lehre*. In D. Carstensen & B. Barrios (Eds.), Campus 2004. Kommen die digitalen Medien an den Hochschulen in die Jahre? (pp. 57-67). Münster: Waxmann.

Kulhavy, R.W., Stock, A. T., Hancock, E. L., Swindell, K. P., & Hammrich, L. (1990). Written feedback: Response certitude and durability. *Contemporary Educational Psychology*, *15*(4), 319-332.

Kulhavy, R.W., & Stock, W.A. (1989).Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*, 279. Doi:10.1007/BF01320096.

Little, J. L., Bjork E.L., Bjork R.A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychological Sciences*, *23*(11), 1337-44. doi: 10.1177/0956797612443370.

Llorens, A. C., Cerdán, R., & Vidal-Abarca, E. (2014). Adaptive formative feedback to improve strategic search decisions in task-oriented reading. *Journal of Computer Assisted Learning*, *30*(3), 233–251.

Llorens, A. C., Vidal-Abarca, E., & Cerdán, R. (2016). Formative feedback to transfer self-regulation of task-oriented reading strategies. *Journal of Computer Assisted Learning*, *32*(4), 314–331.

Llorens, A. C., Gil, L., Vidal-Abarca, E., Martínez, E., Mañá, A., & Gilabert, R., (2011). Evaluación de la competencia lectora: la prueba de Competencia Lectora para Educación Secundaria (CompLEC). *Psicothema*, *23*(4), 808–817.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*, *40*(4), 257-265. Doi: 10.1207/s15326985ep4004_7.

McCrudden, M. T., Magliano, J., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals, and text processing: A mixed methods study. *Contemporary Educational Psychology*, *35*(4), 229–241.

McMartin-Miller, C. (2014). How much feedback is enough?: Instructor practices and student attitudes toward error treatment in second language writing. *Assessing Writing*, *19*, 24-35.

Mory, E. H. (2004). *Feedback research revisited*. In D. Jonassen (Ed.), Handbook of research on educational communications and technology (pp. 745–783). Mahwah: Erlbaum.

Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, *71*, 56–76.

Narciss, S. & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training on written subtraction. *Learning and Instruction*, *16*, 310-322.

Narciss, S., Körndle, H. Reimann, G., & Müller. C. (2004). *Feedback-seeking and feedback efficiency in web-based learning – How do they relate to task and learner characteristics?*. In P. Gerjets, P.A. Kirschner, J. Elen, & R. Joiner (Eds.), Instructional design for effective and enjoyable computer- supported learning. Proceedings of the first joint meeting of the EARLI SIGs Instructional Design and Learning and Instruction with Computers (pp. 377-388). Tuebingen: Knowledge Media Research Center.

Narciss, S., Proske, A. & Körndle, H. (2007). Promoting Self-regulated learning in Web-based Learning Environments. *Computers and Human Behavior*, *23*, 1126-1144.doi:10.1016/j.chb.2006.10.006.

Newman, M. I., Williams, R. G., & Hiller, J. H. (1994). Delay of Information Feedback in an Applied Setting: Effects on Initially Learned and Unlearned Items. *The Journal of Experimental Education*, 42(4), 55-59.

O´Really, T., & McNamaran, D. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal*, *44*(1), 161-196.

O´Reilly, T., Best, R., & McNamara, D. S. (2004). *Self-explanation reading training: Effects for low-knowledge readers*. In K. Forbus, D. Gentner, & T. Regier (Eds.), Proceedings of the 26th Annual Meeting of the Cognitive Science Society (pp. 1053–1058). Mahwah, NJ: Lawrence Erlbaum.

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D.S. Comparing Comprehension Measured by Multiple-Choice and Open-Ended Questions. *Canadian Journal of Experimental Psychology*, *67*(3), 215–227.

Pintrich, P. R., y De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*, 4-13.

Rosenshine, B. (2012). Principles of Effective Instruction: Research-Based Strategies that All Teachers Should Know. *American Educator*, pp. 12-39. Extracted from https://www.aft.org//sites/default/files/periodicals/Rosenshine.pdf

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

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

Skinner, B. F. (1954). *The science of learning and the art of teaching*. Cambridge, Mass, USA.

Surber, J. R., & Anderson, R.C. (1975). Delay-Retention Effect in Natural Classroom Settings. *Journal of Educational Psychology*, *87*(2), 170-173.

