VNIVERSITAT DE VALÈNCIA

FACULTAD DE PSICOLOGÍA

Programa de Doctorado en Lectura y Comprensión

Effectiveness of feedback procedures to improve student's declarative knowledge in task-oriented reading environments.

Eficacia de los procedimientos de retroalimentación para mejorar el conocimiento declarativo de los estudiantes en un entorno de lectura orientado a tareas.

TESIS DOCTORAL CON MENCIÓN INTERNACIONAL PRESENTADA POR

ARANTXA GARCÍA LÓPEZ

DIRIGIDA POR

Dr. EDUARDO VIDAL-ABARCA GÁMEZ

Dr. TOMÁS MARTÍNEZ GIMÉNEZ

Valencia, enero de 2021

A mis padres, a José Andrés y a Jambo,

AKNOWLEDGEMENTS - AGRADECIMIENTOS

En estas líneas quiero expresar mi agradecimiento a todos aquellos que, de un modo u otro, me han acompañado, guiado y apoyado a lo largo de esta etapa.

A Eduardo y Tomás, mis directores, gracias por darme la oportunidad de descubrir el mundo de la investigación y de formar parte de vuestro equipo. Gracias Eduardo por guiarme y preocuparte por mi formación, por darme la oportunidad de asistir y de contribuir en congresos de reputado prestigio, y por transmitirme tu pasión por la investigación y la docencia. Gracias Tomás por toda tu dedicación, tu comprensión y empatía, por los ánimos y por toda la ayuda y esfuerzo que me has dedicado.

A Marian, por guiarme, cuidarme y motivarme. Gracias por todos tus ánimos, por estar pendiente de mi progreso y por compartir con alegría mis avances. Eres una grandísima profesional y parte fundamental de este trabajo. Espero, de corazón, que llegues dónde te propongas porque vales mucho.

Gracias a todos los compañeros y miembros del grupo Psicotext y de la ERI-Lectura. Desde el primer día me hicisteis sentir una más del equipo y he aprendido de todos vosotros semana tras semana a lo largo de estos años. A Raquel, gracias por hacerme llegar ese e-mail en junio de 2015 animándome a solicitar la beca, y por haberte interesado por mi evolución y prestado tu ayuda en todo momento. A Lalo, gracias por compartir tu experiencia conmigo a lo largo de los primeros años, dejarme aprender de ti y guiarme en la escritura de mis primeros artículos científicos. A Ramiro, gracias por tus consejos, por llenar los pasillos de risas y sonrisas, y por compartir momentos tan divertidos como los vividos en el congreso de Oulu, Finlandia. A Amelia y Vicen, gracias por todo vuestro apoyo, ánimo y palabras de aliento sobre todo en los momentos más difíciles, habéis sido un gran apoyo y he aprendido mucho de vosotras. A Antonio, Laura Gil, Javi, Inma, Gemma y Cristina, gracias por vuestros gestos de cariño y vuestros consejos. A Guillem, Dani y Francisco gracias por desarrollar y solucionar los problemas de Read&Learn y hacer tantas modificaciones como hemos pedido.

Gracias a Maria José Cantero, por tus sonrisas y miradas cómplices, por darme todo tu apoyo haciéndome saber que estabas ahí para lo que hiciera falta.

Pablo, gracias por todos los buenos momentos. Me he reído mucho contigo y siempre me has ayudado. No podría haber tenido mejor compañero de despacho que tú.

Alba, gracias por todos los momentos que hemos vivido. Hemos sabido apoyarnos mutuamente durante todo este tiempo.

I would also like to thank the Leibniz-Instituts für Bildungsverläufe for opening the doors of their institute for my 3 months stay. In particular, thank you Cordula Artelt for contributing to my training and Brigitte Finke and Sabine Fischer helping me out before and during my stay.

Sophia, it’s always nice to meet people like you when you are far from home, but it’s even better to have built a real friendship regardless of the distance. Thank you.

A los becarios de investigación Carmen, Nacho, Nadina, Noemí, Marina y Bea, y a las técnico de investigación Laura Royo y Vittoria, por las miradas cómplices en los pasillos, las risas, el cariño y el apoyo a lo largo de esta andadura.

A todas las personas anónimas que, en un avión, un tren o una cena con amigos, me han preguntado “¿a qué te dedicas?”. Gracias a ellos, a su curiosidad, sus preguntas y su sorpresa, le he dado más valor al trabajo realizado, y me han ayudado a reflexionar sobre dónde estoy y dónde quiero llegar a estar.

Gracias al equipo del British School of Valencia por abrirme sus puertas, valorar mi trabajo y experiencia, y por animarme y apoyarme en la etapa final de este recorrido.

A mis amigas, por haber sido un apoyo incondicional en estos años y por haberse preocupado por la evolución de mi trabajo. Por entender que fuera siempre la última en llegar y la primera en irme (los días que iba). Por todos vuestros ánimos e inestimable ayuda. Por cuidar de mí y estar siempre en los momentos más importantes de mi vida. Gracias a todas, os quiero.

José Andrés, gracias por todo tu cariño, comprensión, confianza y apoyo en la fase más dura y complicada de este trabajo. Has sabido sacarme siempre una sonrisa y me has calmado en los momentos más tensos. Me siento muy afortunada de que nuestros caminos se cruzaran y de que compartamos la vida juntos. Te quiero mucho; mucho y bien.

Gracias a mis padres, Ana y José, por vuestro apoyo incondicional, por ayudarme a tomar decisiones, por confiar en mí y en mi valía incluso cuando a mí me costaba hacerlo, por hacer (una vez más) de mis sueños los vuestros y remar conmigo para alcanzar mis metas, por acoger mis altibajos y por tratar de comprender todo lo que os iba contando de mi trabajo. Os quiero mucho.

Finalmente, quiero darle las gracias a Jambo, mi perro, que no se ha separado de mi durante la fase de escritura, me ha dado mucho cariño y me ha obligado a salir con él a despejarme incluso en épocas complicadas de confinamiento y cuarentena ♥.

“I never teach my pupils, I only attempt to

provide the conditions in which they can learn”

Albert Einstein (1879-1955)

INDEX

SUMMARY

The present dissertation aimed to analyze the effectiveness of different types of elaborative feedback messages to help students learn declarative content from a task involving reading a scientific expository text and answering related questions. As a starting point, in Chapter 1 we introduce the topic by presenting the theoretical framework for the basis of this work. We will start by reviewing the issue of learning by answering questions from texts and the most influential theoretical models involved in this matter. Additionally, we focus on the types of questions and their impact on learning according to previous research. We subsequently address the topic of feedback as a teaching tool were we review the types of formative feedback and previous research outcomes. We also explore text availability and response certitude as variables which impact the outcome of instructional feedback. To finish our first chapter, we include a section on how to correct misconceptions according to the recent literature. We will explore several procedures although we will focus on refutational texts and we will discuss how can feedback help in the refutational process during learning. In Chapter 2 we present the current dissertation, explaining the path followed from one study to the next one and the common

procedures among the different studies. We then present our empirical studies in Chapters 3, 4, 5 and 6. The first empirical study of this dissertation (Chapter 3) aims to explore the effectiveness of two types of elaborative feedback vs. corrective feedback when students can decide to search the text for additional information after giving an incorrect answer. Chapter 4 includes a new study that replicates the previous one with a major change: searching was not available at any point of the experimental procedure. For our third study (Chapter 5) we decide to analyze only the effect of elaborated feedback messages and we create this study with three different types of elaborated feedback with the objective of identifying if it is better to direct the information in the feedback message to the right answer, to correct the mistake or using a combination depending on the score. Chapter 6 includes our last experimental study were we analyze the difference between the two most used elaborated feedback messages and we explore how response certitude impacts feedback reading times. Finally, in Chapter 7 we discuss the overall findings from these four empirical studies. Our findings throughout the dissertation outstand the importance of coherence between the students answers and the feedback message given as well as the usefulness depending on the correct or mistaken answer of the student. This coherence effect will help acquiring knowledge tested on a post test. Other variables such as searching and type of question are also discussed. Additionally, we suggest future research directions and we discuss the main practical implications of our findings for education.

CHAPTER 1

THEORETICAL FRAMEWORK

1. LEARNIG BY ANSWERING QUESTIONS FROM TEXTS

Reading and understanding what we read is an essential demand of our society nowadays. Everywhere we go we are meant to understand information set all around us in very different formats and displays with also very different purposes. For example, some daily activities that require reading and understanding are checking a pay sheet, verifying the terms and conditions of a promotion or searching for information about a recent news. As we can see, reading comprehension requires not only understanding single words and sentences but moreover implies constructing and understanding language representations at different levels (phonological, semantic, syntactic, and thematic) (Graesser, 2007).

The changing nature of the reading demands in our society have made a strong impact on research objectives on this area. These changes are driven by the gradual adjustment of the definition of reading literacy though the last decades which has been influenced by the continuous changes in society and advances in innovation. Not that long ago, The United States National Adult Literacy Survey defined it as: “using printed and written information to function in society, to achieve one’s goals, and to develop one’s knowledge and potential” (Kaestle et al., 2001, p.2). For the following years, the OCDE-PISA used a similar definition and it is not until 2009 when PISA reading literacy frameworks took digital reading into account: “Understanding, using and reflecting on written texts, in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society” (OECD, 2018, p 27). Thus, PISA tests assess the ability to properly *understand* and *use* written information for a variety of specific purposes, using diverse textual genres and formats. If we focus on the meaning of *understand,* we can establish a homologous meaning to the traditional concept of reading comprehension (Kintsch, 1998; van den Broek, Young, Tzeng & Linderholm, 1999). The traditional definition of reading comprehension considers the construction of a mental representation of the texts’ content (Kintsch & van Dijk, 1978; Palincsar & Brown, 1984; Van Dijk & Kintsch, 1983) and therefore the activation of the cognitive processes related to comprehension. Additionally, the meaning of *use* refers to employ the textual information for some purpose and for this, readers need to actively self-regulate their reading strategies to make connection between different sources of information. This way, when the learner is reading with and established aim, he creates a mental representation of the task which consists of the relevant information which the learner consideres relevant to attain his objective. (Britt, Rouet, & Durik, 2018; Graesser, 2007; Rouet, Britt, & Durik, 2017; Vidal-Abarca, Martínez, Gil, García, & Máñez, 2019). Hence, reading literacy not only understanding the text or any type of document, but also making use of it by accessing or retrieving information, discarding irrelevant information, or integrating information from diverse sources, among other metacognitive activities.

In educational settings, students face many situations where they have to read in different formats and with different purposes. Students have to be able to understand declarative knowledge from a text, answer correctly questions about it and process feedback messages to gather relevant information which enables them to improve their understanding of specific content (Ness, 2011; Pressley, Wharton-McDonald, Mistretta-Hampston, & Echevarria, 1998).

1.1.Comprehension of texts.

In today's literature, there is no well-established theoretical model to study the relationship established between reading a text and the execution of a tasks associated with it as it may be answering questions about the text. Nevertheless, there has been recent research examining the cognitive processes involved in answering questions from texts (Cerdán, Gilabert, & Vidal-Abarca, 2011; Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gil, 2009; Gil, Martinez, & Vidal-Abarca, 2015; Maña, Vidal-Abarca, Domínguez, Gil, & Cerdán, 2009; Vidal-Abarca, Mañá, & Gil, 2010). The objective of this section is to review cognitive models of comprehension and task-oriented reading, especially in question answering tasks. First, we will describe the basis of the process of comprehension according to the *Construction-Integration* model proposed by Kintsch, (1988). This model explains the origins of how we build a mental representation considering the different informational elements available, although it does not contemplate the resolution of a task. As we cannot directly evaluate the mental representation of the reader and the model does not include a specific task to assess the underlying processes, we will review other theoretical models which have complimented Kintsch's model. Thus, we will review the *TRACE* model proposed by Rouet (2006), which considers the processes that take place during purposeful interaction with complex texts, followed by the *RESOLV* model (Rouet, Britt, & Durik, 2017) focused on the readers' construction and management of objectives during text comprehension tasks. Finally, we will present the *Question-Answering* model (Vidal-Abarca et al., 2019), based on the previous theoretical models exposed before, which considers the main actions and decisions made by the student in question-answering tasks with feedback.

According to the *Construction-Integration model* (from now on, C-I) (Kintsch, 1988) comprehension is the process of building a coherent mental representation of the text information. In order to do this, the reader goes through two phases that culminate with the formation of a mental representation of the text's ideas. On the first phase, *construction*, the reader has to generate a network of ideas regarding the text. The course begins with the reader processing the text by *cycles*, which are units of information with meaning (i.e., a phrase or sentence), that enable the reader to extract *ideas* or *propositions* that connect between then forming the base of the network of ideas. The ideas which are generated by connecting explicit ideas within the text are called *bridge inferences*. This network also includes ideas from the previous background knowledge of the reader which have been activated during the text processing. On the second phase, *integration*, the imprecisely constructed and unintegrated network of ideas is reduced to a set of propositions formed by the most active ideas (i.e., those with more connections) resulting in a mental representation of the text. Both phases, construction-integration, will repeat cyclically as the learner reads the text.

It is essential to understand how information is extracted from the text, but also to understand how the reader interprets this information. There is a general agreement between authors among the three basic levels of representation in comprehension: the surface level, the text base level and the situation model level, ever since it was proposed by Kintsch & van Dijk, (1978; see also van Dijk & Kintsch, 1983). First, the surface level, refers to the literal wording, grammar and syntax that constitutes the text. The reader accesses the literal meaning of the words and constructs a meaning on this basis and considering the grammatical rules. A big limitation of this level of representation is that a reader may consider topical information such as a specific word as important if it appears frequently on the text, but this level it does not yet imply the construction of a network that integrates or relates the different ideas of the text. This results in the second level, the text base, which considers explicit text propositions and small inferences between close ideas from the text as a network ideas and propositions with meaning and inherent coherence. The reader selects, modifies and organizes the propositional elements included in the text, though no part of the reader’s knowledge is included in the model’s textual representation. When these text base elements are combined and integrated with the readers’ general and background knowledge, the third representation level, situation model, emerges resulting in a coherent network of ideas which integrates and intertwines textual ideas and previous knowledge of the reader as high order processes. This way we can accomplish the main goal of comprehension which is to create an accurate situation model (Radvansky, Zwaan, Curiel, & Copeland, 2001) as this involves new knowledge for the learner which may facilitate the reader's process of reconstructing their own knowledge through the integration of the information present in the text.

Thus, the C-I model explains the process of building a mental representation from a text, and the different levels of representation that we can generate depending on the number of textual connections, and between the ideas extracted from the text and those in the background knowledge of the reader. A major limitation of this classic model of comprehension is that it does not consider situations where the reader has to use the text information to do a task, for example, answering questions or writing a recall about the text, which are common tasks in educational settings. We will now revise different models of comprehension that consider the three elements mentioned before: the text, the reader and the task.

The *TRACE* model (*Task-based Relevance Assessment and Content Extraction*) of Rouet (2006), describes the interaction between the reader and the text in a task situation where the learner has to answer questions from a text. According to the model (see Figure 1), when readers uses a text to perform a task, they have two types of resources: information resources and memory resources. On one hand, information resources include the specific task to be performed (i.e. answering questions or making a summary) the texts, and the idea inherent in the answers constructed by the learner. On the other hand, memory resources refer to the mental representation that the reader has built while reading the text, the prior knowledge of the reader activated while processing the text, or the mental elaborations related to the response. While the learner is carrying out the tasks, she will activate the processes based on both types of resources to elaborate the response.

According to the TRACE model, the first step the learner has to do is to examine the requests of the task (step 1) to be able to form a mental model of the assignment (step 2). The reader then will decide if she needs to look up the text to improve her mental representation or if it can be solved with the memory resources and the mental representation obtained while reading the text (step 3). If the reader decides to search the text, she will select the document needed (step 4), process the useful information (step5) and assess its relevance with respect to the requests of the task (step 6). Once the learner considers that the relevant information has been recovered and processed, the learner will elaborate a response model (step 7). If the reader thinks she knows the answer, she will decide not to search and will go directly to step 7. Once the learner has elaborate a response model, she has to decide if the answer fulfills the tasks' requests (step 8). If negative, the learner will go back to step 3; if positive, a final answer considered accurate by the learner will be given and the process will end. It is noteworthy to know that this model has its version for the processing of multiple text documents, the *MD-TRACE* model (*Multiple Document Task-based Relevance Assessment and Content Extraction*) of Rouet and Britt (2011).

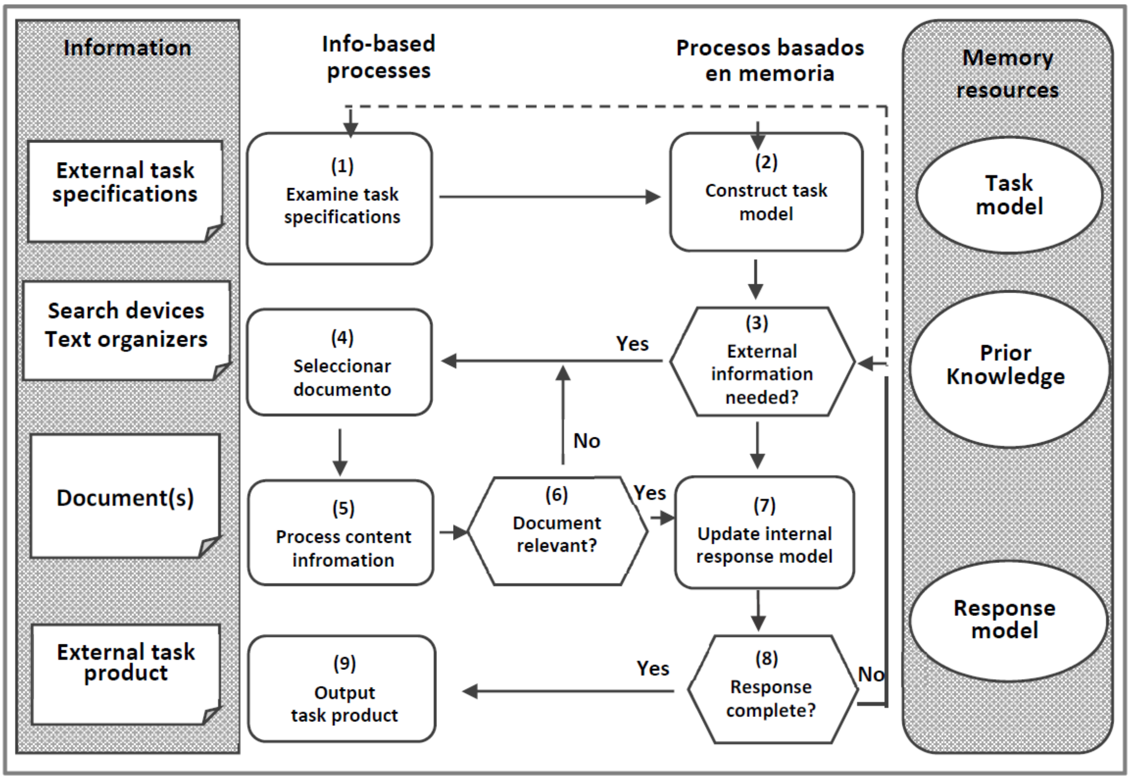


Figure 1. TRACE Model. Adapted from Rouet, 2006, p. 105.

The RESOLV model, REading as problem SOLVing, (Rouet et al., 2017; see also Britt, Rouet, & Durik, 2018) emerges from the need to better understand how readers interpret tasks and how this interpretation has an effect on how readers process the materials. The RESOLV model (see Figure 2) considers reading as an adaptive behavior to reach a specific purpose, set on a specific *physical and social context* with five dimensions which have a direct impact on reading: the request (what is the purpose for reading?), the requester (who is setting this purpose?), the audience (to whom is it directed?), support and obstacles (what resources and information do I have?) and the self (how do I perceive myself regarding this purpose?). Moreover, this model specifies the reader’s resources considered relevant to reading which are four: the preexisting context schemata, knowledge strategies to address them, self-regulation skills, and comprehension and decoding skills. As the previous models, RESOLV describes reading as an activity that contemplates the readers’ representations and processes, defining two types of representations that take part in reading: the *context model* and the *task model*. The context model is formed by the readers’ initial representation of the physical and social context that he considers pertinent. Although it is relatively stable across the task, the context model can be updated in terms of feedback and activity outcome, considering any specific contextual features that where originally considered irrelevant by the reader, as a valuable resource after having failed to achieve a goal (for instance, support materials). Based on his context model, the reader constructs his task model which is his subjective representation of the final objective and the means to achieve it. The task model processes also involve selecting relevant information from the context model, setting plans or actions and reviewing goals, and managing the possible obstacles that may encounter by analyzing the cost-benefits of the actions. Thus, it is frequently updated throughout the reading task and directs the strategic reading behaviors to solve specific reading assignments.

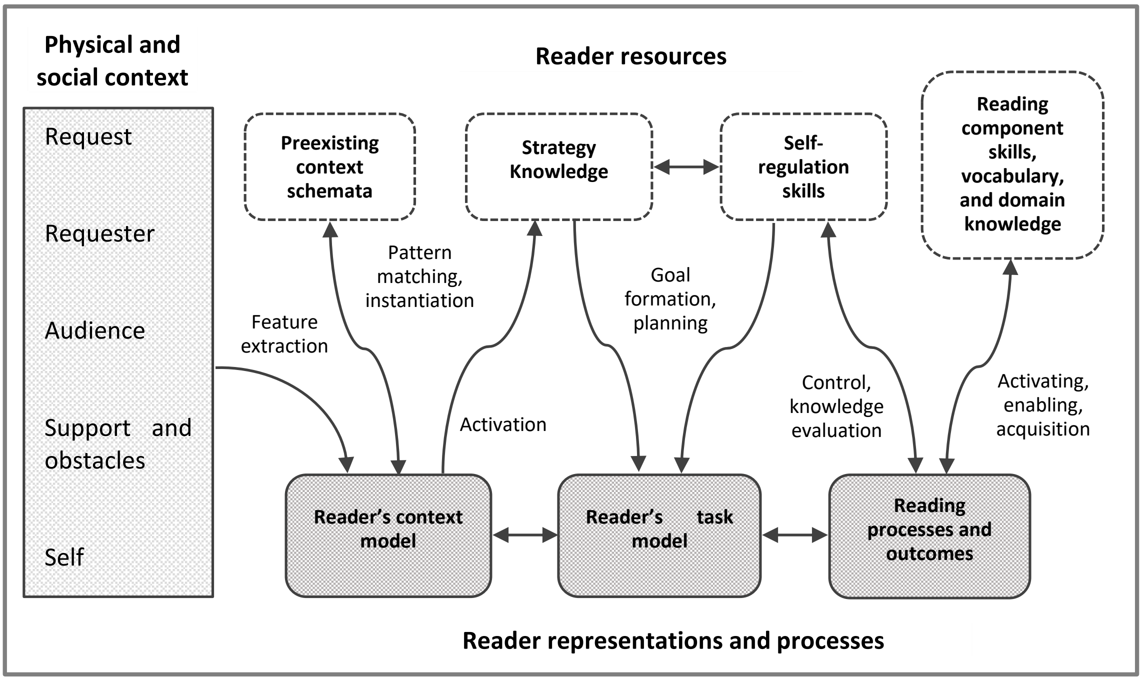


Figure 2. RESOLV Model. Adapted from Rouet, Britt and Durik, 2017

Finally, the Question-Answering model (from now on, Q-A) (Vidal-Abarca et al., 2019), takes into account the main actions and decisions a learner takes when carrying out a task that involves answering learning questions from a text followed by external formative feedback. This model is based on the TRACE and RESOLV model with and additional recursive phase that only occurs when students have external feedback after providing an answer to the task. Our empirical research throughout this dissertation is based on this model. According to the model (see Figure 3), when a learner interprets a task, the first thing she does is create a task model (1a), defined previously by Rouet et al., 2017 as the learners objectives and means to accomplish them. Once the student has organized and integrated his own task model, she is then ready to decide whether she needs to access the text to generate a response or not (1b). This decision implies monitoring processes that occur on the task model already built, which can be modified and updated throughout the task. If the student considers that he is able to give a correct answer with the information she has been able to retain from the text during the initial reading and decides not to search, then an answer is created (1c) and, if the student considers it fits with her task model and she does not have to update it (1d), she will give a response. On the other hand, if the student considers she needs additional information or to review the text in order to give a correct answer, then she will search the text for relevant information (2a). Now the learner must assess what information obtained during the search is relevant (2b) to accomplish the task objective (Mccrudden & Schraw, 2007). This assessment with depend on the task model constructed earlier by the student and will take place in a cyclical manner. The number of cycles will be influenced by the difficulty of the question and by the learners' skills. If she is not sure about the relevance of the information to complete the task, she might need to update the task model (2c) before being able to give an answer by going back to the phase 1a. Once the student gives a response and if there is external feedback available for the student, the final stage of this model starts. When the learner gets external evaluation, independently of the type of feedback given, she will ask herself if there she needs to modify or update her response (3a) and proceed to the elaboration of the new response if yes, or will continue with the next task of finish the task if no.

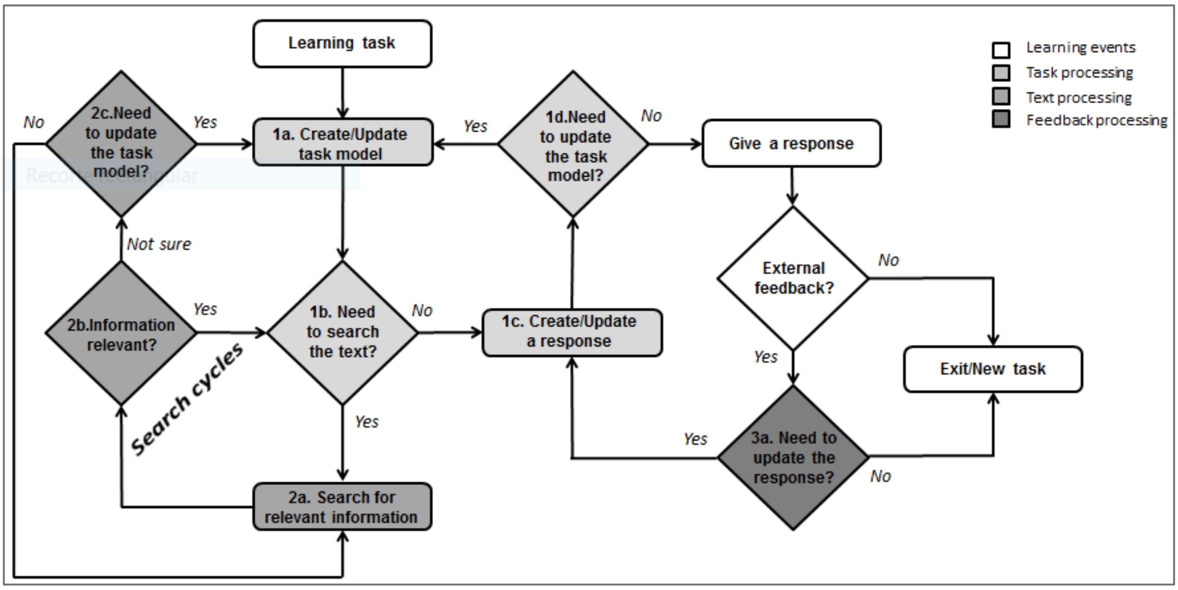


Figure 3. Question-Answering Model, extracted from Vidal-Abarca, Martínez, Gil, García and Mañez (2019).