Tarchi, C. (2010). Reading Comprehension of Informative Texts in Secondary School: A Focus on Direct and Indirect Effects of Reader's Prior Knowledge. *Learning and Individual Differences*, *20*(5), 415-420.

Timmers, C. F., & Veldkamp, B. P. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education, 56*, 923–930. doi:10.1016/j.compedu.2010.11.007.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. San Diego, CA: Academic Press.

Vidal-Abarca, E., Martínez, T., Gil, L., García, A., & Máñez, I. (2018). In Long, D., Magliano, J., Wiemer, K., & Millis, K. K. (Eds). Learning Tasks in Electronic Environments, Advances towards Interactive eTextbooks, Deep Comprehension; Multidisciplinary Approaches to Understanding, Enhancing and Measuring Comprehension. Routledge: New York.

Vidal-Abarca, E., Gilabert, R., Ferrer, A., Ávila, V., Martínez, T., Mañá, A., & Serrano,

M. Á. (2014). TuinLEC, an intelligent tutoring system to improve reading literacy skills/TuinLEC, un tutor inteligente para mejorar la competencia lectora. Infancia y Aprendizaje, 37(1), 25–56.

Vidal-Abarca, E., Martinez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., Ferris, R. (2011). Recording online processes in task-oriented reading with Read&Answer. Behavior Research Methods, 43(1), 179–192.

ANNEXES

ANNEX A: TEXT AND QUESTIONS

STUDY 1 ANNEX B: TEXT AND

QUESTIONS STUDY 2

ANNEX C: PREVIOUS BACKGROUND KNOWLEDGE STUDY 2

ANNEX 1

TEXT STUDY 1

The Earth and Space Foundation

Scientific exploration - the new frontier

The community that focuses its efforts on the exploration of space has largely been different from the community focused on the study and protection of the Earth´s environment, despite the fact that both fields of interest involve what might be referred to as scientific exploration. The reason for this dichotomous existence is chiefly historical. The exploration of the Earth has been occurring over many centuries, and the institutions created to do it are often very different from those founded in the second part of the 20th century to explore space. This separation is also caused by the fact that space exploration has attracted experts from mainly non-biological disciplines-primarily engineers and physicists - but the study of Earth and its environment is a domain heavily populated by biologists.

The separation between the two communities is often reflected in attitudes. In the environmental community it is not uncommon for space exploration to be regarded as a waste of money, distracting governments from solving major environmental problems here at home. In the space exploration community, it is not uncommon for environmentalists to be regarded as introspective people who divert attention from the more expansive visions of the exploration of space - the new frontier. These perceptions can also be negative in consequence because the full potential of both communities can be realized better when they work together to solve problems. For example, those involved in space exploration can provide the satellites to monitor the Earth´s fragile environments, and environmentalists can provide information on the survival of life in extreme environments.

A more accurate view of Earth and space exploration is to see them as a continuum of exploration with many interconnected and mutually beneficial links. The Earth and Space Foundation, a registered charity, was established for the purposes of fostering such links through field research and by direct practical action.

Foundation Research Projects

Projects that have been supported by the Foundation include environmental projects using technologies resulting from space exploration, satellite communication, GPS, remote sensing and power sources. For example, in places where people are faced with destruction of the forests on which their livelihood depends, rather than rejecting economic progress and trying to save the forests on their intrinsic merit, another approach is to enhance the value of the forests. In the past, the Foundation provided a grant to a group of expeditions that used remote sensing to plan eco-tourism routes in the forest of Guatemala, thus providing capital to the local communities through the tourist trade. This novel approach is now making the protection of the forests a sensible economic decision.

The Foundation funds expeditions making astronomical observations from remote, difficult-to-access Earth locations, archaeological field projects and field expeditions. A part of Syria - The Fertile Crescent - was the birthplace of astronomy, accountancy and many other fundamental developments of human civilization. The Foundation helped fund a large archaeology project by the Society for Syrian Archaeology at The University of California, Los Angeles, in collaboration with the Syrian government that used GPS and satellite imagery to locate mounds, containing artefacts and remnants of early civilizations.