We will discuss now the four major assumptions underlying the Q-A model regarding students' task processing and the empirical evidence supporting them. Firstly, the model assumes that when the student forms his task model (1a), he forms and integrates the different propositions in the task, existing differences between students depending on their comprehension skills. A recent study by Cerdán, Pérez, Vidal-Abarca, & Rouet (2019) tested this hypothesis. Students had to read two texts and answer five questions for each of them with the texts available. For half of the questions, they received questions aids which included a paraphrases of the question with simplified rewording and key information relevant to answer the question and facilitate the text search, facilitation this way the formation and integration of propositions within the question. This way, the message was more simple as the quantity of ideas was reduced and the most important information was accentuated. For those questions with aids, students were told to read the question and the corresponding aid and after they could search the text and give a response. Results show the beneficial effect of the aids for the less-skilled comprehenders, as they were helped in their specific strategic processes reducing their difficulties in forming and integrating propositions. Hence, forming and integrating propositions are key comprehension processes (McNamara & Magliano, 2009).

Secondly, the Q-A model assumes that deciding whether or not one needs to search information is a monitoring decision. The student assesses his own level of knowledge to answer a question and, after a monitoring process, decides to search or not to search (Gil et al., 2015; Vidal-Abarca et al., 2010). To measure the precision of this monitoring, researches commonly use response certitude (RC) as it is considered and effective method to gather information on the metacognitive assessment of students' performance during the learning process. Vidal-Abarca et al. (2010) found similarities between RC and no-search decisions regarding monitoring accuracy. In compliance with the metacognitive nature of this decision, they report that when students were sure or quite sure of the correctness of their answer they tended to make no-search decisions. Likewise, (Gil et al., 2015) found that the monitoring accuracy of no-search decisions predicted positively and significantly the learners' performance in a task-oriented reading procedure. Findings also confirmed that the number of search decisions was also a significant predictor of the learners' performance. Results obtained by Martínez, Vidal-Abarca, Gil, & Gilabert (2009) show that less-skilled students decided to search more on easier questions, whereas those skilled students increased the searching decisions on complex question.

Thirdly, the model assumes that deciding if the information searched is relevant (2b) operates on text information and depends on comprehension skills. As we explained previously, the task model construction is highly influenced by comprehension skills such as understanding textual material and the question or task. Continuing with the results of the studies mentioned on the previous assumption of the Q-A model, both Gil et al. (2015) and Vidal-Abarca et al. (2010) found that skilled comprehenders self-regulated the use of relevant information more precisely than less-skilled comprehenders, as they observed that skilled comprehenders emitted a correct answer right after reading a piece of relevant text information more often than less-skilled comprehenders. Along with the previous finding, Martínez et al. (2009)found that skilled students do not practically search the text on easy questions, but they do on moderate and difficult ones as they are more flexible with reading behavior, and have greater ability to discriminate difficulty and invest more resources to complex tasks Contrary to this, results showed that less-skilled students do not understand or give up in complex tasks and invest resources on easier question where they can be successful.

Finally, the fourth assumption considers that the learner may have to update the task model (1d and 2c) This can be done in any of the three phases explained in the model: after deciding if the information searched is relevant, and after generating a response in the first attempt and after receiving the feedback message and updating the response. If the learner does not have to update the task model, he will provide a response. A limitation of this model is that the learner’s answer is considered more or less appropriate depending on the context and the coherence standards, as it may vary as a result of the confluence of task difficulty and the learner’s competence since the student behavior is adaptive in this context (Rouet et al, 2017).

1.2. Type of Questions in Question-Answering tasks.

Many researchers have documented the beneficial effect of reading tasks in academic context to promote learning (Baumann & Bergeron, 1993; Block, 1993; Collins, Brown, & Holum, 1991; Dole, Brown, & Trathen, 1996). Ness, (2011) examined instructional practices with the objective to know how much time is spent on reading tasks and which reading strategies are the most used in classroom settings. Results show that reading tasks took place more frequently than any other type of instruction, accounting for 25% of the total instruction time. The reading comprehension strategy most used was question-answering (8.5%) followed by predicting/prior knowledge (6.1%) and summarization (3.4%). Thus, as questions are the most frequent instructional resource used for assessing reading comprehension, it is important that instructors know that different types of questions may have a different impact on the student's learning process.

Not all types of questions are equally helpful for meaningful learning. Learning is considered a cognitive process that requires actively constructing meaning from the text information by accounting on processes of integration and reorganization of current and prior existing knowledge (Chi, 2009; Fiorella & Mayer, 2016; Renkl, 2014; Wittrock, 2010). Questions should assess the different and relevant representation levels of the text and its content. As we explained in the previous section, there are three representation levels: the surface level, the text base level and the situation model level, thus, depending on the task requirements and the student's level of domain, their mental representation will rely on text base or situation model representations (Kintsch & Franzke, 1995; Mannes & Kintsch, 1987).

As we have explained, the type of tasks administered to students when processing a text determines the mental representation they construct, and therefore, their comprehension and learning through the text. On one hand, text base questions (also referred to as low-level questions by multiple authors, and from now on, TB) are those that demand students to process one or various units of information that appear explicitly or implicitly (involving the resolution of small textual inferences) on the text. These type of questions are generally considered easy by students as they produce high performance, although they may limit deep comprehension as they do not promote the processing of multiple units of information at the same time, but they are ideal for learning of simple concepts (i.e., the capital of a country, or the year of an invention). On the other hand, situation model questions (also known in research as high-level question, and from now on, SM) are those that orient students to connect their prior knowledge with the text information, to connect different and spread out units of information and to make inferences. They are considered more difficult and produce lower performance than text base questions in short term, but facilitate the formation of a mental model that help to produce deeper comprehension as it is likely that the knowledge acquired can be applied to a wide range of similar situations. Thus, this shows the importance of using different type of questions to assess the different representation levels depending on the text information and the instructional objectives, i.e., text-base and situation model, representative of superficial or deep learning.

Many authors have studied these two types of questions in a profound manner, agreeing that questions will be more effective when the level of cognitive processing needed to solve those increases. For instance, SM questions that require the learner to make an elaboration from the text information and apply it to solve a new problem are more beneficial than questions that ask learners to make a summary about the information on the text (Cerdán et al., 2009; Hinze, Wiley, & Pellegrino, 2013; Roelle & Berthold, 2017). This results support the evidence showing the effectiveness of using SM questions to improve learning compared to TB questions (Cerdán et al., 2009; Hamilton, 1985; Ozgungor & Guthrie, 2004; Paris, Lipson, & Wixson, 1983; Sagerman & Mayer, 1987). Finally, research and meta-analysis report positive influences of SM questions on student achievement (e.g. Chin, 2007; Hamaker, 1986; Redfield & Rousseau, 1981).

2. FEEDBACK AS A TEACHING TOOL

Learning and understanding complex content from a text may require instructional support to assure the correct learning of new content. This support provided by the instructor, usually comes as a formative feedback message, defined as any informative message provided by a peer, instructor or computer-based system directed at the student's performance(Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mathan & Koedinger, 2005; Narciss, Sosnovsky, & Andres, 2014) to modify cognitive or behavioral processes and improve learning (Shute, 2008).

Among the different instructional procedures, formative feedback is considered one of the most used psychological intervention to improve learning of written texts and one of the top influences on achievement (Cohen, 1985; Hattie & Gan, 2011) as found in a review of 12 meta-analyses (Hattie, 2009). Formative feedback is used on a daily basis on educational settings. As we discussed before, a common task for learning on the vast majority of subjects implies reading a text that contains new concepts, and answering questions that will facilitate the comprehension of the text ideas. When students answer the questions, they will rely on their previous background knowledge and the mental model built while reading the text. It is possible that the new idea has not been correctly understood, or even that there is erroneous background knowledge that interferes with the text information causing difficulties for the students to understand the text and, therefore, it is likely that the student gives an incorrect answer to the question. In this cases, especially after committing an error, receiving feedback from the teachers is a common and useful practice in education (Hattie, 2009).

Vygotsky's defined the concept of the zone of proximal development (ZPD) as the distance between two different learning stages: i) the level of actual development (from now on, LAD) determined by students' independent problem solving strategies, and ii) the level of potential development (from now on, LPD) determined through problem solving under adult guidance and successful instruction (Vygotsky, 1978). According to this definition, we can establish feedback as an instructional tool used to help students reach the level of potential development. Elaborated feedback messages are instructional tools that interact with the learner and can help him identify and repair his misconceptions (e.g., Pashler, Cepeda, Wixted, & Rohrer, 2005) and retain correct responses (e.g., Butler, Karpicke, & Roediger, 2008) in classroom and digital settings (Hattie & Gan, 2011; Hattie & Timperley, 2007; Narciss et al., 2014). Through this dissertation we will discuss the main results found in feedback research until now and we will present a series of studies that examine the processing of elaborated feedback on a digital environment where students have to learn by answering questions about some texts.

2.1. Types of Formative Feedback

Feedback used in educational backgrounds is an instructional tool considered crucial in its role to improve knowledge and skill acquisition (Anderson, Conrad, & Corbett, 1989; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Epstein et al., 2002; Moreno, 2004). There is no consensus on what type of feedback is more effective to promote learning, but researchers agree on a general classification of feedback according to the specificity of the feedback message. Feedback specificity is defined as the level of information presented in feedback messages (Goodman, Wood, & Hendrickx, 2004). In other words, how much information and what type of information is given to the learner with the objective of increasing student knowledge, skills, and understanding (Shute, 2008). There are numerous types of feedback messages that, regarding their specificity level, can be grouped to one of the following: a) verification feedback, b) correct response feedback, and c) elaborated feedback. As we will see, different authors include different types of messages in each of this three categories, although since the beginning of feedback research the objective of instructional feedback is to act on validation of right answers and correction of errors (Kulhavy, 1977).

On the first place, verification feedback is known as Knowledge of Result (from now on, KR) according to the general classification of feedback. KR feedback messages only include information about the correctness of the student's response. They are characterized for being simple messages with corrective matter. The most common way is by telling the student "Your answer is correct" or "Your answer is incorrect", but KR feedback can also be given to the students by, for example, using colors or a checkmark to indicate the correctness of the answer (Shute, 2008). In the literature, we find other terms describing feedback messages which can be encompassed in verification feedback such as *error flagging,* described as highlighting the errors in a solution without giving the correct answer (Corbett & Anderson, 2001). Another type of verification feedback is *multiple try feedback*, also known as “try again” or “repeat until correct”, which informs the student about an incorrect response and allows them more attempts to answer the question (Clariana, 1990, 1993). Although *multiple try feedback* is considered verification feedback regarding the specificity level, Clariana & Koul (2005) found differences between students in the condition and those in normal KR condition. Narciss (2013) includes all the previous verification feedbacks as KR and includes one new type of feedback within verification messages named Knowledge of Performance. This type of message informs the student about how well they performed the task in a global manner by providing summative feedback after a set of questions or once the finished the whole task by, for example, giving a grade or counting the total number of erroneous answers.

On the second place, correct response feedback is known as Knowledge of Correct Response (from now on, KCR) according to the general classification of feedback. KCR feedback messages are also corrective although they include as well additional information explicitly stating the correct answer in a directive way (Shute, 2008). The most common way to give KCR feedback is by telling the student the correct answer (e.g., The correct statement is “c”, or using green highlight for the correct answer and red highlight for the incorrect statements) but other forms, such as giving a sample solution (Narciss, 2013) are also KCR types of feedback.

Finally, Elaborated Feedback (from now on, EF) can be defined as any additional information which is not KR or KCR (Kulhavy & Stock, 1989). EF provides additional information beyond the students’ current performance (Vidal-Abarca et al., 2019). It embraces a wide range of messages (e.g., explanations, worked-out examples or hints) and may include or not KR or KCR feedback (e.g., Van der Kleij, Feskens, & Eggen, 2015). Different researchers have studied different types of EF messages, among which we can find *Topic Contingent feedback* which involves item verification and a general elaboration on the topic currently studied (Mandernach, 2005), and *Response Contingent Feedback* which includes KCR plus additional information detailing why the correct answer is incorrect or why the incorrect answer is not correct (Gilman, 1969; Gilman, 1969; Whyte, Karolick, Nielsen, Elder, & Hawley, 1995). As Shute (2008) points out, the EF usually is directed to the correct answer by giving hints, cues or prompts (although avoiding stating explicitly the correct answer), but it may be directed to correct the misconception by explain the learner's specific errors. From now on we will refer to these types of EF messages as $EF_{Correct}$ and $\mathrm{EF}_{\mathrm{Incorrect}}$, respectively. On our research we will focus on Response Contingent Feedback with $EF_{Correct}$ and $EF_{Incorrect}$ messages.

2.2. Effectiveness of Formative Feedback

Feedback effectiveness on learning has been extensively studied over the last decades, even though previous empirical evidence has found inconsistent findings (e.g., Jaehnig & Miller, 2007; Kluger & DeNisi, 1996; Shute, 2008; Van Der Kleij, Timmers, & Eggen, 2011). Although there is no consensus on what characteristics should a feedback message have to be effective in task-oriented-reading questions (Narciss & Huth, 2004), it is well established that (1) feedback is useful to reinforce correct ideas and to correct errors (Anderson, Kulhavy, & Andre, 1971, 1972), and (2) even the most simple form of feedback should inform about the correctness or not of the answer and combine it in most cases with an elaboration (Mory, 1992; Shute, 2008). Contrary to this agreement, there are no consistent results about the amount and type of information that elaborated feedback messages should include (Kulhavy, 1977; Schimmel, 1988).

Research indicates that EF messages are more effective than corrective ones (Bangert-Drowns et al., 1991; Van Der Kleij et al., 2011; Whyte et al., 1995) In a recent meta-analysis about effectiveness of different types of feedback in digital environments, Van der Kleij et al. (2015) indicate that the average effect sizes for KR, KCR and EF are .05, .32 and .49, respectively. According to these results, EF seems to be the most effective type of feedback because it allows the learner to shorten the gap between her LAD and the desired level of performance or LPD. KCR shows relatively positive results as it includes the correct answer which may help the student to restructure the representations in his mental model. As we can observe, KR on its own is the least effective type of feedback, as only knowing if the answer is correct or not gives very little information to the student about how to improve or modify the mental model of the student. Noonan (1984) conducted an investigation with six different feedback conditions in computer-assisted instruction and found out that the effectiveness of KR increased if it was given along with an additional elaboration which they called “process explanation” (PsE).

Regarding the effectiveness of KCR, Andre and Thieman (1988) examined the effect of factual and application questions and type of feedback: no feedback, KCR and self-correction feedback. Results indicate that KCR feedback facilitate the students’ performance on the questions while they were answering the task, but did not influence concept learning. Regarding application question, researchers found better concept learning when there was one-day interval between the training phase and the post test, than when the post test was done immediately after.

Another study designed with the aim of analyzing the effectiveness of various types of feedback was conducted by Maier, Wolf and Randler (2016). 261 high school students were randomly assigned to one of three different experimental conditions: KR condition, KR plus EF condition and a control condition without feedback. Participants read science text characterized by being well-structured declarative knowledge and answered a set of questions. They received feedback according to the assigned experimental condition. Afterwards participants in the KR plus EF condition were divided into those who used the feedback message meticulously (A) and those who did not use it (B). The results showed that KR and KR plus EF(A) were more efficient than KR plus EF(B) and no feedback. Although there are some limitation regarding this division, we are interested on the results obtained by these researchers when they consider that students process the feedback message, as not using the message implies not engaging on the task.

The complexity of feedback messages has been studied in detail, as EF is characterized by being longer and more complex than correctives types. There is evidence to believe that more information in the feedback message does not necessary mean that the message will help students to perform better. Considering this, Phye (1979) studied 3 types of elaborated feedbacks which varied on the amount of information given. It was found that the feedback message which included the least amount of information produced greater improvement on post-test, accordingly, the message has to be directed and precise. Thus, more information does not have a facilitative effect as it produces an increase of cognitive load (Phye & Bender, 1989). On the same line as the previous studies, Kulhavy, White, Topp, Chan and Adams (1985) conducted a research in an instructional setting were participants had to read a text and answer multiple choice questions about it after each paragraph. Participants were given a feedback message after their answer, which varied in complexity between the different experimental groups, design using an additive strategy. The less complex message was KCR, whereas the most complex messages included information about why it was an incorrect answer, and a hint to the relevant textual idea needed to answer the question correctly. Result demonstrated that a more complex feedback message was not related to a higher performance or to a better efficiency in correcting students' errors. This disinformation about how complex and what type of information has to contain an EF message has resulted in the fact that numerous studies present inconclusive results regarding the effectiveness of EF in enhancing learning in comprehension tasks that require answering questions from a text (Golke, Dörfler, & Artelt, 2015; Llorens, Vidal-Abarca, & Cerdán, 2016; Llorens, Vidal-Abarca, Cerdán, & Ávila, 2015).

For instance, Golke et al. (2015) examined on their Experiment 1 the impact on EF on deep-level comprehension using a computer-based assessment. Participants were 566 sixth graders assigned to one of the five experimental conditions; three of them had different types of EF designed to improve text comprehension and the other two were control conditions (KR only and no feedback). The three experimental conditions included KR information plus an additional elaboration which consisted of an inference prompt to facilitate the construction of inferences (an error explanation) or a monitoring prompt to help comprehension of the relevant text section. The effectiveness of these types of EF was tested by comparing students results across the five conditions. Participants were asked to read 3 expository and two narrative texts and answer total of 37 questions which appeared on the computer screen. All students, except those in the no feedback group, received the assigned messages after an incorrect answer, after which they had a second attempt without any feedback. Students completed a posttest evaluation immediately after completing the test which consisted on reading another 2 passages (one of each type) and answering a total of 14 questions, and follow up evaluation was completed several weeks later which consisted on reading 4 more texts (two of each type) and answering 30 questions. Both the posttest and the follow up test included different texts and questions. The results showed that there were no significant differences in students' performance between the types of EF compared to the KR feedback or the group without feedback. This is explained by the authors making reference to the low commitment of the students to actively process the feedback provided and they highlight the need to account on motivational factors in feedback interventions.

In a recent study, Butler, Godbole and Marsh (2013) point out that many studies (some of them mentioned above, e.g., Anderson et al., 1972; Andre & Thieman, 1988; Phye, 1979) conclude that increasing the complexity of EF feedback message does not have a positive impact on learning compared to the results obtained using KCR. They discuss that these studies assess retention of the right answer of a question shown previously rather than deep understanding and application of complex material. In this study, Butler. et al. (2013) carried out 2 experiments with 60 and 24 university students respectively. Participants read 10 passages 500 words long, with two critical concepts per text. On a post test, 10 of these concepts was assessed using definition question to measure retention and the other 10 were assess using inference questions to assess transfer of knowledge. Students received KCR, EF (which included KCR and two sentences elaborating on the correct answer), or no feedback (control group), according to their experimental condition. The results in their study demonstrate that both types of feedback helped students on questions assessing retention in a similar manner, and significantly better to not receiving feedback. Nevertheless, EF outperformed KCR in new inferential questions. On this basis, the authors suggest that a reevaluation of the impact on EF on learning is required.

A meta-analysis was carried out by van der Kleij et al. (2015) regarding students' learning outcomes depending on different types of feedback in computer-based environments. The results showed that, regarding higher order learning outcomes such as transfer of learning and application of knowledge, EF as more effective than simple feedback (KR and KCR). The authors also outstand the important practical implications for educational software designers and computer-based learning settings.

Additional research with the objective of determining the impact of EF on students' question-answering performance, was carried out by Máñez, Vidal-Abarca and Martínez (2019) with 75 secondary school students. Participants were randomly assigned to a control group or an EF group. Students had two tasks: i) they had to answer a multiple choice question with 3 distractors and one correct answer, and ii) they had to select the information they considered relevant from the text to answer the question. Those students in the EF group were given a feedback message which included KR on the correctness of their answer (i.e., they were told if the alternative they selected was correct or incorrect), information on the accuracy of their selection of information relevant to answer the question, and monitoring hints on relevant strategies for succeeding on the task. After closing this message, they got KCR on their answer to the question and KR on their selection. Those students in the control group had a placebo feedback message (e.g., "You have answered question 1") and no formative feedback message or information was displayed. Results show that students in the EF group outperformed those students in the control (p = .027) and included less non-relevant information in their selection (p = .002).

To sum up, due to the empirical evidence, elaborated types of feedback are commonly preferred to simple ones (Hattie & Timperley, 2007; Shute, 2008; Van der Kleij et al., 2015), at least for tasks requiring higher-order cognitive processes (Bangert-Drowns et al., 1991; Van der Kleij et al., 2015). Research evidences that the EF must contain personalized information for the learners that helps them create, change or restructure their knowledge and beliefs (Butler & Winne, 1995), establishing relationships between student performance and the standard required for the task concerned (e.g. Kluger & DeNisi, 1996; Narciss & Huth, 2004). Nevertheless, understanding of the EF is not an automatic process, since the learner has to be engaged on the task, willing to read the text and trigger the cognitive and metacognitive processes necessary to integrate that information with their previous knowledge (Bangert-Drowns et al., 1991). Thus, in order to be beneficial, students must monitor their need to process it and engage in processing its content. In case they decide to process it, feedback is expected to challenge students' cognition by comparing their performance with the desired level of performance.

2.3. Text availability

The effect of feedback may be influenced by any factor that can be manipulated throughout the learning process, therefore it is important to study the effect of these external factors and its influence on EF messages (Lefevre & Cox, 2016), although there is not many empirical research relating this two variables. One of the factors that can be most easily controlled in digital settings when using feedback to enhance learning is the text availability. This situational variable can have an effect on the use of feedback in task-oriented reading activities. Textual availability refers to the possibility to refer back to the text to search for additional information or to clarify concepts, between others. In this section we will first discuss how reading and comprehension is influenced by text availability, and we will finish outlining the only study done until this date relating both variables.

Different cognitive processes are involved in question-answering tasks depending if the text will be available to consult while answering the questions about the text or not (see Figure 4, adapted from Ozuru, Best, Bell, Witherspoon, and McNamara, (2007)). On one hand, when a student answers comprehension questions without the possibility to search the text during the task, he relies on his previous background knowledge to help in the formation of a textual mental model during the time he is reading the text (Anderson, 1978; Anderson & Pearson, 1984; Kintsch, 1988). On the other hand, when a student answers a comprehension question having the text available, other factors such as engagement, ability to search and willingness to emerge in a more effortful task emerge. This time, the student relies on his ability to restructure the initial mental model formed during the initial reading phase as the re-reading cycles clarify or include more information that was not included in his task model previously. The text availability enables him to update his internal response model (Rouet, 2006; Rouet et al., 2017).

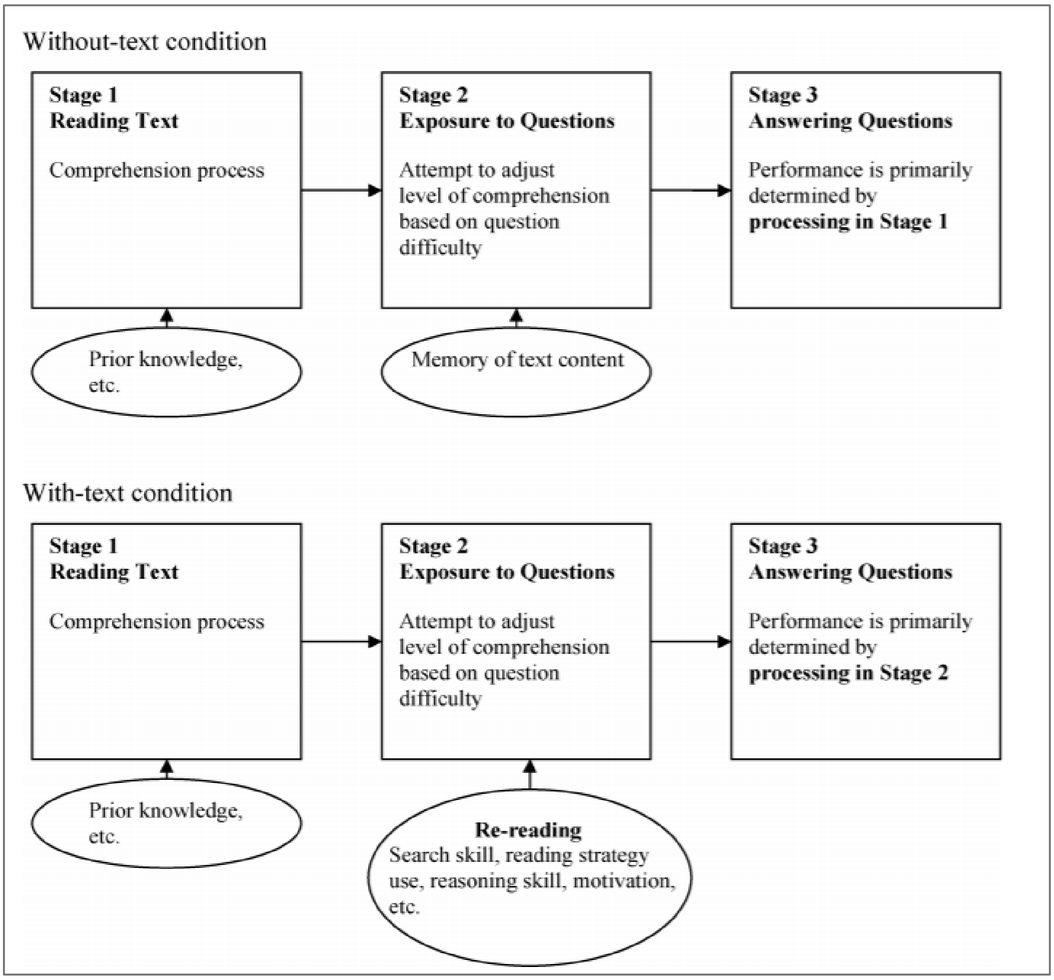


Figure 4. Cognitive processes in question-answering task without text availability and with text availability. Adapted from Ozuru et al., (2007)

Thus, answering questions from a text being able to refer back to it during the task, considers cognitive and metacognitive processes which are not present when the student completes the same task without being able to refer to the text (Artelt, Schiefele, & Schneider, 2001; Ozuru et al., 2007). Consequently, the text model formed by the student has to consider the learning objectives and demands of a given task, as well as its specific characteristics such as availability.

Many researchers have studied the role of text availability in learning. For instance, Ozuru et al., (2007) presented a science expository text to undergraduates. Their task was to read a scientific passage and after, answer question about the content. During the question-answering task, half of the students had the text available so they could refer back to the text to search for additional information or clarify the concepts or ideas involved in the answer before answering, while others had the text unavailable and had to answer relying only on the information retained during the initial reading. Results show that having the option to search on the text improved performance, and that those students without access to the text have similar performance results to chance level.

Likewise, Agarwal, Karpicke, Kang and Roediger, (2008) examined the testing effect with and without access to the text while answering the questions in university students. Results confirm previous findings as text availability led to better performance results than without text availability. The authors also practiced a delayed test, where they found that the beneficial effects of text availability did not persist and both types of conditions produced equivalent retention on a one-week delayed test.