Field research also applies the Earth´s environmental and biological resources to the human exploration and settlement of space. This may include the use of remote environments on Earth, as well as physiological and psychological studies in harsh environments. In one research project, the Foundation provided a grant to an international caving expedition to study the psychology of explorers subjected to long-term isolation in caves in Mexico. The psychometric tests on the cavers were used to enhance US astronaut selection criteria by the NASA Johnson Space Center.

Space-like environments on Earth help us understand how to operate in the space environment. In the Arctic, a 24-kilometre-wide impact crater formed by an asteroid or cornet 23 million years ago has become home to a Mars-analogue programme, The Foundation helped fund the NASA Haughton-Mars Project to use this crater to test communications and exploration technologies in preparation for the human exploration of Mars. The crater, which sits in high Artic permafrost, provides an excellent replica of the physical processes occurring on Mars, a permafrosted, impact-altered planet.

The Foundation's awards

In addition to its fieldwork and scientific activities, the Foundation has award programmes. These include a series of awards for the future human exploration of Mars, a location with a diverse set of exploration challenges. The awards will honour a number of firsts on Mars that include landing on the surface, undertaking an overland expedition to the Martian North Pole, climbing Olympus Mons, the highest mountain in the solar system, and descending to the bottom of Valles Marineris, the deepest canyon on Mars. The Foundation will offer awards for expeditions further out in the solar system once these Mars awards have been claimed.

ANNEX 1

Training learning questions STUDY 1

1. In line with the information in the text, what makes the Foundation´s award programme on Mars different from the rest of projects is...

a) Researchers are currently evaluating projects on natural environments in preparation for human expeditions.

b) Researchers want to investigate communications and their technological efficiency only on Mars.

c) Researchers want to find new flaming items and forms of energy only in the solar system.

d) Researchers want to develop human and technological expeditions on Mars and other possible space lands.

2. Advanced exploration technologies are used in the Syrian project to,

a) Help archaeologists find ancient items.

b) Explore land that is hard to reach.

c) Reduce the impact of archaeological activity.

d) Evaluate some early astronomical theories.

3. What was the main positive impact of the novel approach adopted in the Guatemala project on the locals?

a) Minimizing the need to protect the forests world-wide.

b) Reducing the impact of tourism on archaeological fields.

c) Demonstrating that touristic routes can be profitable.

d) Giving the Foundation greater control over the forests.

4. Scientific exploration can be defined as...

a) The study of the solar system phenomena for human knowledge and technology.

b) The study of advanced technology on similar workfields on Earth for space exploration.

c) The study of archaeology in order to design new technology to explore harsh environments.

d) The study of nature to replicate flaming processes occurring on other planets.

5. According to what the text states, the environment played a very important role for the NASA Haughton-Mars Project because,

a) It provided dry caves made by stone for astronomical observation.

b) It provided a permafrosted land for research on the location of mounds.

c) It provided a frozen-like land for testing new advanced technologies.

d) It provided an archaeological expedition for the location of fossils.

6. How do activities related to environmental community can contribute to space exploration?

a) By allowing researchers to develop on-Earth explorations recreating space-like environments.

b) By recruiting experts like physicists for future extraterrestrial exploration in the solar system.

c) By researching into space ecosystems, space fossils and impact-altered space landscapes.

d) By developing proposals for a world-wide safety policy on human and wildlife management.

7. A beneficial interest for the Foundation regarding both, the environmental and space exploration paradigms is,

a) The establishment of a collaborative framework for non-biological and biological disciplines.

b) The establishment of a collaborative framework for promoting eco-tourism for local communities.

c) The design of new ways of remote controls, GPSs and satellite communications in Syria.

d) To get the public attention into innovative developments of the exploration of the space.

8. The purpose of the Mexico´s Foundation project was,

a) To promote space exploration research for eco-tourism in Guatemala.

b) To promote space exploration research in long-term isolation scenarios.

c) To promote the study of nature and forest protection in ancient civilizations.

d) To promote the continuity of space exploration activities in the solar system.

9. The Earth and Space Foundation priorities regarding the Mars award programmes are,

a) Investing financial capital to recover naturally cultivated areas on Mars.

b) Promoting scientific studies into astronomy and astrology all around the world.

c) Developing advanced overland expeditions on the surface of remote areas like Syria.

d) Promoting projects for exploration research on different landscapes on Mars.

10. The common argument that has divided both communities referred in the text is,

a) Different technological and environmental resources were applied in the NASA Haughton-Mars Project.

b) Governments tend to allocate more money to environmental projects rather than to space exploration.

c) Space exploration and environmental communities differ in terms of technological and human resources.

d) The Earth and Space Charity was finally established later than it was originally intended by the Foundation.

11. The NASA Haughton-Mars Project is a good example of,

a) A programme that focuses on the use of satellite imagery to locate mounds.

b) A platform focused on the study and protection of the Earth's environment.

c) A project that shows how space-like scenarios serve exploration technologies.

d) A project focused on the location of artefacts in dry caves made by stone.

12. According to the author, in an attempt to preserve forests, the Foundation...

a) Provides the community with remote sensing and GPSs to monitor tropical environments.

b) Applies exploration technologies for the preservation of nature and the locals in harsh environments.

c) Advices environmental organizations on the protection of the forest for future economic progress.

d) Develops programmes for the preservation of archaeological field expeditions in harsh environments.

ANNEX 1

POST-TEST PHASE BOOKLET

6 REPEATED MULTIPLE-CHOICE QUESTIONS

DISTRIBUTED VS SPACED FEEDBACK EXPERIMENT 1 (2)

Please fill-in your PARTICIPANT ID and answer some questions regarding your person. The participant ID is necessary to make sure your data remain anonymous. It will also help us to match different answers.

Your PARTICIPANT ID:

1. CODE DIGIT 1:
2. CODE DIGIT 2:
3. CODE DIGIT 3:
4. CODE DIGIT 4:

Some questions about yourself:

5. Age:
6. Gender:

   Female Male

   Other Not given

7. Your mother tongue:

   Spanish English

   Other

8. Your current university level (e.g, fresh year):

GENERAL INSTRUCTIONS

In the following, you will find 6 multiple choice questions related to the text you studied in the previous phase. Read and answer them very carefully! You have 15 minutes for these set of tasks. If you have a question at any time, please raise your hand, and I will come to assist you.

Thanks for your collaboration!

1. In line with the information in the text, what makes the Foundation´s award programme on Mars different from the rest of projects is...

a) Researchers are currently evaluating projects on natural environments in preparation for human expeditions.

b) Researchers want to investigate communications and their technological efficiency only on Mars.

c) Researchers want to find new flaming items and forms of energy only in the solar system.

d) Researchers want to develop human and technological expeditions on Mars and other possible space lands.

2. Advanced exploration technologies are used in the Syrian project to,

a) Help archaeologists find ancient items.

b) Explore land that is hard to reach.

c) Reduce the impact of archaeological activity.

d) Evaluate some early astronomical theories.

3. What was the main positive impact of the novel approach adopted in the Guatemala project on the locals?

a) Minimizing the need to protect the forests world-wide.

b) Reducing the impact of tourism on archaeological fields.

c) Demonstrating that touristic routes can be profitable.

d) Giving the Foundation greater control over the forests.

4. A beneficial interest for the Foundation regarding both, the environmental and space exploration paradigms is,

e) The establishment of a collaborative framework for non-biological and biological disciplines.

f) The establishment of a collaborative framework for promoting eco-tourism for local communities.

g) The design of new ways of remote controls, GPSs and satellite communications in Syria.

h) To get the public attention into innovative developments of the exploration of the space.

5. The purpose of the Mexico´s Foundation project was,

a) To promote space exploration research for eco-tourism in Guatemala.

b) To promote space exploration research in long-term isolation scenarios.

c) To promote the study of nature and forest protection in ancient civilizations.

d) To promote the continuity of space exploration activities in the solar system.

6. The NASA Haughton-Mars Project is a good example of,

a) A programme that focuses on the use of satellite imagery to locate mounds.

b) A platform focused on the study and protection of the Earth's environment.

c) A project that shows how space-like scenarios serve exploration technologies.

d) A project focused on the location of artefacts in dry caves made by stone.

ANNEX 2

TEXT STUDY 2

LA PRESIÓN ATMOSFÉRICA Y EL VIENTO

La Tierra está rodeada por una capa de gases que la separa del espacio vacío que constituye, en su mayor parte, el Universo. Esta capa recibe el nombre de atmósfera y está formada por una mezcla de gases que llamamos aire.

En esta unidad aprenderás algunas características básicas de la atmósfera, entre ellas, qué es presión atmosférica y uno de los fenómenos más importantes relacionados con esta: el viento.