These studies did not contemplate feedback, and there have been few studies that have analyzed the interaction between these variables and you make an introduction to the studies, on what the studies have focused and you already present. The point is to make it clear that available availability studies have rarely included feedback in their studies.

Additionally, having a text available also influences how students use the text information as well as other resources available (Ferrer, Vidal-Abarca, Serrano, & Gilabert, 2017; Higgs, Magliano, Vidal-Abarca, Martínez, & Mcnamara, 2016), as it can be use of instructional EF messages. For example, those students without text availability can only rely on their text model formed on the initial reading, therefore it is more likely that they use any other source of information, such as feedback, more often than students with an available text.

Mañez et al 2017 examined how text availability influenced student's decision to access an EF message when answering multiple choice questions from a scientific text. Half of the students were able to re-read the text for additional information before giving an answer while the other half could not access the text. Results show that those students that had the text unavailable increased the access to EF messages. The authors suggest that this increase is due to the lack of information when answering the question, as the only additional aid they have to enhance understanding is EF. When the text is not available, EF becomes a metacognitive decision influences by the text disponibility, an important factor to consider in the development of digital environments.

Although there has been very few empirical research analyzing the relationship between text availability and feedback, recent results demonstrate that individual metacognitive variables involved in the process of learning are important to consider (Narciss, 2013). The latest research indicates that this variable should be studied in more depths, especially in electronic environments where feedback messages are directed to enhance performance.

2.4. Feedback processing and response certitude

In current feedback frameworks, the learner is considered as an active constructor of knowledge and feedback is contemplated as a tool to provide learners with instructional or tutoring information on their actual state of learning to help them to successfully regulate the learning process (Butler & Winne, 1995; Hattie & Gan, 2011; Hattie &

Timperley, 2007; Maier et al., 2016; Narciss et al., 2014; Shute, 2008). An individual variable widely studied within the framework of feedback research is response certitude (Kulhavy & Stock, 1989), considered a control variable operating at a high level of cognitive organization (Stock, Winston, Behrens, & Harper-marinick, 1989). In this section we will firstly discuss what is response certitude and the general findings related to this variable, and we will finally outline the main findings on the relationship between response certitude and feedback.

According to The Oxford English Dictionary (1933), certitude is defined as: “Subjective certainty; the state of being certain or sure of anything”, and this definition goes in line with the theoretical revision of response certitude in feedback research done throughout this dissertation. Thus, when we study response certitude in feedback research, we are considering a metacognitive judgement of the accessibility of the learner to a specific piece of information at the time he is being asked for his response certitude.

Response certitude has been examined by Kealy and Ritzhaupt (2010), considering that making students rate their response certitude has other beneficial outputs for performance as it makes them engage in the task by actively retrieving textual information and evaluating their answer. This increases the activation of the working memory and strengthens the encoding of the initial response. Additionally, results on their study seem to indicate there is a beneficial effect of response certitude especially when students are able to compare their initial response with the target text for direct comparison. A way to help the students compare their initial response with a more complete answer is using formative feedback as an aid to enhance learning.

Taking response certitude into account, Kulhavy & Stock, 1989 proposed a model to understand how feedback influenced learning from written instruction by combining the learners confidence and the complexity of elaborative feedback messages among other factors. Their objective was to be able to predict the changes from the initial response to the final response as a function of a feedback message. To do so, the authors considered a standard instructional situation including feedback and stated the three different cycles we can find. The first cycle refers to the first attempt to answer the question, the second cycle refers to the elaborated feedback facilitated to the student, and the third cycle refers to the second attempt to answer the question (as shown in Figure 5). In the model, the student states his response certitude on the first cycle before giving a response.

On the first cycle, the learner reads the question and the possible alternatives and makes a cognitive evaluation based on his prior knowledge on his execution on similar tasks, and on the understanding of the studied content for this specific instructional task. In this point, the learner evaluates the different alternatives available to answer the question by comparing the statements between them and with his own mental representation for the task. When this evaluation finishes, the learner produces their initial response, a memorial representation of the response and a response certitude associated to it. On the second cycle, the learner received information on his performance by means of an instructional feedback message which ideally will include verification and elaboration component. The student now makes a cognitive evaluation where he compares the content in the feedback message with his memorial representation and his certitude produced in cycle I. To produce an estimate of discrepancy between the feedback message and the initial response, the student evaluates the degree of match. The bigger the discrepancy, the more effort the learner will employ in aligning the correct response to the information given in the feedback message. To check the magnitude of the discrepancy, the student will first use the information on the verification component of feedback to decide whether or not his initial response is correct. When a student gives an answer with a high response certitude that matches the information on the feedback message, there will be no discrepancy. On the contrary, when a student gives an answer with a high response certitude that does not match the information on the feedback message, discrepancy would be high. The elaborative message will serve as additional information which will be more relevant for the student with high discrepancies. Finally, on the third cycle, the initial question is presented again for a second attempt. Here, the student compares his initial cognitive evaluations (cycle I) with the feedback modification process (cycle II) to produce a final response.

	Instructional situation	**Student's cognitive evaluation**	**Production**
Cycle I	First attempt	Compares demands to experience and content	Initial response Memorial representation of the response Response certitude
Cycle II	Feedback	Compares verification feedback with initial certitude and memorial representation of the response	Internal discrepancy measure
Cycle III	Second attempt	Compares the initial answer to the final response.	Final response

Figure 5. Summary of Kulhavy and Stock's (1989) response certitude model for instruction with feedback.

Considering the explanation of Kulhavy and Stock's model (1989), we can understand the authors' consideration of the metacognitive variable response certitude as a predictive index of comprehension. When the learner states a high response certitude, this suggests that he is able to construct a relationship between the task model and his prior knowledge, understanding the demands and considering he is able to meet the requirements. On the contrary, when a learner states a low response certitude, it suggests that he is unsure to be able to meet the task demands accounting on their task model and prior knowledge.

Research has also been directed to explore the contribution of feedback to improve the predictive value of response certitude. On experiment 1, Butler et al. (2008) found that feedback is very useful when students respond correctly to questions made with low response certitude as it helps in the correction of the initial metacognitive error and it boosts retention of this unsure correct answers. Additionally, on experiment 2 the same authors found that feedback improved retention as students were able to answer correctly on a posttest questions which they answered erroneously on a first attempt, increasing this way the total number of correct responses. Furthermore, they found that feedback can help students make more accurate predictions of their response certitude by diminishing the discrepancy gap between their perceived response certitude and the correctness of the response.

Hence, previous literature shows that feedback is not only useful to correct memory errors, but it has suitable effects to correct metacognitive errors, especially when the student expresses high response certitude (Butler et al., 2008; Butterfield & Metcalfe, 2001; Fazio & Marsh, 2009).

3. CORRECTION OF THE STUDENT'S MISCONCEPTIONS

Researchers have faced a big challenge reviewing how students construct scientific knowledge. When they face new topics, they already have previous conceptions which are, sometimes, erroneous and this makes it harder to acquire and integrate the new knowledge (e.g., Carey, 2009; Chi, 2005; Kendeou, Walsh, Smith, & O’Brien, 2014; Novak, 1988; Trumper, 2001).

The construction of a correct mental model of scientific content normally requires completing and repairing their misconceptions until achieving an adequate mental representation, especially when dealing with complex concepts or when students may have erroneous previous idea (e.g., Gadgil, Nokes-Malach, & Chi, 2012; Kendeou et al., 2014; Posner, Strike, Hewson, & Gertzog, 1982). Repeated studies have shown that it is not an easy task, as mistakes are difficult to repair and hinder students from acquiring accurate knowledge, especially when dealing with scientific content (Carey, 2009; Chi, 2005; Graesser, Léon, & Otero, 2002; Novak, 1988; Vosniadou & Brewer, 1992).

Although it has been shown that misconceptions are extremely resistant to modification (Dole & Smith, 1989; Maria, 2000), research proves that they can be transformed by means of a process called conceptual change (Posner et al., 1982). From a classical point of view, conceptual change was described as the process of correcting or replacing conceptual errors through exposure to more accurate information (see Özdemir & Clark, 2007; Vosniadou, 2013). Conversely, the actual perspective suggest that new scientific knowledge serves to mask, rather than replace, one’s initial intuitions (Dunbar, Fugelsang, & Stein, 2007; Legare, Gelman, Wellman, & Kushnir, 2008; Shtulman & Valcarcel, 2012).

Experts in scientific instruction have rejected the belief that students are “empty vessels” ready to be filled with knowledge, rather, they consider pre-instructional theories that can interfere with the learning process (Carey, 2000; Keil, 2011). Additionally, many science education researchers see the purpose of instruction as helping students to activate correct new acquired ideas (Caravita & Halldén, 1994; Clement, 1993; Dunbar, Fugelsang, & Stein, 2007; Legare & Gelman, 2008). These proposals emphasize the inferential competition between theories rather than the replacement of one theory by another (e.g. Ohlsson, 2009; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). It is considered as a process of competition between the different mental representations at a deeper cognitive level. Therefore, the conceptual change depends on the probability that the student activates and uses mostly elaborate schemas rather than simple and intuitive ones (Potvin, Masson, Lafortune, & Cyr, 2015; Ramsburg & Ohlsson, 2016).

When students face an instructional task covering scientific content, they face a double task. Students must be able to construct new and elaborate representations through the information received which has to be enduringly encoded in memory, while simultaneously, they must be able to recognize and inhibit the oldest, imprecise or incomplete conceptions and retrieve the new correct knowledge (Hynd, Alvermann, & Qian, 1997; Hynd, McWhorter, Phares, & Suttles, 1994; Kendeou et al., 2014; Mason & Gava, 2007). The instructional response to this problem has been diverse and research has examined different procedures to deal with misconceptions and to promote revision of inaccurate ideas (van den Broek & Kendeou, 2008), as one of the major challenges in education is shifting the activation of incorrect knowledge existing previously (Chi, Slotta, & De Leeuw, 1994; Guzzetti, Snyder, Glass, & Gamas, 1993; Vosniadou & Brewer, 1992). A meta-analysis conducted by Guzzetti, Snyder, & Glass (1992) identified that the main interventions that promoted conceptual change were refutation text, the learning cycle, analogies, and discrepant events.

Hence, the instructional tool most studied to deal with conceptual change is *refutational texts* (Sinatra, Broughton, Diakidoy, Kendeou, & Van Den Broek, 2011; Tippett, 2010) The advantages of these refutation texts over the traditional expository texts has been proved (e.g., Braasch, Goldman, y Wiley, 2013; Diakidoy, Kendeou, y Ioannides, 2003; Mikkilä-Erdmann, 2001), however, it involves developing texts which its own internal structure refutes the most common misconceptions instead of dealing with the erroneous ideas each individual student might have. We will look in more detail to refutational texts on the following section of this chapter. After this, we will consider feedback oriented to correct the error as an instructional tool which is effective in the correction of misconception following the guidelines and assumptions inherent in refutational texts.

3.1. Refutational texts

In scientific literature, the problem of conceptual change has been mainly addressed through refutational texts (Kendeou, Braasch, & Bråten, 2016; Sinatra et al., 2011; Tippett, 2010). In this instructional procedure, the learning or conceptual change is carried out through the previous activation of the supposedly erroneous or partial knowledge that the novice students have, in order to discuss and/or correct that error and to provide them with a more correct and complete explanation of the idea after reading the text (Hynd, 2001).

Refutational texts have not been the only way to address the issue of misconceptions and conceptual change. Another group of work was aimed at confronting the students, individually or in groups, to face errors or contradictions that they had to solve (eg. Asterhan & Schwarz, 2007; Howe, Tolmie, Duchak-Tanner, & Rattray, 2000; Schwarz et al., 2000). Another studied instructional tool is solved examples, were the student must explain why the procedure followed for resolution is correct or incorrect. The explanations generated by these verbalizations, both in the correct and incorrect exercises, have proved to be an effective mechanism for conceptual change, immediate learning and transference measures (Calin-Jageman & Ratner, 2005; Siegler, 1995; Siegler & Chen, 2008) yet this instructional technique is mostly applied on mathematics and physics (Schworm & Renkl, 2002) as not every domain can be taught with examples. Nevertheless, although these procedures have produced some positive effects, these have not been consistent across studies and have also shown unstable results (Limón, 2001; Ramsburg & Ohlsson, 2016).

The Knowledge Revision Components Framework (from now on KReC, Kendeou & O’Brien, 2014) explains the mechanisms by which incorrect ideas become less active or practically inactive in the mental representation of the student, while new correct ideas become dominant. It collects five principles that explain the necessary conditions needed for knowledge revision. The KReC starts with the *Principle of coding* which suggests that once the information (correct or incorrect) has been encoded in the long term memory (from now on, LTM), it cannot be deleted although its activation can be reduced (Kintsch, 1988; Ratcliff & McKoon, 1988). Therefore, it can be potentially reactivated and be part of subsequent comprehension or learning processes. Accordingly, the *Principle of passive activation* states that information in the LTM is automatically activated by the mechanism of propagation of the activation. This implies that any information stored in the memory can be potentially reactivated, independently of its suitableness for learning and understanding. Next, the *Principle of co-activation* remarks that the incorrect information and the new correct information must be activated at the same time for the revision to take place (Kendeou, Butterfuss, Kim, & Van Boekel, 2019; Kendeou, Smith, & O’Brien, 2013; Kendeou & Van Den Broek, 2007; O’Brien, Cook, & Guéraud, 2010; van den Broek & Kendeou, 2008). Following, the *Principle of integration* suggests that the review of knowledge can only occur when the new correct information is integrated with the erroneous information stored in the memory. Finally, the *Principle of competing activation* states that as the new correct information increases its activation, this information will compete with the old incorrect one and will come to dominate the integration process. Thus, achieving the coherence effect, or in other words, fitting in the learner’s mental representation is essential to assure the activation. The activation of the old information (erroneous) will decrease when knowledge revision has taken place. Van den Broek & Kendeou, (2008) claimed that engaging in deep cognitive processes is crucial for conceptual change to take place.

As Beker, Kim, Van Boekel, van den Broek, & Kendeou (2019) synthesized, refutational texts must contain three features which are fundamental and characteristic. First, there must be an explicit statement on the incorrect idea. Second, there must be there must be and explicit refutation of this erroneous idea. Third, there must be some kind of explanation of the correct idea (Kendeou et al., 2013, 2014). ¿Are these features what make refutational texts better than expository texts in promoting conceptual change?

As we outlined before, there are consistent empirical results that show the advantages of refutational text over the traditional expository texts (e.g. Braasch, Goldman, & Wiley, 2013; Diakidoy, Kendeou, & Ioannides, 2003; Mikkilä-Erdmann, 2001). If we analyze the internal structure of traditional expository texts, we can observe that this type of text generally explains definitions and concepts. Moreover, they address many different subjects in education, including science, and students many times find them difficult to understand (McKeown, Beck, Sinatra, & Loxterman, 1992). This has a negative effect on knowledge acquisitions and it does not help students that have misconceptions as it does not induce to conceptual change (McKeown et al., 1992). Conversely, refutational texts encourage conceptual change of those partial or erroneous ideas as its own internal structure activates the common erroneous ideas, explicitly refutes them, and clearly mentions the correct conception or idea(Guzzetti et al., 1993; Hynd, 2001). Furthermore, a meta-analysis of the research comparing refutational text and expository texts as learning tools in science classrooms showed that refutational texts were more effective in favouring conceptual change (see Guzzetti et al., 1993).Still, its effect on learning has not been systematically demonstrated (Hynd & Guzzetti, 1998; Mason, Gava, & Boldrin, 2008; Palmer, 2003).

It is interesting to consider the recommendations of use that Nussbaum, Cordova and Rehmat (2017) gather from previous research in refutational texts which can be applied to feedback to foster learning. The first recommendation is to activate student's prior knowledge by, for example, asking them some prior questions related to the topic. A second recommendation is to deal with simple misconceptions instead of dealing with a large topic. Finally, another recommendation is to combine the instructional tools to refute ideas between the by, for example, combining refutational texts with elaborative feedback and refutational modelling.

It is crucial to understand that knowledge revision will take place whenever there is new information that does not fit in the schemas and mental representation of the learner, and so the previous information has to be understood as incorrect to facilitate the activation and retrieval of the new correct information. As we saw in chapter 1, the models of reading comprehension we reviewed and many other models, consider the learner as an active individual that is continuously integrating, revising, updating and outdating the information in his mental representation (Gerrig & O'Brien, 2005; Graesser, Singer, & Trabasso, 1994; Kintsch, 1988; Kintsch & Van Dijk, 1978; Magliano, Trabasso, & Graesser, 1999; McNamara & Magliano, 2009; Rapp & van den Broek, 2005; Rouet, 2006; Rouet et al., 2017; Tzeng, Van Den Broek, Kendeou, & Lee, 2005; Vidal-Abarca et al., 2019; Zwaan, Magliano, & Graesser, 1995).Nevertheless, research in this area has found contradictory results. On one hand, the majority of the findings suggest that refutation texts are an effective instructional tool to promote conceptual change and deal with misconceptions(Ariasi & Mason, 2011; Broughton, Sinatra, & Reynolds, 2010; Guzzetti et al., 1993; Mason & Gava, 2007; Sinatra et al., 2011; Tippett, 2010). On the other hand, other authors have compared both refutation and non-refutational texts without finding significant differences in performance between them (Alvermann & Hague, 1989; Hynd & Guzzetti, 1998). These findings may be due to the different types of refutation texts used, different knowledge domains tested, different age groups, and different methodological approaches such as eye tracking (Ariasi & Mason, 2011; Mikkilä-Erdmann, Penttinen, Anto, & Olkinuora, 2008), think aloud protocols (Broughton et al., 2010; Kendeou & Van Den Broek, 2007; Mccrudden & Schraw, 2007) or reading times (Kendeou et al., 2014; Yazbec, Borovsky, & Kaschak, 2019).

Presumably, the effectiveness of refutational text resides, among others, in the capacity to induce deep engagement and critical thinking due to the cognitive dissonance between the individual's misconceptions and the new correct information (Guzzetti et al., 1993; Mason, Gava, & Boldrin, 2008). Contrary to this, Hynd (2001) suggested that although these texts have been found to be effective, they are not ideal as they may not promote deep engagement and critical thinking. She explained that the inherent structure of these texts is too explicit as it tells the students what to think and what without giving a change for exploration and profound processing. In other words, the structure of this texts do not require the students' attention, so it may be easier to recall but not to learn or transfer the learning to other situations or environments.

Despite the existence of these contradictory results, there is enough evidence to state that students, when faced with a refutation text, activate a series of strategies that facilitate the processes of correction and reconstruction of information (Diakidoy, Mouskounti, Fella, & Ioannides, 2016; Panayiota Kendeou et al., 2014; Sinatra et al., 2011; van den Broek & Kendeou, 2008) that tend to promote their learning process and conceptual change. Some studies have found that this changes are more likely to be conserved if conceptual change has taken place through a refutational text (Mason & Gava, 2007).

Of course, other instructional tools have been studied to deal with conceptual change. This is the cased of solved example procedures, were the student must explain why the procedure followed for resolution is correct or incorrect. The explanations generated by these verbalizations, both in the correct and incorrect exercises, have proved to be an effective mechanism for conceptual change, immediate learning and transference measures (eg. Calin-Jageman y Ratner, 2005; Siegler, 1995; Siegler y Chen, 2008), yet this instructional technique is mostly applied on mathematics and physics (Schworm & Renkl, 2002) as not every domain can be taught with examples.

Additionally, refutation in non-textual procedures have also been empirically tested in the lasts decades. With the objective of evaluating the most appropriate method for facilitating conceptual change, Frède (2008) tested 2 common scientific misconceptions in pre-service elementary teachers by comparing three instructional methods. One was a simple expository text, another one was a refutational text and the last one was refutation modelling activity by testing the hypothesis of the scientific explanation and of the misconceptions. Results indicated that both refutational methods obtained significantly higher immediate posttest scores than reading an expository text. Moreover, one month later they conducted a delayed posttest were they found that the refutational modelling group obtained significantly higher scores than the two textual methods, and the refutational text obtained better results than the expository text.

A previous study conducted by Hynd et al. (1994) compared the instructional effect of reading a refutational text, watching a refutational demonstration and participating in a discussion group by peers. They found that the refutational text was a better tool as it was the only one that improved scores significantly. These researches put into manifest that some other refutational techniques different to textual refutation might as effective or more than traditional refutational texts, while others are not a good instructional tool to deal with misconceptions. Besides, it evidences that the instructional tools that share the underlying cognitive processes and the general principles and assumptions of refutational texts merit further research.

CHAPTER 2

THE CURRENT DISSERTATION

THE CURRENT DISSERTATION

The general objective of the present dissertation is to analyze the effectiveness of different types of elaborative feedback messages to help students learn declarative content from a task involving reading a scientific expository text, in which misconceptions may exist, and answering related questions.

Considering the task our participants will face in the following 4 studies of the present dissertation, we will explain how the assumptions inherent in refutational texts could be transferred to EF messages. Although each experimental study has its own characteristics which will be explained in more detail in the *Method* section of the corresponding chapter, we will briefly outline the experimental procedure to explain the use of refutations in EF. Students will read a scientific text on, for example, atmospheric pressure followed by a set of multiple choice questions of 4 alternatives that they will have to answer without referring back to the text. After validating their answer, they will receive a feedback message. If the answer is correct they will continue with the task and if it is incorrect they will have a second attempt. After answering the second attempt, they get KR feedback. Thus, if we focus on the experimental procedure when the student has committed an error and has to deal with an erroneous idea and favor conceptual change, the feedback message oriented to correct the misconception ($EF_{Incorrect}$) somehow meets the principles and assumptions established in the KReC framework.

First, following the principle of coding, the learner already has an erroneous idea encoded in the LTM that becomes reactivated during the learning process. Continuing with the example on atmospheric pressure, it may be "*Air pressure increases with the elevation on the surface of the earth*", as found by Dilber (2011) as one of the three most common misconceptions identified from the assessments of students' test responses in the topic of atmospheric pressure. According to the principle of passive activation, this statement can be potentially reactivated in an automatic manner when reading the text or when answering the following question:

> *If you try to replicate Torricelli's experiment on top of an 8,000-meter mountain, rather than at sea level where the original was made, what do you think will happen?*
>
> *a) More mercury will come out of the tube into the bucket, dropping more than 760 mm high.*
> *b) Less mercury will come out of the tube into the bucket, lowering less than 760 mm high.*
> *c) The same amount of mercury will come out of the tube into the bucket, remaining 760 mm high.*
> *d) Due to the height, the mercury will almost completely come out of the tube.*

Although the correct answer is statement *a*, those students with the erroneous conception we stated before will select option *b* and will receive the $EF_{Incorrect}$ associated to that statement: *"You failed, if the column of mercury went down less than 760mm you* *would be affirming that more mercury remains inside the tube and for this to happen, the mercury in the bucket should be under greater pressure.”*

Consequently, when a student is answering a multiple choice question and he answers wrongly, $EF_{Incorrect}$ gives an explanation of why his choice is incorrect. This creates a setting were we can find that the wrong choice is active at the same time as the EF which explains why this chosen statement is wrong, following the principle of co-activation. When both conceptions are activated at the same time, the student receives contradictory information creating a cognitive conflict (McCrudden & Kendeou, 2014; van den Broek & Kendeou, 2008) that the student will have the need to solve.

The information in the EF message will activate the error and facilitate the construction of a new response. When he is unable to do this, he will have the chance to search on the text before giving a second answer to facilitate the students’ selection of the correct answer on the second attempt. At this point, we can find that there is an explicit statement of the incorrect idea, an explicit refutation in the $\mathrm{EF}_{\mathrm{Incorrect}}$ message about the incorrect idea, and the activation of the correct idea, following the three fundamental and characteristic features of refutational texts exposed by Beker et al. (2019) and accomplishing the principle of integration. This sequence will increase the activation of the new correct information to favor conceptual change and the principle of activation will be met. The principle of co-activation will therefore not be met when students receive only KR feedback, as they will not receive at any point information which explains them why did they score incorrectly. This will also happen with students which received $EF_{Correct}$ as they will be given information about the correct idea instead on an explanation about their error. In pilot studies, we tested a type of feedback were, after giving an erroneous answer, students received the information on $EF_{Incorrect}$ plus the information on

$EF_{Correct}$, but results showed it did not function adequately, probably due to its complexity and length (Shute, 2008).

The explanation component in the KReC framework is the structure that determines knowledge revision (Kendeou et al., 2014). Likewise, the explanation students receive when they obtain $EF_{Incorrect}$ is comparable to the one in the refutational text, and when this explanation is integrated in the students' mental representation, the information becomes more accurate (Beker et al., 2019; Kendeou et al., 2019; Kendeou, Butterfuss, Van Boekel, & O'Brien, 2017; Panayiota Kendeou et al., 2013, 2014; Van Boekel, Lassonde, O'Brien, & Kendeou, 2017). Therefore, $EF_{Incorrect}$ can be considered as another form of refutation through feedback.

In order to give response to this broad objective, we carried out four studies with specific objectives. All of the studies included, at least, two types of EF messages. The first one, $EF_{Correct}$, gave students information oriented to the correct answer, such as a hint, or a clue. The second one, $EF_{Incorrect}$, explained the student what specific information of the chosen statement was incorrect. The objective of the first study was to compare the two types of EF with a control group with only KR feedback and analyze which type of feedback was more efficient. As we considered that text availability on the second attempt was making a strong impact on the experimental conditions, we decided to carry out a second study. The objective of the second study was to replicate study 1 to examine if there are different results when the text is not available for searching, as the differences in performance could be only due to the influence of feedback. To increase the power of the results, we decided to add another text for study 3 and focus only on three EF conditions: $EF_{Correct}$, Incorrect and $EF_{Combined}$ which showed $EF_{Correct}$ feedback when a student gave a correct answer and $EF_{Incorrect}$ feedback when the answer selected by the student was incorrect. This was done to show coherence between the mental representation of the student and the message shown. Finally, our fourth study wanted to examine in more depth feedback processing times, so experiment 3 was replicated with response certitude as an additional variable.

It is especially relevant for education to study what type of feedback helps students more not only to understand a text and perform better, but to be able to transfer the ideas learnt in the text to similar replicable situations. As we have previously pointed out, in educational setting, and specially in science education, it is very common to learn from reading an expository text and answering questions about it (Ness, 2011). Thus, what happens when a student gives a correct or a wrong answer will help the student to consolidate what he understood from the text or to correct the erroneous conception he has built (Narciss, 2008; Shute, 2008). As we have explained before, feedback has been widely studied although there are no consistent results among researchers on its effectivity, probably because of the enormous variability of types of feedback and experimental situations tested in different studies, although those feedback messages with additional information (EF) are considered the most effective type (Shute, 2008; van der Kleij et al., 2015; Butler et al., 2013). In the following studies, we attempt to give response to what type of EF is more efficient and what conditions does this feedback must have to be able to consolidate knowledge when a student gives a correct answer, and deal with the erroneous conception by refuting their idea after an incorrect answer.