La presión atmosférica

La atmósfera está formada por un conjunto de gases que, como toda materia, está compuesto por partículas que tienen masa y, por lo tanto, peso. Se estima que el aire que compone la atmósfera pesa, aproximadamente, 5.500 billones de toneladas.

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

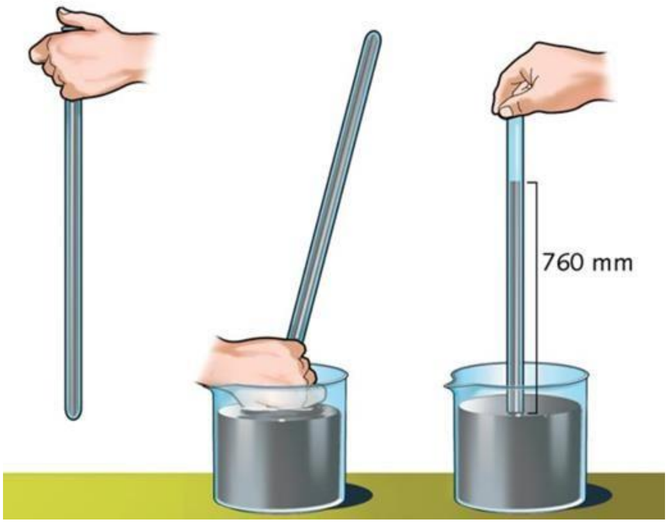
Todos los materiales y los seres vivos que poblamos este planeta estamos sometidos a la presión atmosférica. Si no somos conscientes de este peso del aire es porque ya estamos adaptados porque se ejerce por igual en todas direcciones y nuestros líquidos internos están a la misma presión.

El descubrimiento de la presión atmosférica.

La fuerza del peso del aire no se descubre hasta 1643. Fue el científico italiano Evangelista Torricelli quien ese año hizo un experimento que demostró que el aire ejercía presión.

La imagen 1 muestra cómo lo hizo: en primer lugar, llenó con mercurio un tubo de un metro de longitud cerrado por un extremo y tapó el extremo abierto con el dedo. Luego, lo introdujo invertido en una cubeta llena de mercurio y, finalmente, retiró el dedo con cuidado para que no entrara aire en el tubo. En ese momento, el mercurio descendió hasta una altura de 760 mm sin llegar a vaciar el tubo y dejó un vacío en el extremo cerrado del tubo. Así fue como Torricelli también fue el primero en establecer cómo medir la presión atmosférica.

Imagen 1. Experimento de Torricelli



En la antigüedad, se creía que el vacío no era algo natural y que la naturaleza se resistía a tolerar la ausencia de aire. Torricelli demostró con su experimento que no es esa resistencia al vacío lo que impide al mercurio salir del tubo, sino que es la atmósfera la que ejerce presión sobre el mercurio de la cubeta. Por eso, la columna de mercurio desciende hasta una altura de 760 mm, el punto que iguala la presión que el aire ejerce sobre el mercurio de la cubeta.

El aparato que diseñó Torricelli es el barómetro y se utiliza para medir la presión atmosférica. La presión del aire a nivel de mar equivale a la presión que ejerce una columna de mercurio (Hg) de 760 mm de altura, valor que equivale a una atmósfera (760mmHg = 1 atm). En meteorología se suelen utilizar otras unidades de medida como son el hectopascal (hPa) y el milibar (mb).

La presión atmosférica varía

Como sabes, en los gases las partículas se mueven libremente ocupando todo el volumen. Pero en la atmósfera, al ser un espacio tan grande, la distribución de las partículas que conforman el aire no es uniforme, debido a que las condiciones de cada lugar son diferentes. Así, la densidad de las capas de aire varía según la altura, ya que el tamaño de la capa de aire por encima es diferente. Por eso, en las capas inferiores de la atmósfera, que soportan el peso de todas las que están encima, las partículas de aire se comprimen más, se mueven menos y el aire es más denso (hay más cantidad de partículas de gas por unidad de volumen).

Como la presión depende del peso del aire que tenemos por encima, a medida que ascendemos la presión va disminuyendo. Por eso, la presión que existe en la cima de una montaña de 3.000 m de altitud es menor que la de una playa. Como la presión atmosférica varía con la altura se ha establecido que la presión normal, la equivalente a 1 atm (1013 hPa), es la que se da justo al nivel del mar, porque es la referencia que también utilizamos para indicar la altura de cualquier punto geográfico. Las presiones superiores a ésta se denominan altas presiones y las inferiores, bajas presiones.

La presión atmosférica además de variar con la altura también varía con la temperatura. Cuando los gases se calientan sus partículas se aceleran y tienden a expandirse, separándose y reduciendo su densidad, lo que afectaría a su presión; lógicamente ocurre todo lo contrario cuando se enfrían.

La presión atmosférica y el viento

Las diferencias en la densidad y presión atmosférica de las distintas partes del planeta son las responsables de los vientos y de otros fenómenos meteorológicos.

Las zonas de baja presión atmosférica, llamadas de ciclón o borrasca (imagen 2), se forman por masas de aire caliente que cuando ascienden dejan tras de sí un área de baja densidad. Mientras que en las zonas de anticiclon (imagen 3) son las masas de aire frío, más denso, las que tienden a descender desde las capas altas; causando la compresión de las masas de aire inferior, dando lugar a zonas de alta presión atmosférica.

Imagen 2. Formación de una zona de baja presión o borrasca. El aire caliente asciende y el hueco que deja lo llenan masas de aire vecinas.

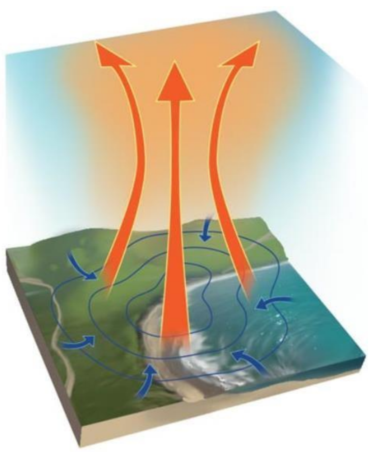
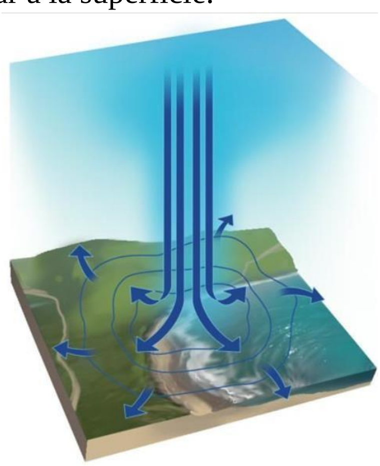


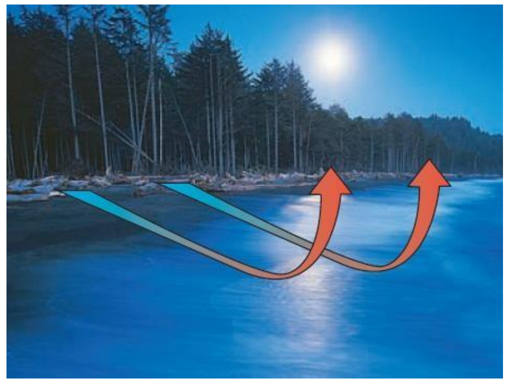
Imagen 3. Formación de una zona de alta presión o anticiclón. El aire frío desciende y comprime el aire inferior que se dispersa al llegar a la superficie.



Si combinamos estos dos fenómenos podemos entender cómo funciona la dinámica de la atmósfera y cómo se produce el viento. El viento es el movimiento de grandes masas de aire a través de la troposfera (la capa inferior de la atmósfera). La existencia de zonas de baja y alta presión ocasiona movimientos de aire que tienden a igualar las presiones de las distintas áreas, provocando corrientes de aire que van desde las zonas de alta presión hasta las zonas de baja presión. De la misma manera se producen otros fenómenos que son más locales y tenues como por ejemplo, las brisas diurnas y las nocturnas (imágenes 4 y 5).

Imagen 4. Durante el día, la tierra se calienta más deprisa que el mar. El aire caliente de la costa asciende y es sustituido por el aire más frío procedente del mar, proceso que da lugar a la brisa diurna.

Imagen 5. Durante la noche, la tierra se enfría más deprisa que el mar. El aire sobre el mar, más caliente, asciende y el lugar de éste es ocupado por el aire más frío procedente de la costa, hecho que da lugar a la brisa nocturna.