CHAPTER 3

STUDY 1. EFFECTIVENESS OF INSTRUCTIONAL FEEDBACK: THE ROLE OF SEARCHING IN CORRECTIVE AND ELABORATED FEEDBACK MESSAGES

1. OBJECTIVES AND HYPOTHESES

The main objective of this study was to analyze different types of feedback and their influence on learning. Thus, our aim was to compare the instructional effects of two types of task centered EF with KR feedback. The first type of EF was aimed at orienting or facilitating the correct response and, the second type of EF was addressed to refute the mistakes in each alternative of the multiple choice questions whereas the control group only received KR feedback to inform students about the correctness of the response.

Our main hypothesis is that the group that receives information explaining their error will be more successful in acquiring correct knowledge. Even though in traditional environments feedback is usually aimed at facilitating students to acquire the correct knowledge, it might not be the most appropriate strategy after the students have made a mistake. When an error has occurred, the students need to repair and understand the problem, but this is not so simple, because students may be unable to manage the error themselves. If immediately after making a mistake the students get information oriented to the correct idea, this could interfere with the reparation process, since there is coactivation of this information together with the erroneous scheme previously activated to respond. This information, more than helping, could be overloading their working memory, hindering learning. Another alternative could be to give information related to their mistake, more coherent with the repair process. During the learning process, error reparation takes place in consecutive phases. To starts, the student must activate the erroneous information and its refutation, along with an explanation, in order to understand the error. Subsequently, the correct idea must be activated (Diakidoy & Kendeou, 2001; Diakidoy et al., 2003; Kendeou & O’Brien, 2014; Kendeou et al., 2014; van den Broek & Kendeou, 2008). When the student has been able to successfully repair the erroneous idea, this one will activate faster on his memory building a new and more founded response. Regarding our main objective, our main predictions are:

First, we expect an effect based on the *usefulness* of the feedback; students will be more interested in EF's messages after failing a question, especially when the message has information relevant to a second attempt. In this sense, we expect that students spend more time reading the feedback when they fail than when they get it right, except in the GC, since the message does not contain any relevant information further to the KR.

Second, we also expect an effect based on the *coherence* of the feedback with the last response given by the student. The messages will be easier to process when they are more related to their answer. In this study, the coherence varies between groups, depending on whether responds correctly or incorrectly. Thus, students in $EF_{Correct}$ will receive a more coherent message after a correct answer, while those in $EF_{Incorrect}$ will be a more coherent feedback after they have failed. Thus, we predict that after a success the students in the $EF_{Incorrect}$ group will spend more time reading the feedback than the students in $EF_{Correct}$, and the opposite will happen when students select an erroneous choice and give a wrong answer.

Third, we expect an interaction between these two effects, *usefulness* and *coherence* of the feedback message. We expect large difference in time processing feedback between success and failure in $EF_{Correct}$, as the effects of usefulness and coherence act in the same direction, as after a correct answer, the feedback message will be read faster as it will be perceived less useful and will be processed in less time as it is more coherent, while after an error the feedback message will be read slower as it will be perceived as more useful and will be processed in more time as it is less coherent. However, in the $EF_{Incorrect}$ group this difference should be moderate, as both effects act in opposite directions, as after a correct answer, the feedback message will be read faster as it will be perceived less useful and will be processed in more time as it is less coherent, while after an error the feedback message will be read slower as it will be perceived as more useful and will be processed in less time as it is more coherent.

Forth, we expect that students who have been given information regarding their mistake will perform better on a second attempt than those who have been given information about the correct answer but have not been helped to understand their mistake or refute an erroneous idea. Thus, we predict that scores on the second attempt will be higher for those students in $EF_{Incorrect}$ condition.

Our second objective was to analyze how students use these elaborative messages when the text is still available for review during the second attempt. Rereading is an extra aid that students have the option to use when the feedback received (automatic and immediate) is not enough to solve their doubts. Therefore, we can expect that the availability of the text affected their performance but, it should not affect the usefulness and coherence of the feedback and their processing. According to this second objective our predictions would be:

First, students will need to revisit the text more often after a wrong answer. In this sense, we expect a greater rereading time when EF turns out to be more incomplete, or particularly, when they have not received any EF. Therefore, we expect longer rereading times in CG, since the other two groups have the information given in the feedback message. We do not expect differences in rereading times between the other two experimental conditions. $EF_{Correct}$ is more difficult to process, so they will have more doubts and need to reread more, but, at the same time, the information in the text is less relevant, as they have already been able to access the key ideas related to the correct idea in the feedback presented to them. The opposite will happen in group $EF_{Incorrect}$, the feedback will make it easier to understand the error and therefore reduce the need for rereading, but the information that can be found in the text could be more relevant, because it is different from the information received in the feedback.

Second, the possibility of consulting the text will improve the performance of the students during the second attempt, although not equally in all three experimental conditions. The group that would benefit most from re-reading the text should be the CG since they have not received elaborated explanatory feedback. Re-readings could be considered a similar aid to $EF_{Correct}$ as they can access information related to the correct answer, but no explanation of why their answer was wrong is provided to the students. Nevertheless, we do not expect equal performance between $EF_{Correct}$ and CG for four reason: a) firstly, there is no obligation to reread after an error. Some students may try to give a new answer without searching, whereas the feedback was always given automatically to all the students enrolled in the experimental condition, b) students who receive $EF_{Correct}$ can also reread the text and improve their performance, c) students in CG have to locate the relevant information and discard the non-relevant one, while the students in $EF_{Correct}$ received relevant information directly, d) the text did not include any refutative information to help students to fix the mistaken idea or concept, this information is only available for $EF_{Incorrect}$. We expect that students in the CG spend more time searching on the text, followed by those in the $EF_{Correct}$ condition, and less time searching made by those students enrolled in the $EF_{Incorrect}$ condition.

Our third objective was to analyze how students use the feedback messages depending on the type of question as the study contains TB and SM questions. In this sense we expect that the EF messages will be more useful for SM questions, as they require a greater degree of reflection (Fenesi, Sana, & Kim, 2014; Gibbs & Simpson, 2005). Our preliminary hypothesis is that students will spend more time reading the feedback in the SM questions, although we believe that this effect will depend on the *usefulness* of the feedback, which will again produce an interaction with the type of feedback received. The usefulness effect suggests that students will be more interested in EF messages for question that require higher cognitive processing to succeed in the task.

Finally, we wanted to test if the usefulness and coherence effect will have an impact on post-test performance scores, as we expect that those students in $EF_{Incorrect}$ condition will outperform those is the other two experimental conditions due to the content of the feedback message.

2. METHOD

2.1. Participants

Participants were 128 first-year undergraduates enrolled in teaching studies (18.1% male and 81.9% female). A total of 9 additional participants were removed from the sample due to: (a) they did not complete one or more of the tests (*n* = 5), or (b) there were technical problems in the data recording (*n* = 4). All participants spoke Spanish as their native language. The study was set as a complementary class activity for those students enrolled in Science Teaching and those who participated in the study received a bonus mark on their final grade of the subject. Students' confidentiality was preserved.

2.2. Design

This study follows a quantitative experimental design where we analyzed and compared the meansof the varables established in the hypothesis. The experiment followed a mixed between subjects' design with three experimental conditions defined by the type of feedback messages shown to the students after answering a multiple choice question related to a science text and the correctness of their answers. Students were assessed on science background knowledge and assigned to one of the three experimental conditions following a matching procedure. Two of the conditions included KR + EF messages and an additional condition with only KR feedback was set as a Control Group. We tested two types of elaborative feedback; one (a) $EF_{Correct}$, orient the student towards the correct response (n = 43) and other (b) $EF_{Incorrect}$ correct, oriented to help the student understand why their answer was incorrect (n = 42). The Control Group also received a feedback message, which contained KR information along followed by a message irrelevant to the task, such as "you've already answered question x" (n = 43).

2.3. Materials

Science Background Knowledge. Students were administered a previous background knowledge questionnaire on science consisting on 30 statements where some of them where correct and others incorrect. Students had to circle “True/False” for each statement they knew the answer and “I don't know” if they were unsure. Half of the items had a strong relation with the topic read on the training phase by the students and the other half tested knowledge of different scientific topics. 17 test items were true while 13 were false, and the maximum test score was up to 30 points. An example of a true question of an item related to the texts' content was: "A particle of a gas always weighs the same regardless of its temperature", while an example of a false statement with no relation to the text topic was: "1 liter of volume of any liquid is equal to 1 kg of weight". (Annex 1) The psychometric properties of the Background Knowledge Questionnaire revealed reasonable indices of reliability (*Cronbach Alpha* = .74) (Rubio, 2018), according to the indices regarding developing and reporting research instruments in science education (Taber, 2018).

Expository Science Text. For the learning phase we used an expository science texts which dealt with the topic *Atmospheric Pressure and the Wind* (1137 words long and 5 illustrations). Students were instructed to read carefully the text as they would not be able to go back to the text once they decided to access the questions.

Learning Questions. Students had to answer a total of 12 multiple choice questions; 6 TB and 6 SM questions. All questions had four different alternatives where only one was correct. The questions were initially tested with several groups of students in open-ended format and the response options were elaborated from the different answers given by the students. Therefore, the erroneous options reflected the typical mistakes made by the students in that question. Students had two attempts to answer each question. TB questions required the student to retrieve information from the text or to make small inferences among consecutive sentences; an example of a TB question would be "*What causes atmospheric pressure?*" where the answer can be extracted directly from the text. Whereas SM questions required high level inferences or the application of this information to new situations; an example of a SM question would be "*If you try to replicate Torricelli's experiment on top of a mountain, at eight thousand meters, instead of at sea level, what do you think will happen?*". The text explains Torricelli's experiment, what atmospheric pressure is and how it varies with altitude, so the student must connect all the relevant information and apply it to that new situation in order to answer correctly the question.

Transfer Questions. For the posttest students answered a total of 16 questions in paper-pencil format: 8 of the questions were repeated from the training phase, but this time as open-ended format questions instead of multiple choice questions (4 TB and 4 SM), and 8 new questions were added in multiple-choice format (4 TB and 4 SM). Correct responses were scored with 1 point, whereas those open-ended questions that were correct but incomplete responses were scored with 0.5 points. As an example of a repeated SM question in open-ended format we can find, "What would happen in Spain if there was a storm in the Mediterranean and an anticyclone in Portugal?" An example of a new SM question in multiple choice format we can find is (alternative d is correct):

> *What will happen with a swollen and tightly sealed elastic balloon if the temperature rises and the air inside is heated?*
>
> a) When heated it will increase its volume.
>
> b) When heated, its volume will decrease.
>
> c) When heated it will weigh less and ascend.
>
> d) Nothing, since the number of particles inside the globe remains constant.

The correct statement is *a* as the student has to transfer the information on the text about what happens to the particles when temperature rises to the specific situation explained in the statement, as there is no section on the text that explains the situation given on this question.

2.4. Procedure

Participants were tested in the computer lab in the *Faculty of Education* in two sessions on two different days. In a first session that lasted 30 minutes, students were tested on science previous background knowledge (PBK) and they were explained the functioning of the *Read&Learn* software. Then they were assigned to the experimental conditions controlling PBK. In a second session that lasted 1 hour, students did the experimental task and completed the posttest. The experimental task is described in Figure 1 as the Learning phase. Students read the text at their will and then answered question-by-question being this a recursive process until the end of the task. Students had two attempts to answer each question. In the first attempt, they could not reread the text to answer the question. After selecting a choice and validating it, they were given KR feedback (e.g., your response is right/wrong). One group simply got this feedback (control), while the other two got additional elaborations according to the experimental condition. The three groups had the option to go back to the text and search for additional information that could help them give a correct answer on the second attempt. After this new attempt, students got new KR feedback and could proceed to the following question. After the learning phase, students watched a 20-minute video of an unrelated topic as a delay activity and then took the posttest.

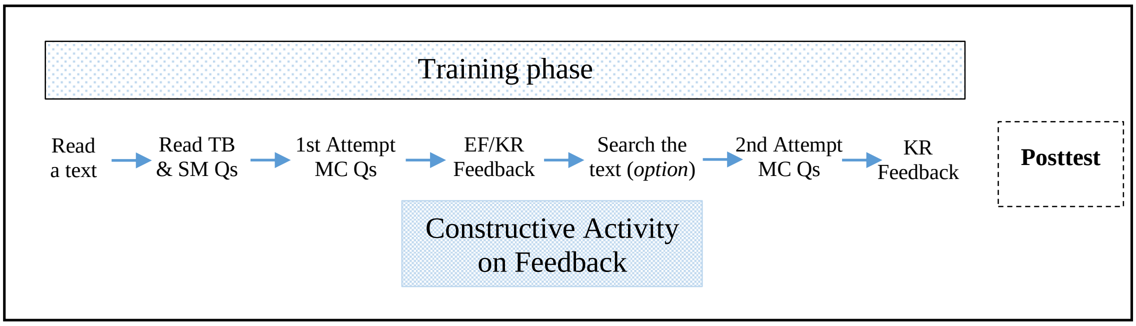


Figure 1. Procedure of the experimental phase in study 1.

2.5. Measures

Initial Time Reading the Text. We calculated the number of milliseconds the student invested in reading the text for the first time, before accessing the questions screen.

Scores on Training Phase. We calculated the number of successful answers given by the student in the first attempt, *First Attempt Performance*, in the second attempt when failing the first attempt, and considering the *Total Final Score*. Each question answered correctly was scored with 1 point and scores for TB and SM questions were broken down.

Success Rate in Second Attempt. We computed the success rate for the second attempt, after failing the first attempt and having read the feedback message. Scores for TB and SM questions were also broken down.

Mean Time Reading Feedback. We computed the mean of milliseconds the students spent reading the feedback message on the first attempt. Those times considered outliers were removed. This variable can vary highly due to small distractions, which do not reflect the time reading feedback. Thus, those values higher than 3 times the interquartile range of their group were cap to that limit, before making the average. These times were broken down into two of our analyses: in one of them based on the success or failure of the question and on other occasions depending on the type of question, TB or SM. In both cases these variables were considered within-subject's variables.

Searching Reading Time. We calculated the number of milliseconds the student is reading the text after receiving the feedback message and while answering the question on a second attempt. This variable also was broken down based on the success or failure of the question, thus considered as a within-subject's variables.

Score on Transfer Test. We calculated the number of successful answers in the transfer test. Scores for TB and SM were also broken down. Each multiple choice question answered correctly was scored with 1 point. Participants' answers to the open-ended questions were scored independently by two raters using a solution index developed by the researchers. Correct responses were scored with 1 point, whereas correct but incomplete responses were scored with 0.5 points. Inter-rater agreement measured with Cohen's kappa was .82. A correction template was created for following studies. Disagreements were discussed until a consensus was reached.

2.6. Read&Learn Software

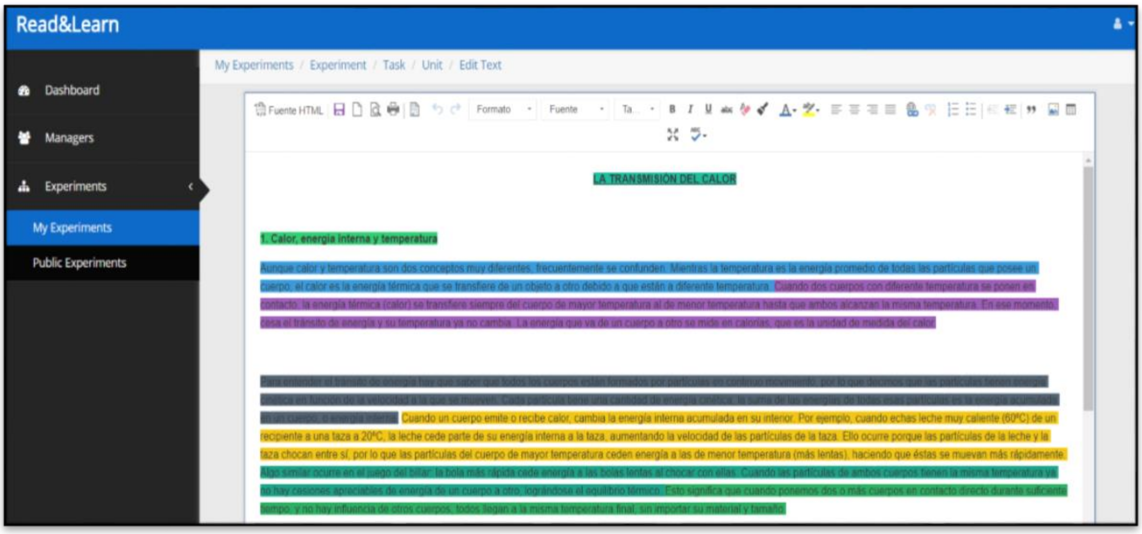
Read&Learn is an online application used to record the student's behavior while performing tasks on digital environments for learning purposes. A common task for learning declarative content in educational setting is by reading a text and answering questions about it. Although this can be done in paper-pencil format, creating the digital environment permits us record the students' behavior. In this situation where a students read texts, perform learning tasks (e.g., answer questions), and may receive feedback, they not only perform actions (e.g., read, select a statement in a multiple choice question), but they also make decisions (e.g., rereading this but not that information, reading this information slow but this other very fast, or changing their response). The technology inherent in *Read&Learn* records these online actions and decisions, creates a file data with a detailed sequence, and transforms the data into indices representative of cognitive operations which can be analyzed to generate a standardized file that aggregates the information from all students to perform statistical analysis.

Read&Learn technology is an extension of *Read&Answer* (Vidal-Abarca et al., 2011), a computer desktop tool designed to evaluate how readers interact with the text and tasks to analyze the processes involved in task-oriented reading. As it is a tool designed for research (although it can also be used in classroom settings), it allows manipulating different aspects of the learning context such as the text availability while answering the questions, type and conditions of feedback (e.g., number of response attempts), or type and conditions of questions (e.g., JOLs prior answering a question).

I will now briefly present some of the most characteristics utilities of *Read&Learn*, how the researcher designs the experiment and how the students see it. One of the most distinguished characteristics of this tool is the masking procedure which enables the researcher to know what sections of the text have been read during more time, the sequence of reading or which paragraphs have been read during search, among other situations. When we decide to mask the text, the student has to unmask the section he wants to read by clicking on it with the mouse, the rest of the text appears blurred (see Figure 2). During the edition of the experiment, the researcher decides how much information will be unmasked with the click by selecting each masking area and assigning a different color to each, as shown in Figure 3. Or course, the masking procedure is not only available to mask the text in the initial reading. The researcher can also decide how do students see the text during search and revision or if the questions and statements appear masked or unmasked (see Figure 4).

Para entender el tránsito de energía hay que saber que todos los cuerpos están formados por partículas en continuo movimiento, por lo que decimos que las partículas tienen energía cinética en función de la velocidad a la que se mueven. Cada partícula tiene una cantidad de energía cinética; la suma de las energías de todas esas partículas es la energía acumulada en un cuerpo, o energía interna.

Figure 2. Student’s view of the text being masked and out section unmasked by clicking.



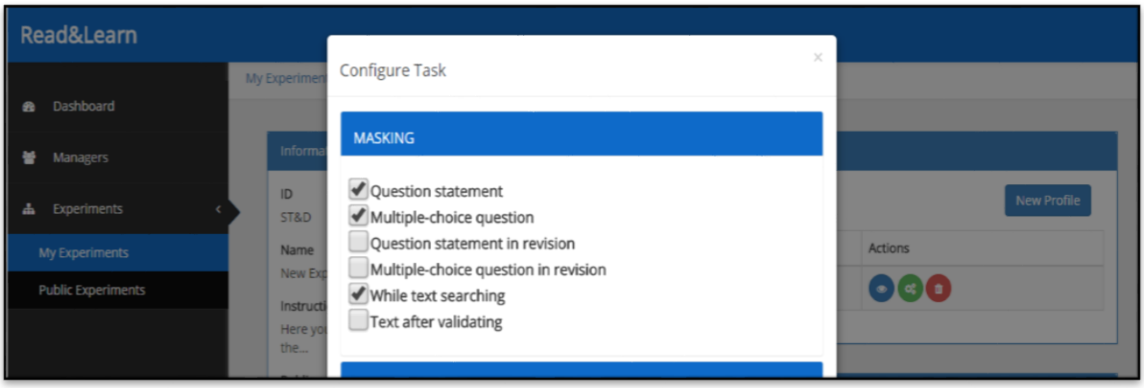
LA TRANSMISIÓN DEL CALOR

1. Calor, energía interna y temperatura

Aunque calor y temperatura son dos conceptos muy diferentes, frecuentemente se confunden. Mientras la temperatura es la energía promedio de todas las partículas que posee un cuerpo, el calor es la energía térmica que se transfiere de un objeto a otro debido a que están a diferente temperatura. Cuando dos cuerpos con diferente temperatura se ponen en contacto, la energía térmica (calor) se transfiere siempre del cuerpo de mayor temperatura al de menor temperatura hasta que ambos alcanzan la misma temperatura. En ese momento, cesa el tránsito de energía y su temperatura ya no cambia. La energía que va de un cuerpo a otro se mide en calorías, que es la unidad de medida del calor.

Para entender el tránsito de energía hay que saber que todos los cuerpos están formados por partículas en continuo movimiento, por lo que decimos que las partículas tienen energía cinética en función de la velocidad a la que se mueven. Cada partícula tiene una cantidad de energía cinética; la suma de las energías de todas esas partículas es la energía acumulada en un cuerpo, o energía interna. Cuando un cuerpo emite o recibe calor, cambia la energía interna acumulada en su interior. Por ejemplo, cuando echas leche muy caliente (60ºC) de un recipiente a una taza a 20ºC, la leche cede parte de su energía interna a la taza, aumentando la velocidad de las partículas de la taza. Ello ocurre porque las partículas de la leche y la taza chocan entre sí, por lo que las partículas del cuerpo de mayor temperatura ceden energía a las de menor temperatura (más lentas), haciendo que éstas se muevan más rápidamente. Algo similar ocurre en el juego del billar: la bola más rápida cede energía a las bolas lentas al chocar con ellas. Cuando las partículas de ambos cuerpos tienen la misma temperatura ya no hay cesiones apreciables de energía de un cuerpo a otro, lográndose el equilibrio térmico. Esto significa que cuando ponemos dos o más cuerpos en contacto directo durante suficiente tiempo, y no hay influencia de otros cuerpos, todos llegan a la misma temperatura final, sin importar su material y tamaño.

Figure 3. Editing screen showing different masking areas with different colors.



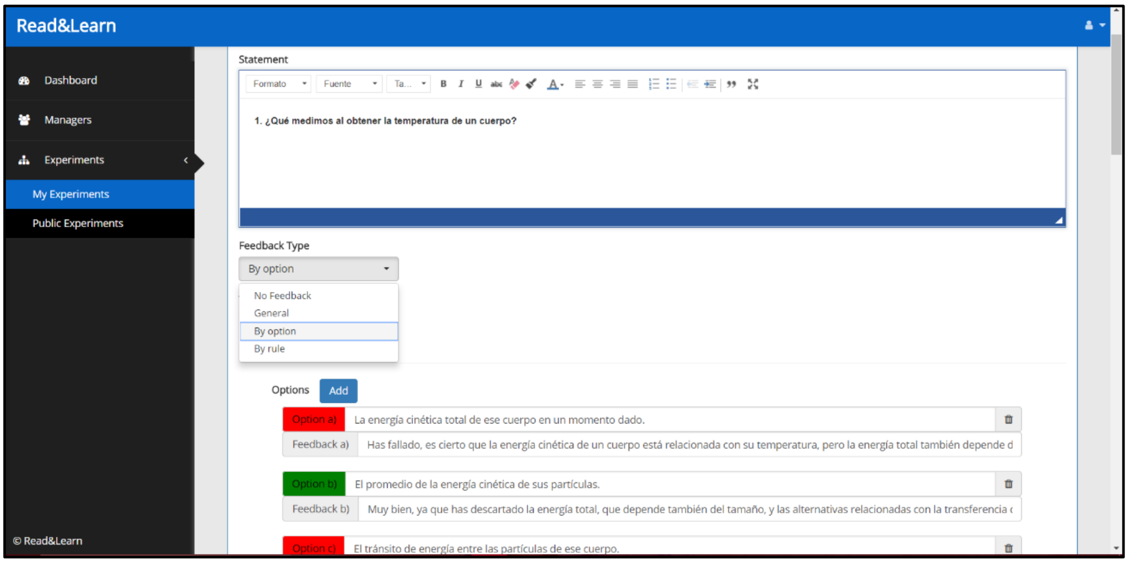
Configure Task

MASKING

☑ Question statement
☑ Multiple-choice question
☐ Question statement in revision
☐ Multiple-choice question in revision
☑ While text searching
☐ Text after validating

Figure 4. Configuration screen of masking options.

Another characteristic used vastly in research and a key element of the software for this dissertation is the possibility to provide automatic feedback to the students after answering a question. When a researcher wants to give feedback to a student, his display appears as in Figure 5. Once the researcher has selected the type of question (in this case, multiple choice), the statement has to be written on the box at the top, and the different options in the lower part. The correct option should be marked in green by clicking on the letter of the option while the rest will be marked in red. To give feedback, first you have to select what type of feedback you want to show the student from the dropdown. In this dissertation we used general feedback for the $EF_{Correct}$ conditions, and by option for $EF_{Incorrect}$ and $EF_{Combined}$ conditions. As you can see on the screenshot in Figure 5, when you choose feedback by option, the feedback message has to be written down right under each of the options the student can select to answer the question. When the student is carrying out the task and receives a feedback message, it will contain the information written by the researcher while editing the experiment and a visual correction of the answer, i.e. a green tick when the option chosen was marked as correct by the researcher (see Figure 6), and a red cross when the option was incorrect. Nevertheless, the presentation of the feedback messages can be adapted to the experimental objectives as the researcher can decide many details such as the timing of the message (after every question, after every 4 question…), the presentation mode (it can be displayed as a written message, a verbal message or even a pictorial message) or establish a new rule to give the feedback message (e.g., after 2 wrong answers).



Statement

1. ¿Qué medimos al obtener la temperatura de un cuerpo?

Option a) La energía cinética total de ese cuerpo en un momento dado.

Feedback a) Has fallado, es cierto que la energía cinética de un cuerpo está relacionada con su temperatura, pero la energía total también depende d

Option b) El promedio de la energía cinética de sus partículas.