ANNEX 2

TRAINING LEARNING QUESTIONS STUDY 2

1. ¿Qué da lugar a la presión atmosférica?

a) El peso de los gases que contiene la atmósfera.

b) La expansión de los gases a través de la atmósfera.

c) El aumento del volumen de las partículas de aire.

d) La resistencia a tolerar la ausencia de aire.

2. ¿Por qué en el experimento de Torricelli el mercurio sólo desciende hasta los 760 mm?

a) Porque se crea un vacío en la parte superior del tubo que impide que siga cayendo.

b) Porque la presión de la atmósfera no es lo suficientemente grande para empujar el mercurio más abajo.

c) Porque a esa altura la fuerza que ejerce la masa del mercurio en la barra es igual a la fuerza que ejerce la columna de aire sobre la cubeta.

d) 240 mm de mercurio es la cantidad que necesitamos para igualar la densidad del mercurio de la cubeta a la densidad del aire.

3. Si Torricelli hubiera cogido un tubo de 2 metros de alto lleno de mercurio, en vez de un tubo de un metro, y hubiera repetido su experimento, ¿qué habría pasado con la columna de mercurio?:

a) Habría quedado a 1760 mm de altura.

b) Habría quedado al doble de altura a 1520 mm.

c) Habría quedado más abajo de 760 mm.

d) Habría quedado a la misma altura, 760 mm.

4. Si intentas replicar el experimento de Torricelli en lo alto de una montaña de ocho mil metros, en vez de al nivel del mar, donde se realizó el original ¿Qué crees que sucederá?

a) Que saldrá más mercurio del tubo a la cubeta, bajando más de 760 mm de altura.

b) Que saldrá menos mercurio desde el tubo a la cubeta, bajando menos de 760 mm de altura.

c) Que saldrá la misma cantidad de mercurio del tubo a la cubeta, quedando a 760 mm de altura.

d) Que debido a la altura saldrá casi completamente el mercurio del tubo.

5. ¿La densidad del aire es igual en todos los puntos de la tierra?

a) Sí, porque los gases se expanden y tienden a ocupar todo el espacio.

b) No, porque varía en función de la temperatura y la altura.

c) No, porque el viento hace que varíe la densidad del aire.

d) Sí, porque la mezcla de gases que componen la atmosfera es siempre igual.

6. ¿Por qué varía la presión atmosférica con la altura?

a) Porque a nivel del mar la presión se ve afectada por la brisa diurna y nocturna.

b) Porque cambia el peso de la columna de aire que tiene por encima.

c) Porque en la cima de una montaña de 3000 metros la presión es mayor al hacer más frio.

d) Porque la presión atmosférica normal es la que se da a nivel del mar.

7. Imagina que llenamos una botella de plástico flexible con aire bastante caliente y la cerramos herméticamente ¿Qué pasará con la botella cuando se enfríe?

a) La botella se hinchará y aumentará su tamaño al enfriarse el aire del interior.

b) La botella pesará más al enfriarse el aire del interior, el cual se hará más denso y pesado.

c) La botella pesará menos, ya que al enfriarse el aire se hará menos denso y pesado.

d) Que la presión atmosférica chafará la botella ya que al enfriarse el aire del interior reducirá su presión.

8. Imagina un globo aerostático abierto por la base, completamente hinchado y ascendiendo al calentar el aire del interior con un quemador ¿Qué ocurre con la densidad del aire dentro del globo?

a) Es igual a la del aire exterior porque el globo está abierto.

b) Es mayor porque aumenta el tamaño de las partículas.

c) Es menor porque el aire está más caliente.

d) Es menor porque el volumen del globo disminuye.

9. ¿Por qué en la atmosfera pesa más el equivalente al volumen de un litro de aire frío que el de un litro de aire caliente?

a) Porque las partículas frías pesan más.

b) Porque hay más partículas al estar más juntas.

c) Porque el aire caliente disminuye su volumen.

d) Porque las partículas del aire se dispersan al enfriarse.

10. ¿Por qué se origina una borrasca?

a) Porque el aire de la superficie es más denso que el aire que le rodea, haciendo que descienda.

b) Porque el aire de la superficie es tan denso como el que le rodea pero está más frío.

c) Porque el aire de la superficie es menos denso que el aire que le rodea, haciendo que ascienda.

d) Porque el aire de la superficie es tan denso como el que le rodea pero está más caliente.