Feedback b) Muy bien, ya que has descartado la energía total, que depende también del tamaño, y las alternativas relacionadas con la transferencia

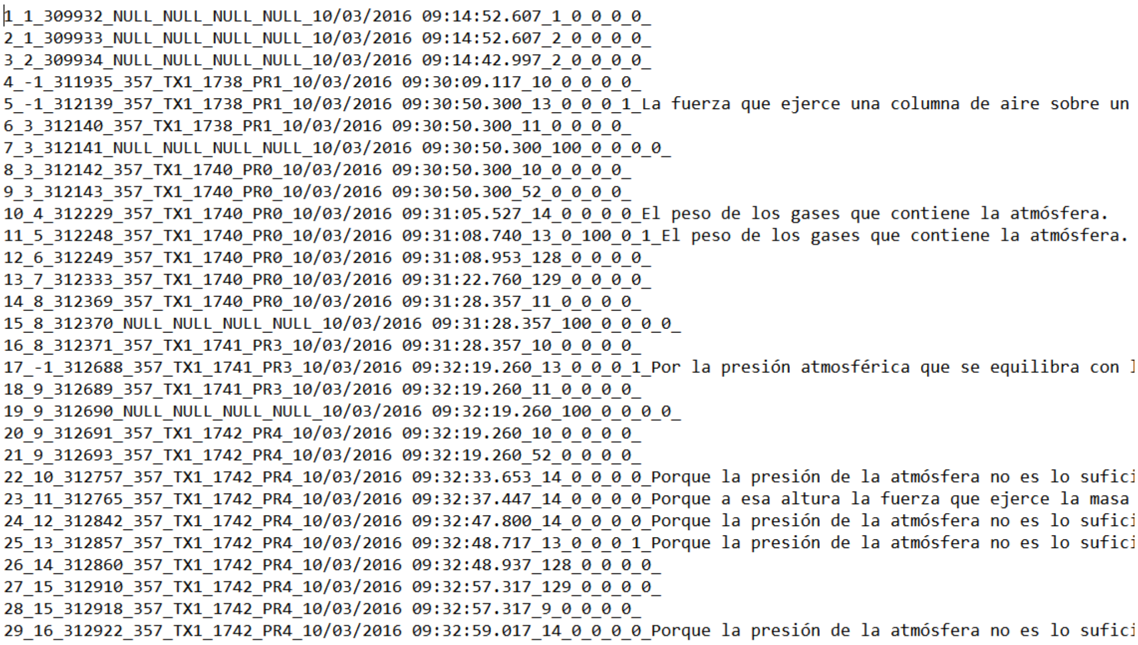
Option c) El tránsito de energía entre las partículas de ese cuerpo.

Figure 5. Researcher’s display to select type of feedback and write the feedback message.

Muy bien, has descartado los objetos fabricados con materiales que transfieren bien el calor y aquellos que al tener muy poca masa no podrían aislar nuestra lata durante mucho tiempo.

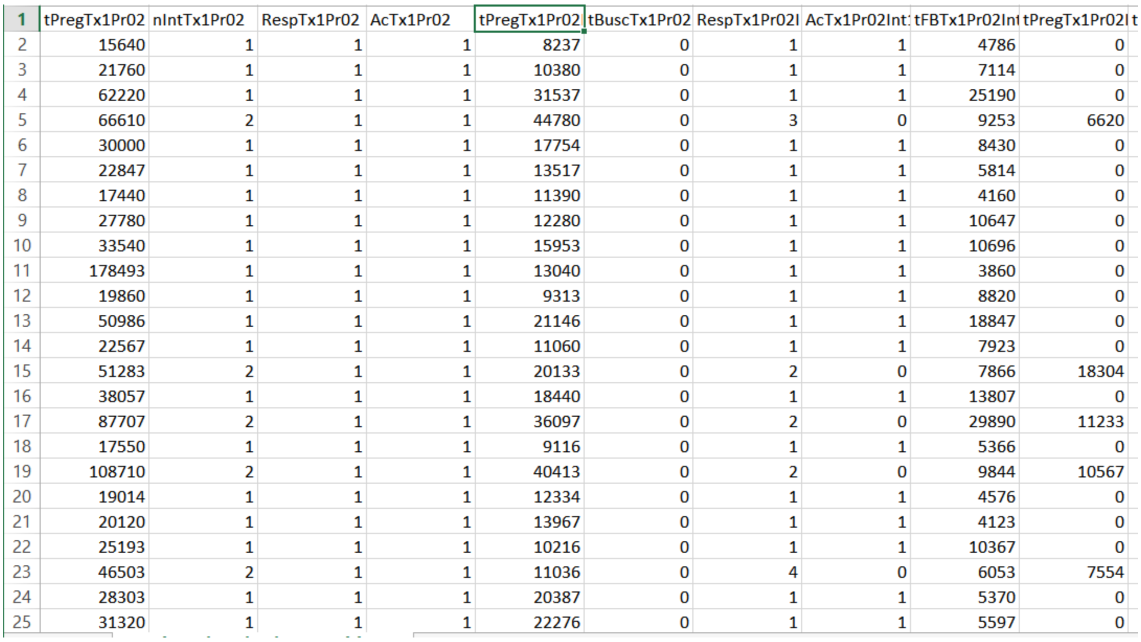
Figure 6. Students screen when receiving a feedback message after a correct answer.

Finally, and vastly relevant for research, it records the readers’ interaction with the system, their behaviors and actions by storing the action and exact millisecond of every click of the mouse, to be able to calculate the processing time of each action by calculating the interval between mouse clicks. This way, it generates log file for each student (see Figure 7). From these log files, a number of indices may be automatically computed. For example: time reading the text initially, number of visits to a specific question, time searching and time reading the feedback message among other indices. It does not register only times and actions, but it also registers context variables such as the number of text and the number of question. We have created different programs which transform the log sequence into a standardized file (see Figure 8). This way, we can analyze these sequential files, compute additional variables and transforms them into a unique file data which can be exported to a statistical package, such as SPSS, for further analysis. This data allows us to analyze the different aspects that manipulated can influence learning while students work on the task set in Read&Learn software.



```
1_1_309932_NULL_NULL_NULL_NULL_10/03/2016 09:14:52.607_1_0_0_0_0_
2_1_309933_NULL_NULL_NULL_NULL_10/03/2016 09:14:52.607_2_0_0_0_0_
3_2_309934_NULL_NULL_NULL_NULL_10/03/2016 09:14:42.997_2_0_0_0_0_
4_-1_311935_357_TX1_1738_PR1_10/03/2016 09:30:09.117_10_0_0_0_0_
5_-1_312139_357_TX1_1738_PR1_10/03/2016 09:30:50.300_13_0_0_0_1_La fuerza que ejerce una columna de aire sobre un
6_3_312140_357_TX1_1738_PR1_10/03/2016 09:30:50.300_11_0_0_0_0_
7_3_312141_NULL_NULL_NULL_NULL_10/03/2016 09:30:50.300_100_0_0_0_0_
8_3_312142_357_TX1_1740_PR0_10/03/2016 09:30:50.300_10_0_0_0_0_
9_3_312143_357_TX1_1740_PR0_10/03/2016 09:30:50.300_52_0_0_0_0_
10_4_312229_357_TX1_1740_PR0_10/03/2016 09:31:05.527_14_0_0_0_0_El peso de los gases que contiene la atmósfera.
11_5_312248_357_TX1_1740_PR0_10/03/2016 09:31:08.740_13_0_100_0_1_El peso de los gases que contiene la atmósfera.
12_6_312249_357_TX1_1740_PR0_10/03/2016 09:31:08.953_128_0_0_0_0_
13_7_312333_357_TX1_1740_PR0_10/03/2016 09:31:22.760_129_0_0_0_0_
14_8_312369_357_TX1_1740_PR0_10/03/2016 09:31:28.357_11_0_0_0_0_
15_8_312370_NULL_NULL_NULL_NULL_10/03/2016 09:31:28.357_100_0_0_0_0_
16_8_312371_357_TX1_1741_PR3_10/03/2016 09:31:28.357_10_0_0_0_0_
17_-1_312688_357_TX1_1741_PR3_10/03/2016 09:32:19.260_13_0_0_0_1_Por la presión atmosférica que se equilibra con
18_9_312689_357_TX1_1741_PR3_10/03/2016 09:32:19.260_11_0_0_0_0_
19_9_312690_NULL_NULL_NULL_NULL_10/03/2016 09:32:19.260_100_0_0_0_0_
20_9_312691_357_TX1_1742_PR4_10/03/2016 09:32:19.260_10_0_0_0_0_
21_9_312693_357_TX1_1742_PR4_10/03/2016 09:32:19.260_52_0_0_0_0_
22_10_312757_357_TX1_1742_PR4_10/03/2016 09:32:33.653_14_0_0_0_0_Porque la presión de la atmósfera no es lo sufici
23_11_312765_357_TX1_1742_PR4_10/03/2016 09:32:37.447_14_0_0_0_0_Porque a esa altura la fuerza que ejerce la masa
24_12_312842_357_TX1_1742_PR4_10/03/2016 09:32:47.800_14_0_0_0_0_Porque la presión de la atmósfera no es lo sufici
25_13_312857_357_TX1_1742_PR4_10/03/2016 09:32:48.717_13_0_0_0_1_Porque la presión de la atmósfera no es lo sufici
26_14_312860_357_TX1_1742_PR4_10/03/2016 09:32:48.937_128_0_0_0_0_
27_15_312910_357_TX1_1742_PR4_10/03/2016 09:32:57.317_129_0_0_0_0_
28_15_312918_357_TX1_1742_PR4_10/03/2016 09:32:57.317_9_0_0_0_0_
29_16_312922_357_TX1_1742_PR4_10/03/2016 09:32:59.017_14_0_0_0_0_Porque la presión de la atmósfera no es lo sufici
```

Figure 7. Log file with the sequence of one students' data.



1	tPregTx1Pr02	nIntTx1Pr02	RespTx1Pr02	AcTx1Pr02	tPregTx1Pr02	tBuscTx1Pr02	RespTx1Pr02I	AcTx1Pr02Int	tFBTx1Pr02Int	tPregTx1Pr02I	t
2	15640	1	1	1	8237	0	1	1	4786	0	
3	21760	1	1	1	10380	0	1	1	7114	0	
4	62220	1	1	1	31537	0	1	1	25190	0	
5	66610	2	1	1	44780	0	3	0	9253	6620	
6	30000	1	1	1	17754	0	1	1	8430	0	
7	22847	1	1	1	13517	0	1	1	5814	0	
8	17440	1	1	1	11390	0	1	1	4160	0	
9	27780	1	1	1	12280	0	1	1	10647	0	
10	33540	1	1	1	15953	0	1	1	10696	0	
11	178493	1	1	1	13040	0	1	1	3860	0	
12	19860	1	1	1	9313	0	1	1	8820	0	
13	50986	1	1	1	21146	0	1	1	18847	0	
14	22567	1	1	1	11060	0	1	1	7923	0	
15	51283	2	1	1	20133	0	2	0	7866	18304	
16	38057	1	1	1	18440	0	1	1	13807	0	
17	87707	2	1	1	36097	0	2	0	29890	11233	
18	17550	1	1	1	9116	0	1	1	5366	0	
19	108710	2	1	1	40413	0	2	0	9844	10567	
20	19014	1	1	1	12334	0	1	1	4576	0	
21	20120	1	1	1	13967	0	1	1	4123	0	
22	25193	1	1	1	10216	0	1	1	10367	0	
23	46503	2	1	1	11036	0	4	0	6053	7554	
24	28303	1	1	1	20387	0	1	1	5370	0	
25	31320	1	1	1	22276	0	1	1	5597	0	

Figure 8. Standardized file computed from log file.

3. RESULTS

Firstly, preliminary analyses were computed to assure the comparability of the three experimental conditions. To do this, we computed 3 one-way ANOVAs including as independent variable the three experimental conditions ($EF_{Correct}$, $EF_{Incorrect}$, and *CG*). No differences were found in the student’s *Science Background Knowledge* [F(2,124) = .058, *p* = .944] or *Initial Reading Time* [F(2,124) = 1.114, *p* = .332]. Therefore, the differences found in the groups cannot be attributed to previous knowledge or to the time dedicated to studying the ideas on the text before receiving the feedback message. This is also reflected in the scores of *First Attempt Performance*, where we also found no differences between the groups [F(2,124) = 1.389, *p* = .253] before the students were exposed to the different feedback messages.

For testing the three firsts predictions regarding the time reading feedback depending on success or failure, an 3x2 ANOVA mixed design was computed, including their simple main effects. The *Mean Time Reading Feedback* was introduced as repeated-measures with two values, according to the correction on the first attempt (right-wrong) and all three conditions were kept as a between-subjects factor.

Our first prediction has its aim in determining whether the time spent on feedback was different depending on the success of the task (usefulness effect). In this case, we found significant differences in the *Mean Time Reading Feedback* [F(1,122) = 40.902, $p < .001$, $\eta_p^2 = .251$], as after a correct answer the reading was faster (M = 4,943, SD = 3.775) than after a question failed (M = 7.771, SD = 7.613), confirming our prediction; students pay more attention to feedback after a mistake, when they perceive the information relevant to help them give a new answer.

Our second prediction was related to the ease with which the feedback would be processed according to the coherence between the feedback information and the last selected response. After answering the question successfully, the *Means Time Reading Feedback* were not different between $EF_{Correct}$ (M = 5.848, SD = 4.558) and $EF_{Incorrect}$ (M = 6.415, SD = 3.766). Obviously, these two groups differed from the *GC* (M=2.675, SD=0.769), as these last students did not receive any elaborated explanation in their feedback. Results show that the *coherence* effect was not observed after answering correctly. However, after an error, we can observe significant differences in the time dedicated to processing the feedback, according to the coherence of the message received (see graph 2). When the student gave a wrong answer, the time spent reading feedback was significantly different between all the groups (always $p < .001$). The $EF_{Correct}$ group, which received explanatory feedback not related to his last answer, was the one that needed significantly more time to process the feedback (M = 14.021, SD = 9.145), followed by $EF_{Incorrect}$ (M = 7.915, SD = 3.981), which received information consistent with their response, and *CG* (M = 1.674, SD = .429), which did not provide any kind of elaboration.

However, the interaction between these two factors (usefulness and coherence) was significant [F(2,122) = 36.720, $p < .001$, $\eta_p^2 = .376$.] As we expected, the *CG* does not devote much time to reading the feedback message when it is right (M = 2.675, SD = 0.769) nor when it's wrong (M = 1.674, SD = .429) Although it even seems to read faster after an error, the difference is not significant ($p = .196$). In group $EF_{Incorrect}$ we found a trade-off between these two effects, with no significant difference in the time taken between when they were right (M = 6.415, SD = 3.766) and when they were wrong (M = 7.915, SD = 3.981). However, in the $EF_{Correct}$ group, both effects go in the same direction, increasing the differences (being significant $p < .001$) between a right answer (M = 5.848, SD = 4.558) and a wrong answer (M = 14.021, SD = 9.145) (see Figure 9).

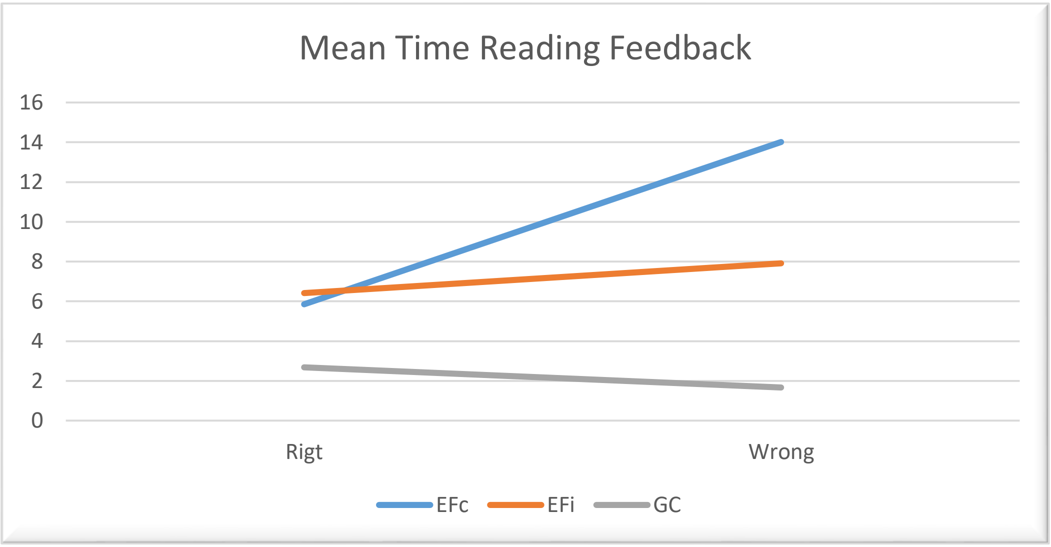


Figure 9: Mean time reading feedback according to success in the answer.

Our second objective was related to analyze the re-reading of the text. A new 3x2 mixed ANOVA was performed: The *Searching Reading Time* was introduced as repeated-measures with two values, according to the correction on the first attempt (*right-wrong*), and the three experimental groups were introduced as a factor. According to our prediction, we found that students spend more *searching* [F(1,125) = 63.836, $p < .001$, $\eta_p^2 = .338$] when they fail (M = 177.348, SD = 231.017) than when they respond correctly (M = 24,206, SD = 52.062). The effect between groups was also significant [F(2,125) = 8.219, $p < .001$, $\eta_p^2 = .116$], although only the differences between the GC (M = 131.650 SD = 124.036) and the other two groups were found to be significant since $EF_{Correct}$ (M = 65.284, SD = 92.413) and $EF_{Incorrect}$ (M = 52.311, SD=67.216) were engaged in re-reading behavior for similar amount of time. The interaction between the time when the answer was right or wrong and the experimental conditions also was significant [F(2,125) = 5.727, $p = .004$, $\eta_p^2 = .084$]. As we expected, it was the control group that needed the most time to re-reading, after an error (see Figure 10), but not after a right answer.

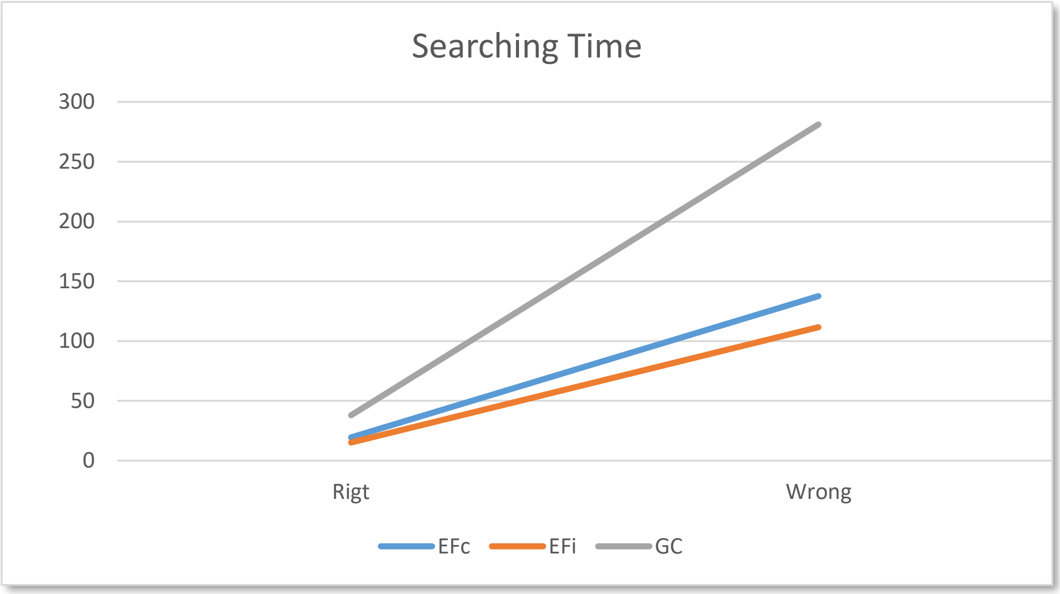


Figure 10: *Searching Time* on the text between groups according to success in the answer.

Next, we conducted a one-way ANOVA to analyze the *Success Rate in Second Attempt*, after receiving feedback or re-reading the text, for each of the three experimental groups. This result was also significant [F(2,122) = 3.213, *p* = .044, η^2 = .050]. In the Tukey *posthoc* tests, only the difference between *CG* (M = .538, SD = .279) and $\mathrm{EF}_{\mathrm{Incorrect}}$ (M = .680, SD = .278) was significant (p=.040). $EF_{Correct}$ (M = .638, SD = .236) did not differ significantly from the other two experimental groups. As we expected in our prediction, *GC* was the most benefited by being able to consult the text, but this aid was not enough to achieve the performance of group $EF_{Incorrect}$, which received information about the mistake made. However, we did not expect that the availability of the text would reduce the differences between $EF_{Incorrect}$ and $EF_{Correct}$ since this aid was available to both groups and they should re-read the text in a similar proportion.

In regards to our third specific hypothesis, we analyzed the *Level of the Question* (TB and SM) by computing a 3x2 mixed ANOVA were the type of question was our within subjects variable and the experimental condition was our between subjects variable. We predicted that students will spend more time reading the feedback message after SM question than after TB question. Regarding the variable *Mean Time Reading Feedback*, depending on the *Type of Question* the feedback was read faster on TB questions (M = 5.970, SD = 6.318) than on SM questions (M = 6.562, SD = 4.971), although the difference did not reach significant value F(1,90) = 3.317, *p* = .072. Nor did the interaction of *Type of Question* and experimental conditions show any significant results [F(2,90) = 1.646, *p* = .199] about the time spent reading the feedback.

We finally analyzed the differences in performance in the post-test by computing an one-way ANOVA to compare total scores on post-test among the three different experimental conditions. Results show no significant differences in Score on *Transfer Test* between the three groups (F<1) as the mean performance in $EF_{Correct}$ (M = 4.151, SD = 1.972), $EF_{Incorrect}$ (M = 4.162, SD = 1.956) and CG (M = 3.659, SD = 1.950) differed only slightly.

4. DISCUSSION AND CONCLUSIONS

Learning and understanding scientific content from a text may require instructional support to assure the correct learning process of new content. This support provided by the instructor, usually comes as a formative feedback message, defined as any informative message provided by a peer, instructor or computer-based system directed at the student's performance (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mathan & Koediger, 2005; Narciss, 2014) to modify cognitive or behavioral processes and improve learning (Shute, 2008). Thus, the present study examines the effects of different feedback messages on undergraduate students' performance on a task-oriented reading activity. Additionally, this investigation studies how students process these EF messages depending on (a) the *usefulness*, (b) the *coherence*, (c) the text *availability*, and (d) the *type of question* they have to respond. Three types of feedback messages were used, all of which received information about the success or failure of the question they had just answered (KR) and two of the groups also received an additional explanation. The $EF_{Correct}$ group always received information related to the correct idea, while the $EF_{Incorrect}$ group received information aimed at refuting any possible conceptual error. Finally, there was no elaborative explanation in the *Control Group* as it only included information on the correctness of their answer.

Our first prediction was related to the *usefulness* effect, since it has been observed that the time students will spend reading the FE is related to the degree to which the information is relevant and useful to perform the task (Stobart, 2008). This fact was confirmed in our studies, since students who received some kind of explanation ($EF_{Incorrect}$ and $EF_{Correct}$) spent significantly more time reading the feedback after an error when the information could be used to give a new answer.

Our second prediction was that the feedback will be easier to process when the information received was consistent with the last response given by the student (effect of *coherence*). In this study the coherence varied between experimental conditions, depending on whether the answer given was correct or incorrect. Thus, students enrolled in $EF_{Correct}$, received consistent information after a right answer, but inconsistent after an error. The opposite situation happened in the group $EF_{Incorrect}$, the information they received after a wrong answer was coherent, but not after a right answer when the message reinforces them with an explanation discarding the statements of the other options from the multiple-choice question.

It is noteworthy to point out that the *coherence* effect cannot be interpreted without simultaneously considering the *usefulness* of the feedback. In our experimental design, these two effects did not always act in the same direction and could be compensated or amplified depending on the EF condition. Therefore, our third prediction was related to the interaction between these variables. The results showed this significant interaction with large effect sizes. Thus, we can observe how the $EF_{Incorrect}$ group does not spend much more time reading the messages after having failed, as we might expect for its usefulness. In this group, usefulness and coherence had different directions, compensating each other. However, for students in $EF_{Correct}$ condition after an error, these two effects are enhanced, increasing processing times not only because the message is more relevant, but also because it is more complex for them to process as it is not consistent with their last given response. This seems to confirm the need to give a message consistent with the response given by the student, especially after an error, since when the student makes a mistake it cannot be directly replaced by a correct idea. In the learning of science, it is necessary that the student understands the error and can repair it before he can build a more adjusted knowledge (Carey, 2000). Studies with refutation texts (Diakidoy et al., 2003; Panayiota Kendeou & O'Brien, 2014; Panayiota Kendeou et al., 2014; van den Broek & Kendeou, 2008) show us how error reparation takes place in successive phases. First, the student must activate the erroneous information and its refutation by means of an explanation to understand the error. It is essential that the erroneous information, the refutation and an explanation coexist in time to understand the error, just then, the correct idea can be activated. Our data seems to fit this model, because after an error they process faster the messages directed to their refutation (M = 7.915, SD = 3.981). When we tried to activate the correct information immediately after an error, results lead to longer processing times of feedback (M = 14.021, SD = 9.145). So, when feedback is immediate it must not only be relevant but must be *coherent* with the mental process that the student is engaging in at that moment. However, in an academic context, it is common to ignore the mistake made by the student and try to help them to understand the right idea, which could interfere with their learning process.

It is important to note that we should not assume that shorter processing reading times of feedback is enough to confirm the inherent hypothesis that suggests that *coherent* information is easier to process. It could be possible that the feedback was read faster merely because it was perceived as less relevant and student were able to discard that information sooner. Hence, if this assumption was accurate, then the students discarding that information should diminish perform obtaining similar results to those students in the CG, who did not receive elaborative information. Nevertheless, the results obtained allow us to discard this option, as we found that students who received feedback oriented to correct the error, processed the message more quickly and performed better on the second attempt. The results in this study show that (a) elaborative feedback, under certain conditions, increases the probability of improving students answer on a second attempt more than just KR, and that (b) it is more useful an explanation about the mistake made than to guide the student towards the correct answer. We think that the most plausible explanation of this last fact is related to the process of refutation, as set out in the KReC model (Kendeou & O’Brien, 2014), given that before assimilating a correct idea the student must repair the error by combining the information of the error with its refutation and an explanation. This model could be used to understand how the student processes the EF after answering questions.

Regarding our second objective, we believed that students will spend more time re-reading the text when the feedback message did not give enough information, or didn't accomplish the aim of heling the student. We hypothesized that those students with only KR will refer to the text more time than those students in the EF conditions as the lack of a feedback message which includes an elaboration will increase the need to search for additional information on the text. Also, we hypothesized that the text will be visited more by those students who read feedback messages that did not clarify completely their ideas, so those who received a feedback message inconsistent with their mental representation after an error will be more prone to re-reading behavior. According to our results, it is important to outstand that when students did not have any additional information beyond the KR, higher rates of time rereading the text were observed. This shows that the EF is used by students when it contains relevant information, and this decreases their need to reread the text. On the other hand, this data show that the text availability is also as a useful complementary aid for the second attempt as searching the text can reduce the lack of additional information for those students with only KR feedback. Thus, the beneficial effect of only KR feedback in this experiment has been counterbalanced by the use of searches to the text. The influence of the *Availability* of the text affected each of the experimental groups differently. First, the *Re-reading Time*, in the text, varied between groups and showed significant differences between CG (131 seconds) and the $EF_{Correct}$ (65 sec.) and $EF_{Incorrect}$ (53 sec,), but there were no significant differences between these two groups. Re-reading the text was a behavior almost exclusive after making an error (M=177.348, SD=231.017), with very little access to the text after a right answer (M=24,206, SD=52.062), and more especially in the group that had not received any explanation (*CG*). Therefore, the interaction between experimental groups and the correct or incorrect answer to the question was also significant as greater use of rereadings depended simultaneously on whether the answer was wrong and whether he had received any EF or not. However, in a similar study by Mañez (2019), in which the students made the decision to look at an EF message or not, it was found that access to the EF was more frequent when they had not been able to consult the text beforehand. Therefore, it seems that these are two similar types of aids and that when the student receives one of them, the possibility of using the other is reduced.