11. ¿Qué ocurriría en España si hubiera una borrasca en el Mediterráneo y un anticiclón en Portugal?

a) En España al estar fuera de la zona de borrasca y de anticiclón habría un tiempo estable y sin viento.

b) En una distancia tan amplía no influiría una sobre la otra.

c) Habría movimientos de masas de aire para que la borrasca y el anticiclón se fueran separando.

d) Habría movimientos de masas de aire desplazándose desde Portugal hacia el Mediterráneo.

12. En la orilla del mar, durante el verano a mediodía, a veces podemos notar un poco de brisa ¿A qué se debe?

a) A la entrada de aire que proviene del mar refrescando la playa.

b) A la salida de aire del interior hacia el mar refrescando la playa.

c) A los movimientos de grandes masas de aire debidos a la presión.

d) A que la tierra después del mediodía está más fría que el mar.

ANNEX 3

PREVIOUS BACKGROUND KNOWLEDGE QUESTIONNAIRE

Nombre………………………………………………………………………………………
Edad………… Sexo…………… Curso………Grupo……………………....................
Colegio/Instituto…………………………………………………………………………….

INSTRUCCIONES

1. Para cada frase, marca con una **X** la opción verdadera (V) o falsa (F) según tus conocimientos.
2. Contesta cuando estés realmente seguro/a de la respuesta. Si no lo estás, marca NS (No sé).
3. Recuerda que cada error resta puntuación en la nota global de la prueba. Por eso, si no tienes seguridad sobre la respuesta, marca NS.

1	Dos objetos con el mismo volumen tienen la misma densidad.	V	F	NS
2	El aire de la atmósfera no pesa.	V	F	NS
3	La temperatura es una propiedad de la materia.	V	F	NS
4	El vapor de agua que contiene el aire forma las nubes.	V	F	NS
5	La densidad es la relación entre la masa y el volumen.	V	F	NS
6	Los gases se componen de partículas que se mueven libremente.	V	F	NS
7	El movimiento de traslación de la Tierra da lugar al día y la noche.	V	F	NS
8	Cuando un gas se comprime el tamaño de las partículas que lo componen también disminuye.	V	F	NS
9	La teoría geocéntrica describe la posición que tiene la Tierra en el universo.	V	F	NS
10	Todo gas cerrado en un recipiente provoca una fuerza sobre sus paredes.	V	F	NS
11	1 litro de volumen de cualquier líquido es igual a 1 kg de peso.	V	F	NS
12	Una partícula de un gas pesa siempre igual independientemente de su temperatura.	V	F	NS
13	La unidad mml/g mide la densidad.	V	F	NS
14	Las partículas de los gases están en constante movimiento.	V	F	NS
15	La presión del aire se ejerce sólo hacia abajo.	V	F	NS
16	La atmósfera está compuesta por partículas en estado gaseoso.	V	F	NS
17	La Tierra es el planeta más cercano al Sol.	V	F	NS
18	Las partículas de los gases se mueven más deprisa por el efecto del calor.	V	F	NS
19	Un litro de mercurio pesa lo mismo que un litro de agua.	V	F	NS

20	Cuando disminuye la temperatura de un gas disminuye su presión.	V	F	NS
21	El peso es la fuerza con la que la Tierra atrae un cuerpo.	V	F	NS
22	El peso de un gas dentro de un recipiente depende de la temperatura a la que se encuentre.	V	F	NS
23	Las mareas se deben principalmente a la atracción que la Luna y el Sol ejercen sobre la Tierra.	V	F	NS
24	Un gas sólo puede estar formado por partículas idénticas.	V	F	NS
25	Cuando ponemos en contacto un cuerpo frio y otro caliente, la energía pasa del cuerpo caliente al frio.	V	F	NS
26	Al comprimir un gas aumenta su densidad.	V	F	NS
27	Podemos conocer el volumen de un sólido si lo sumergimos en un líquido y medimos la cantidad de líquido que se desplaza.	V	F	NS
28	El tipo de partículas que componen un gas no influye en su peso.	V	F	NS
29	Los cuerpos en equilibrio térmico tienen la misma temperatura.	V	F	NS
30	El hielo flota en el agua porque su densidad es mayor.	V	F	NS