Furthermore, in this study it has been especially relevant to have found significant differences in the *Success Rate During the Second Attempt*, as those students who received a refutation of their error after failure, outperformed those who only received KR feedback messages and spent less time reading the feedback message than those in $EF_{Correct}$ condition. This contributes to the understanding of the importance of a refutational tool to boost knowledge acquisition during the learning phase.

Finally, to examine the *Usefulness* of different types of feedback depending on the type of question, we hypothesized that students will spend more time reading the feedback in the SM questions as they involve a higher number of connections and inferences from the text ideas and the students' prior knowledge. Contrary to our expectations, there were no significant differences in time spent reading the feedback message after a TB or a SM question. We also predicted that this difference would be notable across conditions, but our data shows no significant differences depending on the experimental condition. We also reject our last hypothesis which stated that there will be difference in Total Score in the Posttest across condition, though our results show that these differences are not significant.

In summary, in this study the *usefulness* effect is confirmed, students invest more time processing the feedback message when this is more useful when they had a second attempt to answer. Furthermore, this has been extended to the use of rereading the text. Students spend more time searching after a wrong answer, especially if they have not received explanatory feedback. Furthermore, this second study also confirmed the interaction between *usefulness* and *coherence*. In the $EF_{Correct}$ group, we observed a strong increase in the time spent on feedback after an error, when is most useful and when the message that they have received was less consistent with their last response. While in the $EF_{Incorrect}$ group, after an error, they did not significantly increase the time dedicated to read the feedback. Thus, although the feedback was more useful after an error, it was also more coherent, compensating each other for these effects. On the other hand, the possibility of consulting they might compensate the lack of EF in the *CG* and might be useful for students in EF conditions to verify their new response. Differences were only found between the highest performing group ($EF_{Incorrect}$) and the lowest-performing group (*CG*). Additionally, we found no differences in time processing feedback depending on the type of question, and post-test results show no differences in performance after a short delay.

To sum up, main results of this study indicate that (a) when students lack elaborative information about their performance, they carry out more searches on the text (b) in a second attempt, those students which received refutational information scores better on a second attempt than those who only got information about the correctness of their answers, (c) students with only corrective information spent less time reading the feedback message than students on the other two conditions as it was expected due to the difference in the length of the message with respect to EF. Considering EF conditions,

(d) after giving an erroneous answer students spend more time reading the feedback message than after giving a correct answer, although the time is reduced if the information helps in the refutational process, (e) when students succeed on a question and are giving feedback oriented to the correct answer, they read it in less time as if fits their mental representation, and (f)the type of question does not affect significantly feedback reading times, but it does affect searching behavior as those students with only corrective information make significantly more searches for SM questions than for TB questions.

Hence, it is important to note that students need to refute their previous ideas and that these cannot simply be replaced by new information (Caravita & Halldén, 1994; Dunbar, Fugelsang, & Stein, 2007; Kendeou & O'Brien, 2014). So, error-oriented feedback is a more efficient and adjusted aid and learners improved their performance, before drawing up a new answer. This is especially important in electronic environments, where KR and EF are given at the same time and immediately after the answer, when the student still has the wrong information active in their memory. In this context, the explanation about their error facilitates its reparation and the elaboration of a new response. On the other hand, the availability of text can compensate for the lack of feedback, although it is less effective and efficient than error-driven feedback.

Although this study has been useful to explore some issues of interest that had not been addressed previously, it is not exempt from limitations such as the presence of only one text and the sample size. Additionally, the possibility to search of the text before the second attempt might have diminished the effect of feedback in performance as the lack of feedback or the lack of coherence between the feedback message and the students' mental model to answer the question could be counterbalanced by more searches to the text. This way, it is harder to isolate the effect of instructional feedback to study the usefulness and benefits of this tool. Consequently, we think that this experiment should be complemented by a future experiment that could replicate this experiment but without giving the option to search the text, thus the only aid the students will have would be the feedback messages on the three experimental conditions. This way, feedback would be the only instructional aid to help students'.

CHAPTER 4

STUDY 2. THE IMPACT OF NO-SEARCHING ON THE EFFECTIVENESS OF INSTRUCTIONAL CORRECTIVE AND ELABORATED FEEDBACK MESSAGES

1. OBJECTIVES AND HYPOTHESES

The present study aimed at replicating the experimental procedure in Study 1 with a major difference in the absence of text availability. On the previous study, we found that the text availability to search for information before the second attempt compensated the lack of information between conditions. The control group spent significantly more time than the other two EF conditions re-reading the text after making an error on the first attempt. Also, students in $EF_{Correct}$ condition spent more time searching than $EF_{Incorrect}$ condition but mean time differences did not reach significance level. For this reason, we have decided to make another study to test the same variables as in Study 1 but without letting students search on the text at any point. Therefore, the only aid students will have will be the feedback messages. Our main objective is to examine the differences between conditions when these differences can only be due to the different feedback messages the students receive depending on the condition they are enrolled in. Thus, we expect that the differences between conditions are greater as there is no influence of the text revision.

As we aim to replicate the findings in study 1 and investigate the extent to which feedback on its own is a powerful instructional tool, our objectives are the same as in the previous study. Our first objective is to replicate and confirm the effects of *usefulness* and *coherence* of feedback and as well as their interaction.

Concerning the type of question, although we found no significant differences in Study 1, we expect that the EF messages will be more useful for SM questions as students can not visit the text for additional information after failure and that, therefore, the time spent reading the messages in these types if question is longer. We expect an interaction with the experimental condition as *CG* will have no additional aids.

Regarding results on *Success Rate On Second Attempt,* we expect to find significant differences across conditions as the only aid before answering a second attempt will be the feedback message. Due to the interaction between the Usefulness and the *Coherence* effect, we predict that those students in $EF_{Incorrect}$ will have a higher success rate on the second attempt than those in $EF_{Correct}$. Due to the lack of aids, we predict that *CG* will show the smallest improvement.

Finally, we expect that these differences will also be visible after a short delay. Thus, we predict that there will be differences in *Score on Transfer Test*across conditions, were those students in $EF_{Incorrect}$ will outperform the other two experimental conditions.

2. METHOD

2.1. Participants

One hundred and twelve undergraduates enrolled in teaching studies and psychology (17% male and 83% female) participated in the study. All students were

Spanish native speakers. The study was set as a complementary class activity. For those students enrolled in Teaching Studies, it was set as a class activity for the subject Science Teaching, and for those enrolled in Psychology, it was set as a class activity for the subject Educational and Instructional Psychology. Participants received a bonus mark on their final grade of the subject. Students' confidentiality was preserved.

2.2. Design

The experiment design was identical to the one in Study 1, but these time students were not allowed to go back to search on the text after reading the feedback message. The study followed a between subjects' design with three experimental conditions defined by the type of feedback messages shown to the students after answering a multiple-choice question related to a science text, although to test some of our hypothesis, we carried out a mixed design, introducing some of our variables as within subject's variables. Students were assessed on science background knowledge and assigned to one of the three experimental conditions following a matching procedure. Two of the conditions included KR + EF messages and an additional condition with only KR feedback was set as a Control Group. We tested two types of elaborative feedback; one (a) $EF_{Correct}$, orient the student towards the correct response (n = 36) and other (b) $EF_{Incorrect}$, oriented to help the student understand why their answer was incorrect (n = 35). The *Control Group* also received a feedback message, which contained KR information along followed by a message irrelevant to the task, such as "you've already answered question x" (n = 41).

2.3. Materials

As in study 1, students were tested on *Science Background Knowledge*, they read a text on atmospheric pressure and answered 12 learning questions and 16 transfer questions. Students were tested using the *Real&Learn* software.

2.4. Procedure

In session one, students were tested on science previous background knowledge. Then they were assigned to the experimental conditions controlling their prior knowledge. In session two, students did the experimental task and posttest. The experimental task is described in Figure 1 as the Learning phase. Students read the text at their will and then answered question-by-question. They had two attempts to answer each question. After selecting a choice and validating it, they were given KR feedback (e.g., your response is right/wrong). One group simply got this feedback (control), while the other two got additional EFs ($EF_{Correct}$ and $EF_{Incorrect}$). As in study 1, $EF_{Correct}$ included information not present in the text aimed at helping students to make the inferences to give the right response, whereas that in $EF_{Incorrect}$ explained why the wrong choice was not correct, without giving clues toward the right choice. Students could read the message at their will and then shut it down. This was the only information students had before they had the 2nd attempt to answer the question, as they software was designed to not allow the student to go back to the text and search for additional information. After these new attempt students got new KR feedback and could proceed to the following question. After the learning phase, students watched a 20-minute video of an unrelated topic as a delay activity and then took the posttest.

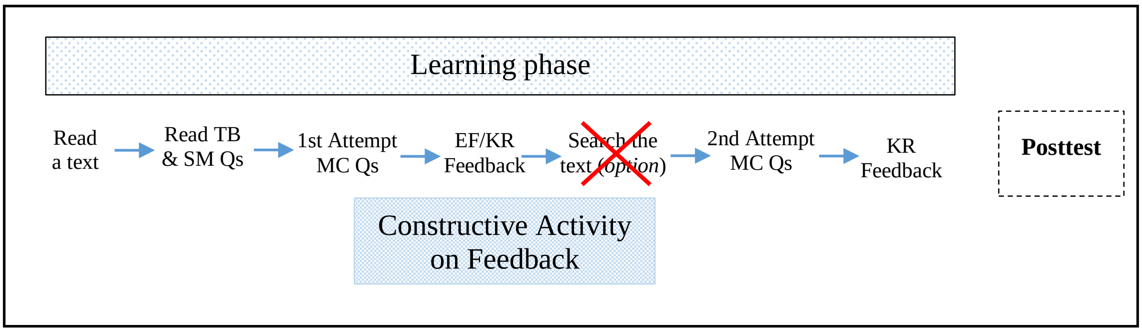


Figure 1. Procedure of the experimental phase in study 2.

2.5. Measures

We computed the following online measures: *Initial Time Reading the Text*, *First Attempt Performance*, *Success Rate in Second Attempt*, *Mean Time Reading Feedback* and *Score on Transfer Test*. We used *Read&Learn Software* as instructional tool to collect the data. See the section “Materials and measures” of study 1 for a detailed description of the materials, the measures, and the software.

3. RESULTS

To replicate the results found in study 1, the same set of analysis were computed, with exception of those including *Searching Reading Times*, as this variable was not measured due to the characteristics of this new experimental design.

Initially, to verify the comparability of the three experimental groups we conducted a series of one-way ANOVAS, between subjects, taking as an independent variable these three conditions ($EF_{Correct}$, $EF_{Incorrect}$, and *CG*). For these groups, we found no differences in *Science Background Knowledge* [F(2,117) = .035, p = .966] or *Initial Reading Time* [F(2,117) = 1.005, p = .369]. Therefore, the differences found in the groups cannot be attributed to previous knowledge or to the time dedicated to studying the material. This is also reflected in the level of *First Attempt Performance*, before reading the feedback, where there were also no differences between the groups [F(2,116) = .883, p = .416].

Regarding our first prediction directed to confirm the effects of *usefulness* and *coherence,* as well as their interaction, we found the following results.

The replication of the *usefulness* effect as tested by carrying out a 3x2 mixed ANOVA were the success on the question was our within subjects variable and the experimental condition was our between subjects variable The *Mean Time Reading Feedback* also depended on whether the answer was *right* (M = 4.418, SD =3.418) or *wrong* (M = 6.282, SD = 6.609) was significant [F(1,122)=40.902, p<.001, η_p^2=.251] when the text is not available, as in study 1. This seems to confirm that the KR previously provided to the students determines in which grade they analyze the subsequent explanations of their answer, at least when there is a new possibility to respond. Students pay more attention to the feedback message when they know they have given an incorrect answer.

The replication of the *coherence* effect. We found partial replication of the coherence effect as this time we find differences between experimental conditions not only after failure, but also after giving a correct answer. After giving a correct question, the *Means Time Reading Feedback* were: $EF_{Correct}$ (M = 4.660, SD = 3.626), $EF_{Incorrect}$ (M = 6.153, SD = 3.980), and *CG* (M = 2.577. SD = 0.796). Comparisons of these means, based on Tukey, show that there are only differences between the *CG* and the other two experimental groups, but we found no significant differences between the groups that had received EF. After giving a wrong answer, the feedback reading time differences across conditions show a slight increment; $EF_{Correct}$ (M = 10.243, SD = 8.248), $EF_{Incorrect}$ (M = 7.165, SD = 5.046) and, *CG* (M = 1.606, SD = 0.398). In this case, the differences between

EF conditions and the *CG* were once again significant, but contrary to the findings on the previous study, differences between EF experimental groups are close to but do not reach significance level (p = .053). In both cases, although the data goes in the direction of our predictions, we cannot conclude that the effect of *coherence* can be demonstrated in this study.

The replication of the interaction between these two factors (*usefulness* and *coherence*), in this second study, was again significant [F(2,113) = 14.733, p < .001; η_p^2 = .207]. It is seen how the $EF_{Correct}$ group needs to increase feedback reading times after a question fails, an increase that does not occur in $EF_{Inorrect}$, nor in *GC* (see Figure 2). Since in the $EF_{Correct}$ group the effects of *coherence* and *usefulness* go in the same direction, accentuating the differences (being significant p < =.001) in *Mean Time Reading Feedback* between *right* (M = 4.660, SD = 3.626) and *wrong* (M = 10.243, SD = 8.248). While in the $EF_{Incorrect}$ group, these effects have opposite directions, canceling each other out: *right* (M = 6.153, SD = 3.979) and *wrong* (M = 7.165, SD = 5.046). On the other hand, in the *CG*, where we did not expect any of these effects, no significant differences were found between *right* (M = 2.577, SD = 0.796) and *wrong* (M = 1.605, SD = 0.398).

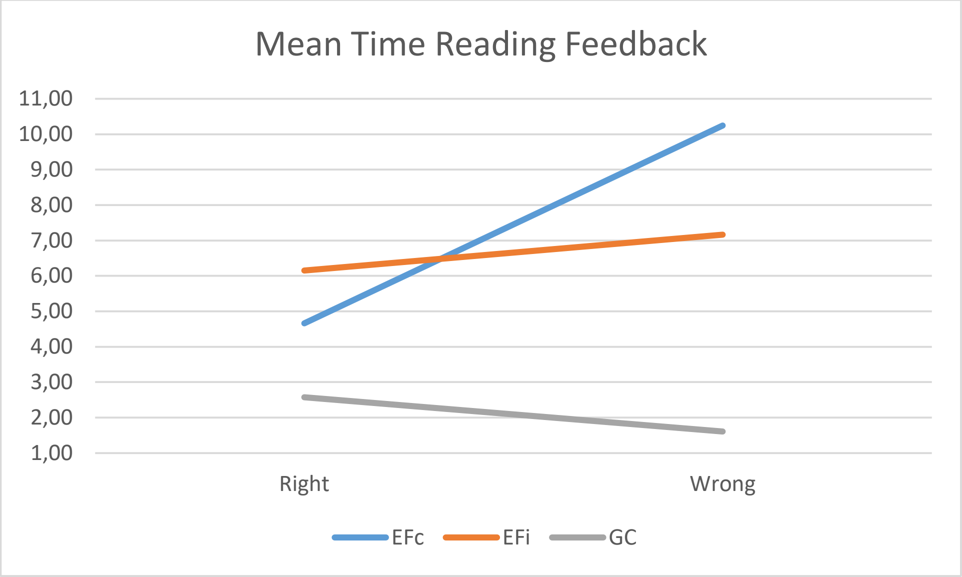


Figure 2: Mean Time Reading Feedback according respond Right or Wrong between conditions.

To test our second specific hypothesis regarding the influence of the type of question and the type of feedback message, an new 3x2 mixed ANOVA was carried out considering the type of question our within subjects variable and the experimental condition was our between subjects variable. Considering the variable *Mean Time Reading Feedback*, we did not find significant differences depending on the *Type of Question* [F(1,86) = 1.198, p = .277)], as students spent a mean time of 4.784 (SD = 3.388) seconds reading the feedback message after TB questions and a mean time of 5.184 (SD = 4.513) seconds reading the feedback message after a SM question, nor their interaction with the experimental conditions [F(2,86) = 3.027, *p* = .054,] although it was shown to be close to significance. The *CG* and $EF_{Incorrect}$ showed a very similar behavior towards the TB and SM questions, while the $EF_{Correct}$ group increased the time dedicated to read feedback on the SM (see Figure 3).

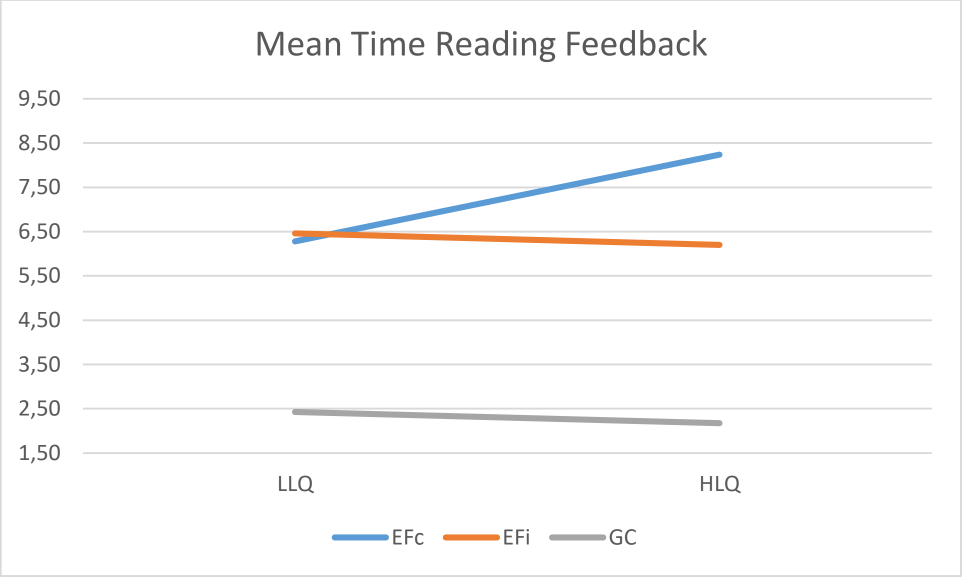


Figure 3: Mean Time Reading Feedback according to type of question (LLQ vs. HLQ) between conditions.

Lastly, we were interested in analyzing the success rates after reading the different feedback messages. The *Success Rate in Second Attempt* was introduced as a dependent variable in a new *one-way* ANOVA, with all three experimental conditions as a factor. We found a significant difference between groups [F(2,116) = 18.212, $p < .001$, $\eta_p^2 = .244$], according to our prediction. In post hoc comparison there were significant differences ($p = .021$) between $EF_{Incorrect}$ success rate (M = .655, SD = .264) and $EF_{Correct}$ group (M = .507, SD = .241). And, in the same way, these two groups showed significant differences ($p < .001$, in both cases) with the *GC* (M = .327, SD = .210). Finally, no differences on *Score on Transfer Test* were found between conditions (F<1), as those students enrolled in $EF_{Correct}$ obtained a total performance score of 3.055 (SD = 0.524), those students in $EF_{Incorrect}$ scored 3.109 (SD = 0.541) and those in the CG scored 2.978 (SD = 0.460)

4. DISCUSSION AND CONCLUSION

The present study aims at investigating the effect of instructional feedback when this is the only aid students have when facing a question answering task. The objective of this design is to isolate the effect of feedback to reduce other causes (such as searching) which may have an impact on the differences between conditions due to a variable different than feedback, the one we are interested in studying. Restricting the text availability, the effectiveness of the different types of feedback was shown in a clearer way, as the deficits in the information received could not be compensated with an additional aid which was a new re-reading of the text. In addition, by replicating study one, we study how students process these EF messages depending on the correctness of their answers and type of question they have to respond over a control condition with only KR feedback.

Regarding the main results in this study, students who got refutational information about their answer obtained higher scores than students who only received KR feedback. Moreover, after eliminating the possibility of consulting the text, a reduction in success rates is observed, although this reduction did not show a homogeneous pattern either, since in the *CG* the success rate was reduced in the second attempt from .55 to .33, on their part in $EF_{Correct}$ the decrease was from .64 to .51, while in the $EF_{Incorrect}$ group practically the same results were obtained; .68 and .65. These results seem to show that the more useful the information received in the feedback message is, the more influence it has on a second response. This is especially relevant as in this second study the differences on performance between the groups increased depending on the feedback received, showing that giving information about why the chosen statement is mistaken, produces an execution significantly superior to the other two conditions, as well as a greater execution on the $EF_{Correct}$ over the *CG*. This seems to demonstrate the hypothesis that when students fail, and before giving a new answer, they need to refute their previous ideas. These ideas cannot be simply replaced by new information (Caravita & Halldén, 1994; Dunbar, Fugelsang, & Stein, 2007), but there is a set of underlying processes involved which we would like to study in following investigations. Nevertheless, it is important to note that differences are seen on the second attempt, but no differences depending on the experimental condition has been registered on the transfer test.

As we have seen, the usefulness of the feedback message is an important variable that can affect the reading of feedback messages, but it is not the only one. *Feedback Reading Times* will also be affected by the easiness with which the message can be processed. According to our initial hypothesis, the feedback message will be easier to process when the information received is consistent with the response given by the student and helps in the refutation process, if necessary. In our study this *coherence* varied between experimental conditions depending on whether the answer given was correct or incorrect. Those students who received information oriented to the correct response had a coherent message with their mental model when they gave a correct answer, but a dissonant message when they gave an incorrect answer as the information given was distant from the mental model created by the students to answer the question. The contrary happened to students who got refutational information, as an inconsistent message for them was shown when they scored correctly because an explanation about why the other statements were wrong was given, but when they scored incorrectly, a message explained why the chosen statement was correct and was close to the mental model created by the student.

The *coherence* effect was only partially verified in our results. When comparing the times dedicated to processing the feedback, in both studies, between $EF_{Correct}$ (M=14.021, SD=9.145) and $EF_{Incorrect}$ (M=7.915, SD=3.981). But, when the text was not available, after an error, these differences were reduced. Given that the questions and the messages were identical, in both studies, we consider that this large difference, between studies, must not be related to coherence but with the proper availability of the text, which we will discuss at a later point. The differences with the *CG* are not relevant as this group received a neutral message.

The results of our work seem to demonstrate the fact that the *coherence* with the last given answer is an important factor as well as the *usefulness* of the message. Both the coherence and the usefulness determine the ease with which the feedback message is processed, producing an interaction between these two variables. Thus, observing the results, we see that when we explain why a given answer is incorrect, the learners spends less time reading the messages after having failed. In these cases, the information is more useful (for the second attempt) and it is more coherent (with the students' mental model), which facilitates its processing. The contrary happens after scoring correctly, as the message is less relevant because students have already scored correctly on this question, and less coherent as it explains why the other statements are incorrect. In the results, these differences between relevance and coherence are compensated for each other, causing students to spend the same amount of time reading the messages after a success or an error. However, for students with information on the correct idea these two effects are enhanced, decreasing and increasing processing times, depending on whether the first answer was correct or incorrect. Thus, after a success, the information received is not only less relevant, but also more coherent, since it is aimed at reinforcing their given response, which translates into a reduction in the time dedicated to processing feedback. However, in this group, after an error, the time students need to process the feedback increases significantly, not only because the message is more relevant, but also because it is more complex for them to process, as it is not consistent with their last given response.

As seen in the previous study, these results also seem to support the students' need to refute the error before giving a new response. The studies with refutation texts (Diakidoy & Kendeou, 2001; Diakidoy et al., 2003; Kendeou & O'Brien, 2014; Kendeou et al., 2014; van den Broek & Kendeou, 2008), show us the successive phases to repair errors in the mental model. The student must first activate the erroneous information and a refutation which involves an explanation of why the information is wrong followed by the activation of the correct idea. If we extrapolate this process to the context of our study, the students after an erroneous answer would have to direct their activity to refute and understand why their answer is incorrect, and only after finishing this process, they will start to elaborate a new answer. Since the EF is received together with the KR, after a mistake, the most relevant process at that time should be aimed at correcting the error, as was the case in the $EF_{Incorrect}$ group. However, if after a wrong answer we provide information related to the correct response, interference with the refutation process may occur over the capacity of the working memory, which could be shown in longer processing times of feedback, as observed in our studies. On the other hand, we must not forget that this longer processing time does not translate into better performance. The idea of interference is reinforced by the fact that the results show that the second attempt is more likely to be successful in the $EF_{Incorrect}$ condition, when the refutation process has been favored, despite having shorter reading times.

Finally, as far as *Question Type* is concerned, results show that although students spend more time reading feedback in SM questions, this difference does not reach significance level, as well as their interaction with experimental conditions.

Nevertheless, it is important to note other limitations inherent to this experiment that we should try to solve on our next study. First, this experimental procedure were students cannot search the text to obtain more information or to clarify their ideas is useful to accomplish the aim of this study to learn about feedback, but it is very different to what would happen in a classroom in a natural setting, as question-answering tasks to learn often allow students to search for further information. Moreover, if we want to understand the underlying processes on elaborated feedback processing, we should focus only on EF. We have already seen notable differences between KR and EF that assure that this last type of feedback is more helpful for students. Additionally, we have also commented that both $EF_{Incorrect}$ and $EF_{Correct}$ have a coherence limitation either when students score correctly or when they score incorrectly, respectively. We should give a type of feedback which is always relevant and coherent independently of the student’s success on the task. Additionally, having only 1 text on the experiment reduces the power of the results and including another text will increase validity to our results. Finally, we have not found any differences on transfer tests so far, so we should revise the transfer test to assure that the questions are reliable and valid.

CHAPTER 5

STUDY 3. ELABORATED FEEDBACK MESSAGE TO ENHANCE THE STUDENT’S PERFORMANCE: IS IT BETTER TO DIRECT THE INFORMATION TO THE RIGHT ANSWER, TO CORRECT A MISTAKE OR TO USE A COMBINATION OF BOTH?

1. OBJECTIVES AND HYPOTHESES

Providing students with effective instructional feedback is one of the goals of the educational community. To do so, it is important to know and understand the inherent processes associated to knowledge construction, and especially to conceptual error reparation when dealing with misconceptions in complex concepts (Gadgil, Nokes-Malach, & Chi, 2012; Kendeou, Walsh, Smith, & O'Brien, 2014; Posner, Strike, Hewson, & Gertzog, 1982).

The learning process of feedback in intelligent learning environments model (LP-FILE model) proposed by Timms, DeVelle, Lay (2016) is a decision-making model which considers the steps in which students perceive, decode, receive and give meaning to information in the feedback message. The general assumption of the model is that the student has to make an error when answering a question in order to consider revision. If the student does not detect the error, external feedback is provided, and the student must perceive it and decode the message. Then, the student has to understand the meaning that has to fit with the task model created by the student. If the formative feedback works correctly, it will help the student to correct his mistake and create an adequate mental model to respond correctly to the question demands.

In studies 1 and 2, students had to read a science text and then had two attempts to answer text-based (TB) and situation model (SM) multiple-choice questions. After the first attempt all students received Knowledge of Results feedback (KR), whereas the two experimental groups got additional EF, either $EF_{Correct}$ (which oriented the students towards the correct answer regardless of their score) or $EF_{Incorrect}$ (which gave information explaining why the chosen alternative was incorrect). Results in studies 1 and 2 showed that both types of EF are more efficient than KR feedback, although the possibility to search on the text helps diminishing these differences as the lack of information, is counterbalanced with more time searching the text to get additional information, but KR feedback on its own does not improve students' performance as much as an elaboration. $EF_{Incorrect}$ is significantly more efficient than $EF_{Correct}$ as students spend less time reading the feedback message and score better on a second attempt. These results are explained by suggesting that feedback is closer to the mental representation built by the student and therefore the explanation is coherent and easier to understand. A limitation we observed in our first two studies was that $EF_{Incorrect}$ was provided independently of failure or success on the first attempt. Thus, when the answer was correct, $EF_{Incorrect}$ informed students about the correctness of their choice but also explained why the wrong alternatives were incorrect.

Consequently, it may be rational to give $EF_{Incorrect}$ after failure in the first attempt, but $EF_{Correct}$ after success, as it is contra-intuitive to activate erroneous ideas. This way, the feedback message will always be coherent with the student's mental representation as it, either to explain why the answer was incorrect, or to enrich it when it was correct, making communication more relevant for the student and adjusted to their performance.

With this study, we aim at overcoming the limitations found in the previous studies by adding an additional text to increase the power of the study and prove that the results found until now are not text-dependent. We also revised and modified the questions in the learning phase and in the transfer test. To do this we, we did a pilot study where we presented the text in paper format and the questions in open-ended format. We selected those questions where we found a greater incidence of misconceptions and we included as options in the multiple choice question the correct statement plus the three wrong statements most written by the students. Additionally, we associated each of the questions in the training phase to a question on the post-test addressed to test the same textual idea. Finally, research indicates that EF messages are more effective than corrective ones (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Van der Kleij et al; 2011, Vidal-Abarca, Martínez, Ferrer & García, 2017; Whyte, Karolick, Neilsen, Elder & Hawley, 1995) therefore we have also eliminated the KR feedback condition, as it was shown to be less efficient than EF and we have replaced it with a new EF condition ($EF_{Combined}$). This new category gives response to a limitation found previously as none of the two EF types on the previous studies were coherent with the students' mental model after success and after failure. This way, $EF_{Combined}$ will orient the student towards the correct response after giving a right answer reinforcing his mental model by using a paraphrase of the idea on the correct answer and will help the student repair his misconception after giving a wrong answer by explaining why the chosen statement is not correct.

The objective of the present study was to deepen the research to examine what type of elaborative feedback is more efficient to help students learn in question-answering tasks: $EF_{Correct}$, $EF_{Incorrect}$ or $EF_{Combined}$ (when students give a correct answer they will receive $EF_{Correct}$ and when they give an incorrect answer they will receive $EF_{Incorrect}$), to explore which type is more effective to help students learn when reading a science text and answering questions about the text, and to test learning in a transfer test after a short delay. Considering the revised literature, we predict that $EF_{combined}$ will be the most efficient type of feedback. Although the message after failure should be as efficient as $EF_{Incorrect}$ as it is the same message, and therefore the information given to the student fits with his mental representation, so it is easier to process and to understand, the $EF_{Combined}$ eliminates the unnatural message students received after succeeding on the question. This third option benefits of the message that work better on each of the two EF conditions examined in the previous studies.

Considering earlier results of our experiments and previous literature, we expect to replicate the usefulness and coherence effect, as well as the interaction between them.

First, regarding the *usefulness* effect we expect that students continue finding more relevant the information after giving a wrong answer than after scoring correctly. Thus, they will spend more amount of time reading elaborative feedback messages after failing a question than after succeeding. When students fail a question the information is seen as more relevant and useful for giving a new correct answer.

Second, we expect differences in *Time Reading Feedback* depending on the coherence effect with the last given response by the student, inherent in each experimental condition. We expect that those students in $EF_{Correct}$ will process faster the feedback message after success and slower after failure. We expect the opposite for those students enrolled in $EF_{Incorrect}$ as we expect feedback processing times after success to be higher than in the other groups and shorter processing times after failure. Finally, we expect that $EF_{Incorrect}$ shows shorter processing times both after success and after failure as the message presented is always coherent with the students' mental representation. Thus, we predict that after a success the students in the $EF_{Inorrect}$ group will spend more time reading the feedback than the students in $EF_{Correct}$ and $EF_{Incorrect}$, and after failure, the students in the $EF_{Correct}$ group will spend more time reading the feedback than the students in $EF_{Incorrect}$ and $EF_{Incorrect}$.

Additionally we predict that there will be an interaction between the *usefulness* and the *coherence* effects of elaborated feedback messages. The *coherence* once again varied depending on the experimental condition although the usefulness depended on the correctness of student's answers. As in studies 1 and 2, $EF_{Correct}$ presents a coherent message after success but an incoherent message after failure. Thus, we expect that this group shows longer processing times of feedback messages after failure as information is more useful and more complex to process. The contrary will happen after success as information will be less useful for students but easier to process. Like in studies 1 and 2, $EF_{Incorrect}$ shows a coherent message after students give an incorrect answer but an incoherent message after success. Thus, we expect that this group shows reduced processing times of feedback messages after failure as information is more useful and easier to process as it fits the student mental model. The contrary will happen after success, information will be less useful but harder to process. Finally, for the new $EF_{Combined}$ condition we expect shorter processing times in all cases as the message will only vary in its usefulness but will always be coherent. When students give a correct answer, the message will not be useful, but it will be easy to process as it will fit the student’s mental representation, and when a student gives a wrong answer, the message will be useful and it will be easy to process.

Fourthly, as we found no significant differences regarding *Time Searching* in previous studies between experimental conditions that administered EF to students, we hypothesize that time searching the text will be reduced from the previous experiments as all conditions were given elaborated information about the students’ performance. The possibility of searching the text is an additional aid that the student will decide to use if they consider the feedback message does not include sufficient relevant information to answer the second attempt directly. Thus, we hypothesize that students will decide to do very few searchers, and this will not variate across experimental conditions.

Fifth, although we did not find differences in *Time Reading Feedback* in studies 1 and 2, we predict that due to the increase of texts and questions, we will find differences in *Time Reading Feedback* depending on the *Type of Question*. Students will spend more amount of time reading feedback messages on SM questions than on TB questions, as they involve more complex connections between textual ideas and the creation of inferences or even background knowledge information.

Our last hypothesis suggests that there will be significant differences on Success Rate on a Second attempt and Total post-test performance across conditions. In reference to success rate in a second attempt, we expect that students in $EF_{Incorrect}$ and $EF_{Combined}$ conditions have a greater improvement rate than those students in $EF_{Correct}$ as they have been given information to understand and correct their misconception. Regarding post test score, we expect that those participants in $EF_{Combined}$ condition will score higher than those in $EF_{Incorrect}$ and $EF_{Correct}$ in total post test results, as the feedback messages provided to this condition were the most adjusted to the participant’s mental representation of the textual ideas helping them to consolidate correct answers and correct misconceptions after erroneous answers.

2. METHOD

2.1. Participants

A sample of 113 first-year undergraduates enrolled in teaching studies (20% male and 80% female), participated in the study without incidences. All students were Spanish native speakers. The data of seven students were excluded from the analysis due to absences during one of the phases (n = 5) or software failures (n = 2). The study was set as a voluntary class activity for those students who participated in the study and they received extra credit. Students' confidentiality was preserved. Students were assessed on science background knowledge and assigned to one of the three experimental conditions following a matching procedure. There were 40 students in the $EF_{Correct}$ condition (EF to orient the student towards the correct response), 35 students on the $EF_{Incorrect}$ condition (EF to help the student understand why their answer was incorrect), and 38 students in the $EF_{Combined}$ condition (combination of the previous two.)

2.2. Design

The experiment followed a mixed between subjects' design with three experimental conditions defined by the type of feedback messages shown to the students after answering multiple choice questions related to two science texts, and the correctness of their answers intoduced as a within subject's variable depending on *success* or *failure*.

The experiment followed a mixed between subjects' design with three experimental conditions defined by the type of feedback messages shown to the students

after answering a multiple choice question related to a science text depending on *success* or *failure*. This study follows a quantitative experimental design where we analyzed and compared the means of the variables established in the hypothesis. Students cience background knowledge was previously tested and then assigned following a matching procedure to one of the three experimental conditions. The three experimental conditions included KR + EF messages. We tested the combination of three elaborative feedback messages; one (a) $EF_{Correct}$, orient the student towards the correct response in success and failure (n = 40), another one (b) $EF_{Incorrect}$, oriented to help the student understand why their answer was incorrect after success and after failure (n = 35), and the last one (c) $EF_{Combined}$ orient the student towards the correct response in success and oriented to help the student understand why their answer was incorrect after failure (n = 38).

2.3. Experimental conditions: Elaborative Feedback

EF messages varied depending on the experimental condition. Those students in the $EF_{Correct}$ condition received a feedback message which guided the students towards the correct idea. For example, for the previous examples we discussed before, when students gave an incorrect answer for the TB question, they received this message: “You are wrong, the atmosphere is formed by a series of layers of air one on top of another. Therefore, the number of layers we have above varies according to the height at which we are.” When students in the $EF_{Correct}$ condition gave a correct answer, the message was the same except for the KR information at the beginning which said “Well done”. Those students in the $EF_{Incorrect}$ condition received a different feedback message depending on the alternative they selected. Considering the examples SM question we saw before, students received this $EF_{Incorrect}$ message when they were right (option *d*): “Well done, you have discarded objects made with materials that transfer heat well”. When students gave an incorrect answer, the feedback message was addressed to explain the specific characteristics of each alternative which made it be an incorrect option, so a different message was presented depending on the wrong alternative selected. Continuing with the same SM example, those students which selected option *a* would receive this feedback message “You are wrong, the aluminum foil is a good thermal conductor and there would be a rapid transfer of heat from the environment to our can”, those which selected option *b* would receive this feedback message “You are wrong, metals are good conductors of heat and there would be a rapid transfer of heat from the environment to the can”, and those students which selected option *c* would receive this feedback message “You are wrong, although it is true that plastic is not a good conductor, it would not be an effective barrier against environmental heat”. Finally, those students in $EF_{Combined}$ received $EF_{Correct}$ when their answer was right and $EF_{Incorrect}$ when their answer was wrong.

2.4. Materials

Science Background Knowledge. Students were administered a new previous background knowledge questionnaire on science consisting of 30 statements where some of them were correct and others incorrect. This questionnaire was modified to include questions on the topic of both texts that will appear on the training phase as well as unrelated questions to the texts. Students had to circle “True/False” for each statement they knew the answer and “I don’t know” if they were unsure. The items were classified into 3 categories depending on the relatedness of the statement to the topic of the text they will read on the training phase. Thus, 9 items had a strong relation with the topic “Heat and Temperature”, 10 items had a strong relation with the topic “Atmospheric Pressure” and 11 items tested knowledge of different scientific topics. 15 test items were true while 15 were false, and the maximum test score was up to 30 points. An example of a false statement of an item related to the text content on “Heat and Temperature” was: “There are materials like metal that are always cooler than other materials like wool”. An example of a true statement of an item related to the text content on “Atmospheric Pressure” was: “Gasses are formed by particles that move freely”, while an example of a true statement with no relation to the text topics was: “Mercury is the closest planet to the Sun”. The psychometric properties of the Background Knowledge Questionnaire revealed moderate indices of reliability (*Cronbach Alpha* = .61), according to the indices regarding developing and reporting research instruments in science education (Taber, 2018).

Expository Science Texts. For the learning phase we used two expository science texts. One of the texts dealt with the topic *Atmospheric Pressure and the Wind* (854 words long and 4 images; a revised and shortened version of the same text as Study 1 and 2) and the other one dealt with the topic *Heat and Temperature* (908 words long). Both texts were shown to the students following a masking procedure as seen in Figure 4a were letters appeared blurred and students must click on the segment they want to read to make it clear. Only one segment could be read at a time as clicking on a new one will automatically mask the previous one. Students were instructed to read carefully the texts as they would not be able to go back to the text once they decided to access the questions,

Learning Questions. Students had to answer a total of 4 TB and 4 SM questions for each text. Questions were formulated in multiple choice format where each question contained a statement and 4 different alternatives where only one of them was correct and the other 3 were distractors. Distractor alternatives were chosen from a bank list of answers in open ended format done in a previous pilot study. Students had to select the correct option and validate their answer. An example of a TB question on the topic of *Atmospheric Pressure and the Wind* is:

Why does atmospheric pressure vary with height?
a) Because at sea level the pressure is more affected by the wind.
b) Because it changes the weight of the column of air that it has above.
c) Because the pressure increases with the height when it gets colder.
d) Because the normal atmospheric pressure is that which occurs at sea level.

The correct alternative is *b* and the answer can be explicitly found in the text in the following paragraph *"...the density of the air layers varies according to the height, the air of the lower layers being denser because they support the air weight of the layers that are above."*. An example of a SM question on the topic of *Heat and Temperature* is:

If we want to keep a soda can cold to go to the beach, what should we wrap it in to keep it cooler?
a) Aluminum foil.
b) A metal box.
c) A plastic bag.
d) A wool cloth.

The correct alternative is *d* and the student has to transfer the information on the text about the characteristics of good insulators and thermal conductors to the specific situation explained in the statement, as the answer does not appear explicitly on the text.

Transfer questions. Students answered 4 open ended TB and 4 open ended SM open-ended question in paper-pencil format for each text. Students had to write short concise answers and they were given a score of 1/0.5/0 depending on the correctness of the answer. Each question in the transfer phase was associated to a question in the learning phase due to matching text ideas necessary to answer the question correctly. Continuing with the previous example, the associated TB question was "Is the density of air the same in all points of the earth? Why?", as the same segment of the text contains the relevant information to answer the question correctly. Regarding the associated SM question in the transfer test, the question was "When you increase the temperature at one end of a thermal conductor, what happens in the rest of that conductor?" as the information regarding the characteristics thermal conductors had to be applied to this situation too.

2.5. Procedure

Students were assessed during two sessions on 2 different days. In the first session, students were tested on previous background knowledge. The session lasted 15 minutes approximately. Students were matched to one of the three experimental conditions, considering their scores on the previous background knowledge questionnaire. For the second session, students went to the computer-lab classroom to proceed with the learning phase task and the transfer test individually. This session lasted approximately 90 minutes and is outlined in Figure 1. First, once they logged on *Read&Learn* they were instructed on how to use the interface, how to read using the masking/unmasking procedure and how the feedback messages will appear. Then, they had to read two texts and answer to 8 multiple choice questions per text plus an additional question about the utility of the feedback messages shown. The order of the texts was counterbalanced, so half of the students read "*Atmospheric Pressure and the Wind"* first, while the other half read "*Heat and Temperature*". Students had two attempts to answer the question correctly. In the first attempt, they could not go back to reread the text. After they answered, they were given feedback which varied depending on the condition assigned. If their answer was right, they could continue with the next question, whereas if it was wrong they had a second attempt. During this attempt, they could decide to reread the text. Once they answered, a corrective feedback message appeared in all 3 experimental conditions, then students could continue with the next question. Each question was displayed at a time, and participants could not move from one question to another. Once they finished the training phase, a short delay activity took place before students could move on to the transfer test. This activity consisted of watching a TED talk of an unrelated topic and lasted approximately 20 minutes. Once the video was finished, students turned off the computers and answered in paper-pencil format the 8 open ended questions related to each text in counterbalanced order.

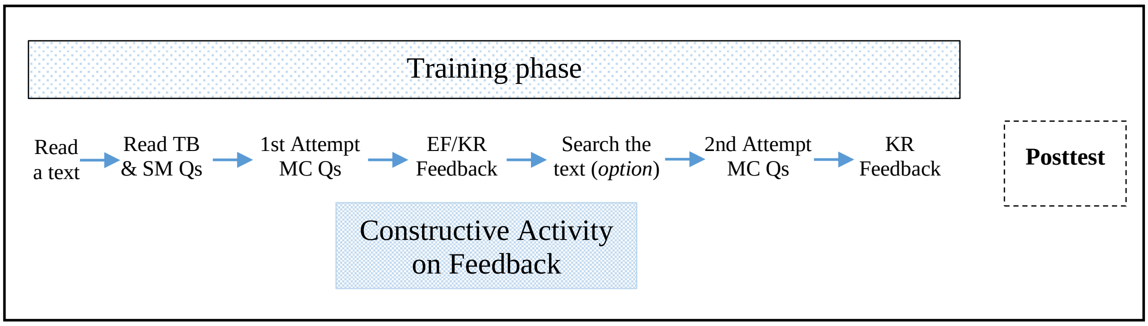


Figure 1. Procedure of the experimental phase in study 3.

2.6. Measures

Initial Time Reading the Text. We calculated the number of milliseconds the student is reading each text for the first time (i.e., before moving to questions).

Scores on Training Phase. We calculated the number of successful answers given by the student in the first attempt, *First Attempt Performance*, in the second attempt when failing the first attempt, and considering the *Total Final Score*. Each question answered correctly was scored with 1 point and scores for TB and SM questions were broken down.

Success Rate on the second attempt. We computed the success rate for the second attempt, after failing the first attempt and having read the feedback message. Scores for TB and SM were also broken down.

Mean Time Reading Feedback. We computed the mean of milliseconds the students spent reading the feedback message on the first attempt. Those times considered outliers were removed. This variable can vary highly due to small distractions, which do not reflect the time reading feedback. Thus, those values higher than 3 times the interquartile range of their group were cap to that limit, before making the average. These times were broken down into two of our analyses: in one of them based on the success or failure of the question and on other occasions depending on the type of question, TB or SM. In both cases these variables were considered within-subject’s variables.

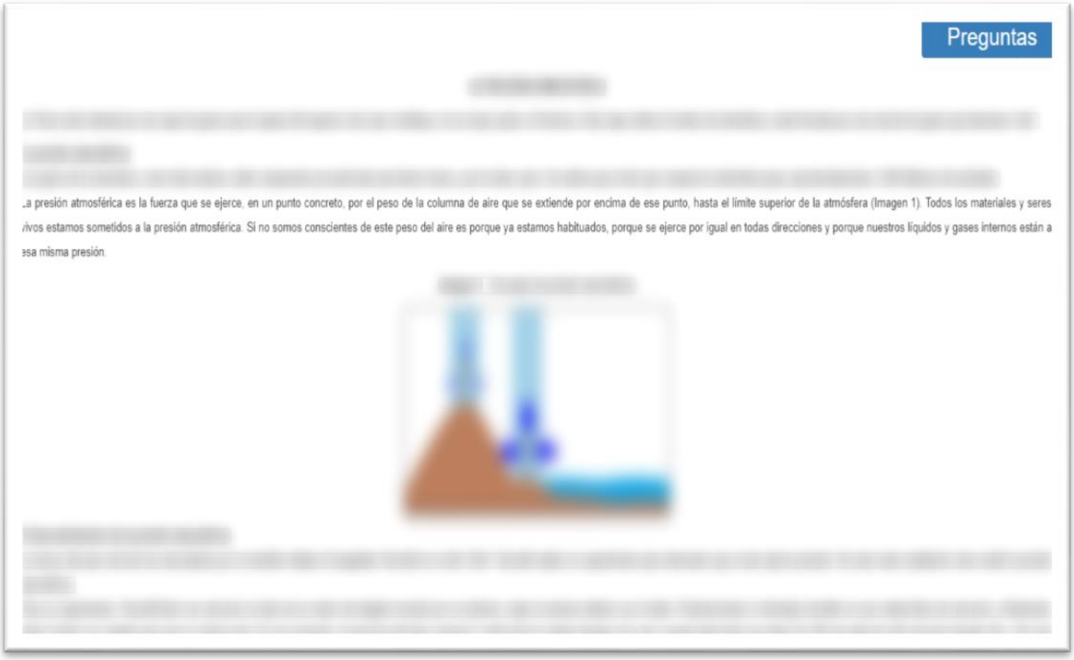
Searching Reading Time. We calculated the number of milliseconds the student is reading the text after receiving the feedback message and while answering the question on a second attempt. This variable also was broken down based on the success or failure of the question, thus considered as a within-subject’s variables.

Score on Transfer Test. We calculated the number of successful answers in the transfer test. Scores for TB and SM were also broken down. Participants’ answers to the open-ended questions were scored independently by two raters using a solution index developed by the researchers. Correct responses were scored with 1 point, whereas correct but incomplete responses were scored with 0.5 points. Inter-rater agreement measured with Cohen's kappa was .714. Disagreements were discussed until a consensus was reached.

2.7. Read&Learn Software

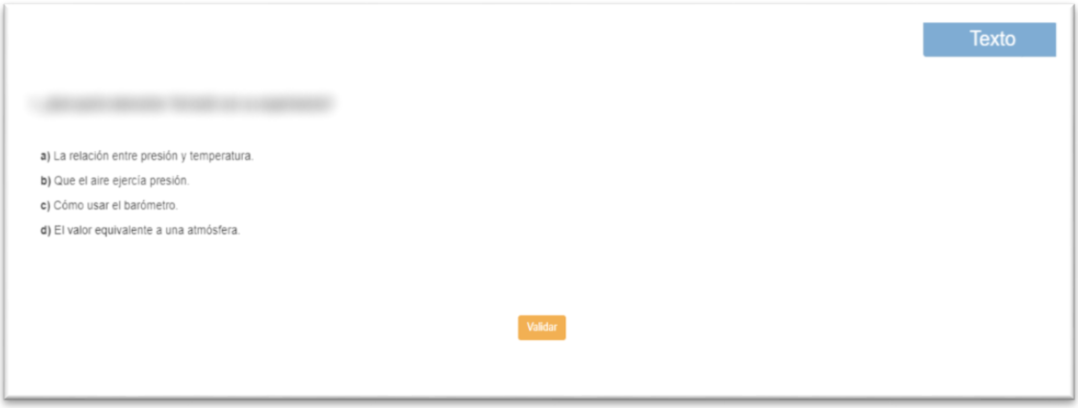
“Read and Learn” software. A new modified version of the computer-based system that records students’ interactive behaviors, *Read&Learn*, was used to complete the main task. This tool enables to trace online students’ behaviors and processing times while performing reading activities and answering questions in the training phase. This software is able to provide students with different feedback messages according to the students’ execution and performance. The *Read&Learn* template used in this investigation displays text and questions on different screens (Figure 2a and 2b) as well as providing EF messages adapted to the question and alternative (figure 2c).

a)



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 (Imagen 1). Todos los materiales y seres vivos estamos sometidos a la presión atmosférica. Si no somos conscientes de este peso del aire es porque ya estamos habituados, porque se ejerce por igual en todas direcciones y porque nuestros líquidos y gases internos están a esa misma presión.

b)



a) La relación entre presión y temperatura.
b) Que el aire ejercía presión.
c) Cómo usar el barómetro.
d) El valor equivalente a una atmósfera.

c)

Has fallado, con su experimento, Torricelli medía el peso de la atmósfera en un punto concreto.

Tienes un segundo intento para responder.

Figure 2: Screen shot of the training phase: a) the text screen with one segment unmasked while the rest appear masked, b) the question screen with masked statement and unmasked alternatives, and c) pop-up box of the EF message when students gave a wrong answer on the first attempt.

3. RESULTS

Data distribution of pretest scores, training phase scores, posttest scores and feedback reading times were inspected before conducting the main analyses. Preliminary one-way ANOVAs were carried out to control that there were no differences between experimental conditions in variables which were not manipulated between groups. We found no significant differences between conditions on *Science Background Knowledge* (*F*<1), on *Initial Time Reading the Text* (*F*<1) *or First Attempt Performance* (*F*<1).

In order to test our three first hypothesis regarding the usefulness and coherence effect as well as their interaction, an 3x2 ANOVA mixed design was computed. The *Mean Time Reading Feedback* was introduced as a repeated-measures with two values, according to the correction on the first attempt *(right-wrong*) and it was kept all three conditions as a between-subjects factor.

Our first prediction was regarding the *usefulness* effect. We stated that time reading feedback will be different depending on the success on the task. As predicted, we found main effects on time spent reading the feedback message after success and failure [$F(1, 110) = 13.625$, $p = .000$, $\eta^2_p = .11$] showing that students devote more time to reading feedback messages after knowing they gave a wrong answer ($M = 4.07$, $SD = 2.67$), than after knowing that their answer was correct ($M = 3.04$, $SD = 1.71$). As we hypothesized, students dedicate more time to reading the feedback message when they give a wrong answer, showing that their engagement in reading the EF is higher than when they give a correct answer due to the usefulness effect.

Regarding our second prediction, we hypothesized that the time spent reading the feedback message will depend on the coherence of this message with the student's mental representation. Thus, those incoherent messages will take more time to be processed by students than those messages which are in line with their thoughts. Results show that after giving a correct answer, the *Means Time Reading Feedback* were: $EF_{Correct}$ (M = 2.954, SD = 1.810), $EF_{Incorrect}$ (M = 3.067, SD = 1.895), and $EF_{Combined}$ (M = 3.094, SD = 1.430). No significant differences were found between conditions in this case as $F < 1$. After missing the question, the *Means Time Reading Feedback* differences rise slightly; $EF_{Correct}$ (M = 4.879, SD = 3.468), $EF_{Incorrect}$ (M = 3.623, SD = 1.715) and, $EF_{Combined}$ (M = 3.367, SD = 2.243). In this case, the difference between experimental conditions does not become significant. In both cases, although the results obtained show a clear direction towards our predictions, we cannot conclude that the effect of *coherence* can be demonstrated in this study.

Consequently, in regards to the interaction between *usefulness* in terms of *right-wrong* answers and these experimental groups, with different levels of *coherence*, a marginally significant interactive effect of time spent reading the feedback message after success and failure and type of feedback was also no significant, [$F(2, 110) = 2.909$, $p = .059$, $\eta^2_p = .05$] (see Figure 3). Although results do not reach significance level, we can observe that those students in $EF_{Correct}$ condition devote more time to reading the EF after failure than those students in $EF_{Incorrect}$ and $EF_{Combined}$ conditions.

Figure 3. Mean time reading feedback after success and after failure by type of feedback.

Our fourth hypothesis predicted that the time searching by students will diminish as all the experimental conditions included an EF which gave explanation related to the question, thus minimizing the time spent on re-reading the text. A mixed 3x2 ANOVA considering *Searching Reading Time*, as dependent variable and *success-error* in the question as within-subject's variable and the three experimental conditions as between-subjects show that there are no differences across conditions in all cases F < 1. Interestingly, results show that in some conditions there were no searches made during the whole experiment, as it was the case in $\text{EF}_{\text{Incorrect}}$ where the mean time searching registered was 0 seconds; in $\text{EF}_{\text{Correct}}$ the mean time searching was 0.525 (SD = 1.536) second and in $\text{EF}_{\text{Combined}}$ 0.473 (SD = 0.951) seconds.This supports our hypothesis as it shows that students perceive the feedback message as containing enough relevant information to successfully give an answer in the second attempt.

To test our fifth hypothesis that predicted that that students will spend more time reading the feedback messages after responding a SM question than a TB question, we proceeded to compute the analysis to test for differences in feedback reading times across conditions depending on the type of question. A new mixed 3x2 ANOVA with type of question (TB and SM) as within-subject variables and experimental condition as between-subject variable was conducted. Results show main effects on the type of question, $F(1, 110) = 16.292$, $p = .000$, $\eta^2_p = .13$, as students spent more time reading the feedback message after a SM question ($M = 3.742$, $SD = 1.751$) than after a TB question ($M = 3.133$, $SD = 1.572$), though there was no interaction between type of question and experimental condition ($F <1$).

In our fifth hypothesis, we were interested in analyzing the success rates after reading the different feedback messages. The *Success Rate in Second Attempt* was introduced as a dependent variable in a one-way ANOVA with all three experimental conditions as a factor. We did not find a significant difference between groups, in this case contrary to our prediction all three groups had a similar performance; $EF_{Correct}$ group (M = .57, SD = .221), $EF_{Incorrect}$ (M = .0.55, SD = .249) and $EF_{Combimed}$ (M = .0.57, SD = .244).

Finally, we examined differences in performance on post-test, *Score on Transfer Test*, depending on the type of feedback the student’s received. Results of a new one-way

ANOVA with these three groups revealed that there were significant differences in transfer test scores across experimental conditions [$F(2, 104) = 4.688$, $p = .011$, $\eta^2_p = .08$]. Post hoc Bonferroni-corrected pairwise comparisons indicated that, students in $EF_{Combimed}$ ($M = 8.300$, $SD = 2.215$) obtained better results than those in $EF_{Correct}$ ($M = 6.621$, $SD = 2.622$) ($p = .01$). No differences were found between other group comparisons as incorrect had a mean score of 7.162 (SD = 2.405).

A new mixed 3x2 ANOVA was conducted to test for differences in TB and SM question, to explore if there is a stronger impact depending on the type of question. To do so, we considered as within subject variable the type of questions and as between subjects variable the experimental condition they were enrolled in. We found significant differences between the scores in TB (M = 4.070, SD = 1.412) and SM (M = 3.449, SD = 1.489) questions [F(1, 106) = 15.515, $p < .000$, $\eta^2_p = .130$] Regarding the interaction between the question type and the experimental condition, we found no significant differences [F(2, 106) = 1.422, $p = .246$] There were no significant differences in post-test results accounting on condition for TB questions. In the contrary, considering only SM questions, results revealed that there were significant differences in transfer test scores for SM questions [F(2, 106) = 5.197, $p = .007$, $\eta^2_p = .09$]. Post hoc Bonferroni-corrected pairwise comparisons indicated that, students in $EF_{Combimed}$ ($M = 4.025$, $SD = 1.363$) obtained better results than those in $EF_{Correct}$ ($M = 3.06$, $SD = 1.738$) ($p = .01$) and $EF_{Incorrect}$ ($M = 3.147$, $SD = 1.555$) ($p = .03$). No differences were found between the last two groups.

4. DISCUSSION AND CONCLUSIONS

Herein, we presented a study that examined the efficacy of different elaborative feedback messages using Read&*Learn*, a specialized software to register students' actions and reading times. One of the most popular activities students complete in school settings to learn declarative knowledge is answering comprehension questions about a written document. When students engage in these activities, formative feedback is an important tool that may offer the students with immediate information about the correctness of their answer, hints and explanations and any kind of instructional information that can guide them towards knowledge acquisition (e.g., Fox, Klein, Entink & Timmers, 2014; Llorens et al., 2016; Timmers & Veldkamp, 2011).

The general aims of the study were first, to extend the research to examine what type of elaborative feedback message is more efficient to help students learn in question-answering tasks. In study 1 and 2, we compared KR plus $EF_{Correct}$ and KR plus $EF_{Incorrect}$ with only KR feedback finding out that both EF messages are more efficient than only KR messages. According to previous research, EF is usually delivered along with corrective feedback (Van der Kleij et al., 2015). We have already examined the extent to which corrective feedback influences EF processing (with and without an available text) considering that in this third study, the only KR condition will not give us additional information regarding our general objectives. Analyzing the type of feedback provided to students on Studies 1 and 2, we realized that none of the EF conditions was always consistent and coherent with the students' mental model of the question, as each condition gave an inconsistent message when the student gave a correct answer or when his answer was incorrect. Therefore, in this study we decided to overcome this limitation and substitute the KR condition for an $EF_{Combimed}$ condition, that gave the $EF_{Correct}$ message when the student' answer was correct, and the $EF_{Incorrect}$ message when they have an erroneous answer. This way, the *coherence* effect will always be met independently of the usefulness effect (scoring correctly or wrongly). Thus, in this new experimental condition the feedback message will always be coherent and consistent with the mental model the student built to answer the question. Our second general objective was to replicate the findings on the previous studies overcoming the limitations inherent in the study design by increasing the number of texts to obtain more powerful and generalizable findings and design a posttest linked to the text ideas evaluated on the training phase.

In reference to our first hypothesis, we predicted that according to the findings in studies 1 and 2, students would spend more time reading the feedback message after they failed the question than when they succeeded. Result once again confirm our hypothesis regarding the usefulness effect as students dedicated more time to engage in feedback reading behavior when they gave a wrong answer than when their answer was correct. This suggests that students occupy more time at discerning why their answer is incorrect and they devote more resources to understanding failure, whereas information about a correct answer is not perceived as challenging by the student (Maier et al., 2016). Additionally, this can be explained by the LP-FILE model (Timmers et al., 2016) which suggests that an error has to occur in order to engage in revision behaviors

Regarding our second prediction, this time we cannot confirm our hypothesis as the *coherence* effect has not appeared to be significant in this last study, although result show a similar pattern to those obtained until now. Thus, especially after an error occurs, we can continue seeing similar results, those students with a dissonant message to their mental model, increase the feedback reading times. As we explained on previous studies, if the student does not create an appropriate mental model to answer the question successfully, it will take more time for them to process and understand a feedback message which orients them to the correct answer by giving hints which are not coherent with their mental representation ($EF_{Correct}$), than an explanation which fits in with their mental representation and that they find coherent and easier to understand ($EF_{Incorrect}$ and $EF_{Combimed}$). Moreover, we should not discard the possibility of a decrease in time spent reading the feedback message as a consequence of increasing the number of texts and questions producing greater fatigue on the students engaged on the task.

Our third prediction hypothesized that there would be an interaction between the usefulness and the coherence effect. Contrary to previous studies, data shows that this time results have not been significant although the probability is close to significance. We did expect this interaction to replicate studies 1 and 2 but possibly differences between studies in the relatedness of experimental conditions, the number of texts number of questions without increasing the sample size has had an impact on the power of the analysis. Mean scores, though, continue showing that the usefulness of the message (more useful after failure) and the coherence of the message (if it is related to the answer given) are important factors in determining the information on the feedback message.

Concerning our fourth hypothesis, we predicted that as all the experimental condition included some kind on explanation in their elaboration, students will decide not to search as much to gather additional information from the text. As we predicted, giving students EF messages reduces searching decisions drastically. We even found that some conditions did not decide to search at any point. Although we do not know the reason for this to occur, it may be due to an increased fatigue as this study increased the total number of questions and had an additional text to the previous studies. Nevertheless, these results show that a feedback that contains an explanation related to the question is perceived as sufficiently informative on its own by the students. These findings are contrary to previous research that has shown that an available text is an essential tool for teaching when Formative Feedback is provided during question-answering tasks (Llorens, Vidal-Abarca, & Cerdán, 2016; Llorens, Cerdán, & Vidal-Abarca, 2014; Máñez, Vidal-Abarca & Martínez, 2019; Máñez, Vidal-Abarca, Martínez & Kendeou, 2017). Thus, we can establish that it depends on the type of instructional feedback given to the students and their perceived capacity to help in a second attempt.

In reference to our fifth hypothesis, contrary to the results obtained on the previous two studies, we obtain significant differences between the time spent reading TB and SM question, as students spend significantly more time reading the feedback addresses to SM question. Probably this is due to the complexity of the questions and the easiness to understand the feedback messages. SM questions involve more cognitive strategies to successfully complete the task than TB question, therefore, they are more complex. Despite these differences, we cannot establish any relationship with the impact of the different types of EF messages as we found no differences considering the experimental condition of the student.

Finally, our fifth hypothesis considered the result on *the Success Rate on Second Attempt* and on the posttest. No differences were found on the *Success Rate on Second Attempt* showing that all experimental conditions showed a similar pattern of success in a second attempt. We expected that students in the $EF_{Combimed}$ condition obtained better results in posttest than those in the other two experimental conditions. Results comparing the posttest scores depending on the experimental condition showed that for SM questions, those students in $EF_{Combimed}$ condition obtained significantly better results than those students in $EF_{Correct}$ and $EF_{Incorrect}$ conditions. The difference in posttest, though, can only be seen in SM question, which imply deep learning courses and more complex cognitive processing due to the relation and elaboration on the relevant concept information from text, inference strategies and the reliance on previous background knowledge (Cerdán & Vidal-Abarca, 2008; Cerdán et al., 2009). This finding suggests that an EF message which is coherent and consistent with the student mental model is beneficial for complex question, and it can promote transfer of learning of the new knowledge acquired during the training phase to new situations or scenarios in the posttest (Butler et al., 2013; Van der Kleij et al., 2015). It is very important to note that independently of the correctness of students' response (usefulness effect), what most benefits students to acquire knowledge after some delay, is that the message reinforces his answer when it is correct, and that the message explains why his chosen statement is incorrect after failing a question (coherence effect).

One important limitation of our study is that students, sometimes and in some question, spend very little time reading the elaborated feedback message. We have not found any explanation for this, but in future research we plan to ask their *Response Certitude* for each question, to examine if we find any explanation to why sometimes the feedback reading times are so short. To sum up, the findings of this study help us understand the role of formative feedback and the characteristics it should have to be efficient to help knowledge acquisition and transfer. It helps us to understand how complex formative feedback is processed by students in digital environments when they have to learn through the common task of answering questions from a text. The most interesting results for educational practice are that feedback should de coherent and consistent with the mental model created by the student when answering the question.

CHAPTER 6

STUDY 4. DIRECTING ELABORATED FEEDBACK MESSAGES TO CORRECT MISCONCEPTIONS: WHAT SHOULD AND EFFECTIVE MESSAGE INCLUDE?

1. OBJECTIVES AND HYPOTHESES

How students process a feedback message is one of the focuses of feedback research as it is a key element to understand this common tool in education settings and to make an effective instructional use of it. Throughout this dissertation we have been studying what type of elaborative instructional feedback is more effective to acquire declarative knowledge. On this final study we attempt to give response to the findings and limitations on study 3.

Our first objective is to replicate the results obtained in Study 3 with the difference that this study will only have 2 experimental conditions: $EF_{Correct}$ and $EF_{Combimed}$. We have decided to dispense with $EF_{Incorrect}$ as the only difference between the feedback messages in this condition and $EF_{Combimed}$ is when students give a correct answer, and we consider that the message given to students in $EF_{Incorrect}$ after a correct answer can be confusing and lacks research interest for our current objective. Therefore, in this investigation we would like to determine what type of feedback message is more effective considering variables such as the time spent reading the feedback message, the scores on first and second attempt and the scores on a posttest after some delay.

One of the findings we would like to investigate is the short amount of time that some students invest in reading the feedback message. Curiously, a very fast processing is not related with worse results on the second attempt, so our second objective in this study is to examine this effect. To do so, we have elaborated a *Response Certitude Scale* to study under what circumstances of their judgement of learning, they spend more or less time reading the feedback message. Response certitude is defined as the students' subjective judgement of certainty perceived by a learner regarding a piece of information or a task (Kulhavy et al., 1990) and it is an important individual variable to consider (Kulhavy & Stock, 1989).

Previous investigations report that response certitude has numerous limitations when students are told to rate in a continuum scale their confidence in giving a correct answer (Mory, 1992; 2004). This is shown by Nelson (1988) when he explains that student's response of certitude was related to their "feeling of knowing", a variable found to be subjective and unstable. It is easy to imagen that when rating a response certitude in a scale from 1 to 10, an answer of 8 will be perceived as different for different students. Also, having an 8 does not give us as researchers many information on the cognitive processes underlying this decision or any kind of objective information. Also, as it is a subjective criterion, students may over- or underestimate how confident they are, as it is difficult to be sure when choosing only one digit to express your certainty. To deal with the limitations listed above, we computed a 4 statement response certitude that allows us to gather more information and to create a more realistic picture about the students' thought and knowledge, as each statement expressed a different degree of ambiguity. For example, the first statement the students could select was "I am sure of the correct response". This statement suggest that the student has absolutely no doubt about the correct answer when answering the question, thus if the student finally gives an incorrect answer it is likely that he feels bewildered and pays more attention to feedback messages or uses additional aids. The next statement students could choose was “I am pretty sure but another alternative makes me doubt”. Contrary to the previous case, here the student expresses that he has a clear first option which he thinks is correct, although there is a second answer that could be correct. In this case, it is very likely that if he scores incorrectly on his first attempt, he will rush to select his second option as it is the one perceived a possibly correct, thus feedback reading times will probably be lower. The third statement they could choose was “I am not very sure because I doubt between two alternatives”. Here the student expresses doubt between two options, so he does not have one clear answer that predominates over the other, although after giving a wrong answer they will be likely to select their second option. Finally, the last statement says, “I am not sure because I doubt between various alternatives”. This suggests that the student does not have a clear first answer but also does not have a clear second answer in case of failing the first attempt. It is therefore more likely to use instructional aids as a mean to overcome this lack of knowledge.

Considering all this, we ask ourselves several research questions. Do students always spend the same amount of time reading the feedback message, or does it vary depending on if the student doubts between only two alternatives or if she is completely sure of the answer? Our prediction regarding this first objective is that those students who show very fast feedback reading times were doubting between two options but those with longer reading times express their doubt between several options or have no doubts of which statement is correct, independently of the experimental condition.

In view of the results in our previous experiments and the revised literature on this topic, we proceed to make six additional hypotheses to give response to the second broad objective of this study.

Firstly, we hypothesized that we will replicate the results in studies 1 and 2 that showed a consistent pattern on time reading feedback depending on the *usefulness* and *coherence* effect as well as their interaction. Concerning the usefulness effect, we predicted that participants would spend more amount of time reading feedback messages after they failed a question than after they succeed. We also predict that this will also depend on the experimental condition the students are enrolled in. We expect to find no significant differences between conditions after success, as the message they both received was identical and adjusted to their mental representation of the question. On the contrary, we do expect to find significant differences between conditions after failure as $EF_{Correct}$ does not include a coherent explanation with the student's mental representation whereas $EF_{Combimed}$ explains why the chosen statement is incorrect adjusting the explanation to the student's mental model.

As in previous studies, we expect that students spend more time reading the feedback message when this one is not consistent with their mental representation as they will find it harder to process ($EF_{Correct}$) than if the message gives an explanation adjusted to the student's mental representation ($EF_{Combimed}$). We expect that reducing the number of experimental conditions to the most relevant ones we find significant difference between these two conditions regarding the time spent reading the feedback message.

Additionally, we expect significant differences in the interaction of usefulness and coherence effect in time reading feedback. Although both experimental conditions are partially similar as after succeeding the EF was the same in both conditions (low usefulness and high coherence), we expect that the coherence in high usefulness feedback messages (after failure) shows significant differences across conditions as coherence between experimental groups varies.

Our fourth hypothesis, consistent with the findings on Study 3, we predict that there will be practically no searches made by any of the two conditions, and thus the difference between conditions will not be significant. Previous research has shown that the presence of elaborative feedback reduces drastically searching decisions when feedback is perceived as useful by the student.

Fifth, we hypothesize that as well as on our previous study, students will spend more amount of time reading feedback messages on SM questions than on TB questions. This difference will also depend on the experimental condition after failure (as it is the only case where they receive different EF explanations) as those participants on $EF_{Combimed}$ condition, will spend less time reading feedback messages in SM questions than those in $EF_{Correct}$.

In reference to response certitude, our fifth hypothesis was that those statement in the RC that expressed that another alternative of the question is making them doubt, would be more selected than those with complete certainty of the response of doubting between more than one alternative. Moreover, we predicted that students who show very fast feedback reading times were doubting between two options but those with longer reading times express their doubt between several options, independently of the experimental condition they are enrolled in. Reading times would also be longer on the cases where students were completely sure of the correct answer but then appeared to be wrong.

Finally, in reference to *Success Rate in Second Attempt*, we expect to replicate the findings of the previous study where we found no differences between experimental conditions. Nevertheless, we consider that all of the previous conditions will have an impact on learning in a posttest due to the recursive process of the task. We hypothesized that there will be significant differences on the post-test scores between conditions in favor of those in $EF_{Combined}$ condition, that will obtain higher performance scores. This is due to the influence of feedback and the coherence between the message and the mental representation on the student. The fact of receiving coherent information in the EF message with the active mental representation of the textual ideas as the student has understood it, after succeeding and after failing, will make students in $EF_{Combimed}$ retain successful answers and repair the error after committing a mistakes.

2. METHOD

2.1. Participants

A total of 97 first-year undergraduate's students enrolled in teaching studies completed the whole study without incidences. All students were Spanish native speakers. The data of two students were excluded from the analysis due to absences during one of the phases. The study was set as a voluntary class activity for those students who participated in the study, and they received extra credit. Students' confidentiality was preserved.

2.2. Design

The experiment followed a between subject's design with two experimental conditions that distinguished by the focus of elaborative feedback messages exposed to the students. Students were assessed on science background knowledge and assigned to one of the two experimental conditions following a matching procedure. There were 46 students in the $EF_{Correct}$ condition (EF to orient the student towards the correct response after failure, and reinforced the idea of their answer after a correct response) and 51 students on the $EF_{Combimed}$ condition (EF to help the student understand why their answer was incorrect after failure, and reinforced the idea of their answer after a correct response).

Students received different feedback messages depending on the experimental condition they were enrolled in. When students got a correct answer, the feedback message was always a paraphrase of the statement, to strengthen the important idea of the question by rewording the correct answer. This happened to all students, independently of their experimental condition. On the contrary, when students scored incorrectly an answer, the feedback message that appeared as a pop-up box varied depending on the experimental condition. Those students in the $EF_{Correct}$ condition received an elaborated feedback message which directed them towards the correct idea or concept, without stating the correct answer explicitly. Those students in the $EF_{Combimed}$ condition received a different elaborated feedback message depending on the alternative they selected when answering the question. This EF message explained why the chosen alternative was incorrect, without giving any information or clues on which one is the correct answer.

2.3. Materials

As in study 3, students were tested on *Science Background Knowledge*, a questionnaire with 30 True/False statements from three different topics: Atmospheric pressure, Heat and Temperature and General scientific knowledge.

For the learning task, they read one text on “Atmospheric Pressure and the Wind” and another one on “Heat and Temperature”, and then answered 8 Multiple choice questions per text on the learning phase and 8 open ended questions per text for the transfer test. Additionally, before answering the question, students had to choose a statement regarding their response certitude. Question statements and alternatives were not masked due to the characteristics of the experimental design. Students were tested using an improved version on the *Real&Learn* software. See the section “Materials” of study 3 for a detailed description of the materials and the software.

The measures considered for this study are the same as the ones explained in Study 3: *Initial Time Reading the Text*;*First Attempt Performance, Success Rate in Second Attempt* , Mean *Time Reading Feedback*; *Searching Reading Time* and *Score on Transfer Test*. Additionally, we included:

Mean Number of Searches which indicates the mean number of times that a student accesses the text to search. This will help us shed light on the results regarding the time spent searching the text.

Response Certitude. Thus, we registered the *Response Certitude*. This is an ordinal variable but, in some analyses, we have made use as a categorical variable, as we were interested in examining both, the relation and the differences between the categories of this response certitude scale.

2.4. Procedure

Students were tested during two sessions on 2 alternate days. In the first session, students were tested on previous knowledge using a 30 items questionnaire, *Science Background Knowledge*. The session lasted 15 minutes approximately. Students were matched to one of the two experimental conditions ($EF_{Correct}$ or $EF_{Combimed}$), taking into account their scores on the previous background knowledge questionnaire, to assure an equal level of science background knowledge between the experimental conditions.

The second session took place in the computer-lab classroom, were students performed the learning phase task and the transfer test individually. This session lasted approximately 90 minutes.

The procedure was the same as in Study 3 (see the section “procedure” in Study 3, for a detailed description) with one fundamental variation. This time, students had to select a statement regarding their response certitude to answer correctly that question, right before answering it (see Figure 1). In other words, students first saw the question and had to select an option regarding how confident they were in answering correctly, and right after, they had to select the correct answer for the question. This was done for each question on both texts. Once they finished the training phase, a short delay activity took place before students could move on to the transfer test. This activity consisted of watching a TED talk of an unrelated topic and lasted approximately 20 minutes. Once the video was finished, students turned off the computers and answered in paper-pencil format the 8 open ended questions related to each text in counterbalanced order.

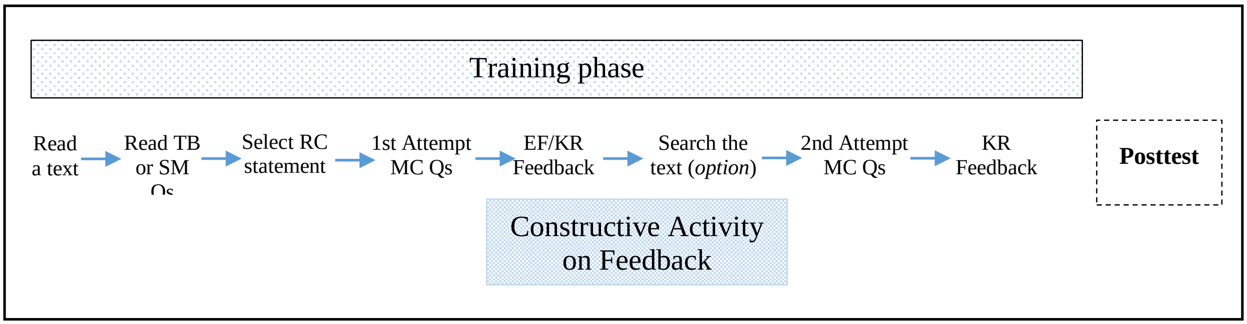


Figure 1: Procedure of the experimental phase in study 4.

2.5. Measures

The measures considered for this study are the same as the ones explained in Study 3: *Initial Time Reading the Text*;*First Attempt Performance*, *Success Rate in Second*

Attempt , Mean *Time Reading Feedback*; *Searching Reading Time* and *Score on Transfer Test*. Additionally, we included:

Mean Number of Searches which indicates the mean number of times that a student accesses the text to search. This will help us shed light on the results regarding the time spent searching the text.

*Response Certitude.W*e registered the *Response Certitude*, an ordinal variable but, in some analyses, we have made use as a categorical variable, as we were interested in examining both, the relation and the differences between the categories of this response certitude scale.

3. RESULTS

Preliminary analyses were carried out before conducting the main analyses to check the data distribution of the test scores in the three phases of the study (*Pretest*, *Training* and *Posttest*). Also, we tested for differences between conditions computing a Student t-test on those variables that were equal across conditions as there was no experimental manipulation involved to prove that *Science Background Knowledge*and *Initial Time Reading the Text* had no impact on the studied variables ($p > .05$).

To test our first three hypotheses, we analyzed *Mean Time Reading Feedback* messages by the students in each condition after success and after failure. To do so, we computed a 2x2 Mixed-ANOVA introducing as dependent variable *Time Reading Feedback*, as between-subject variable the experimental condition and within-subject variable the success or failure in answering.

To test our first hypothesis regarding the usefulness effect, we compared the *Mean Time Reading Feedback* depending on the success of the task. In this case, we found significant differences [F(1,90) = 28.678, p = .000, η^2 = .242], showing a strong effect size. When students gave a correct answer, information on the feedback message was read faster (M = 3.03, SD = 1.62) than after an incorrect answer (M = 4.73, SD = 3.15), confirming our first hypothesis.

To test the second hypothesis regarding the easiness of the processing of the feedback message due to the *coherence* effect, we compared the *Mean Time Reading Feedback* across experimental conditions when they failed the question. Results indicate that those students in $EF_{Correct}$ spend significantly more time reading the feedback message (M = 5.450, SD = 3.086) than those in $EF_{Combined}$ (M = 4.017, SD = 3.086) p=.028. In this new experiment, when they get the question right, both groups receive the same feedback message, thus not observing significant differences in the time spent on feedback in that situation.

Yet, the third hypothesis regarding the interaction between the usefulness and coherence effect tested previously was not significant [F(1,90) = 1.939, p = .167]. Although the $EF_{Correct}$ increases, in greater proportion, the time spent reading feedback after an error, this increase is not significant.. This happenedin Study 3 as result were closed to significance level although did not reach it . The effect may be since the conditions are not identical to the first experiments, as here there are no $EF_{Incorrect}$ condition since the success was incoherent by raising the time, and here both conditions are consistent in success. Additionally, it may be due to the increased fatigue of the students as a consequence of increasing the demands of the activity due to including and additional text and more questions in total.

To give response to our fourth hypothesis, we examined the effect of searching. We predicted that there will be very few searches as EF messages are perceived to have enough information to answer on a second attempt without using an additional aid such as searching. Thus, we hypothesized that the *Searching Reading Time* done by any of the two experimental conditions would be very low. In fact, data shows this as the *Mean Number of Searches* done by students in $EF_{Correct}$ condition is 0.15 (*SD* = 0.47) and in $EF_{Combined}$ 0.23 (*SD* = 0.81). This difference was not significant, through a t test, between experimental groups ($p > .05$). To gather more information, we computed a new Student t-test to examine if these results were identical. If we analyzed the time searching depending on the *type of question*, paired sample t-test. and we found no differences in the *Searching Reading Time* depending on TB or SM questions.

In order to test for differences in feedback reading times on TB and SM questions, we proceeded to compute Mean *Time Reading Feedback* across conditions. As we mentioned in our hypothesis, these analyses are only done after failure as experimental conditions receive equal feedback messages after success. Thus, we computed a 2x2 Mixed-ANOVA introducing as dependent variable Mean *Time Reading Feedback*, as between-subject variable the experimental conditions and as within-subject variable the *Type of Question (TB or SM).* First, we found significant differences in the *Type of Question* [$F(1,95) = 32.436$, $p = .000$, $\eta^2_p = .255$] proving that students spend less *Time Reading Feedback* in TB questions (M = 2.92, SD = 3.64) than in SM questions (M = 7.46, SD = 7.37) showing a mean effect size. According to our expectations, students spend more time Reading the feedback message in SM questions. Nevertheless, no significant differences were found in the interaction [$F(1,95) = .859$, $p = .356$],. In the $EF_{Combined}$ condition, students spend less time in TB questions and more in SM questions in similar proportion as those students enrolled in $EF_{Correct}$ condition. Therefore, there is no difference between both experimental conditions,

Fourthly, regarding response certitude (RC), the RC selected more times in both experimental condition is the second option, "*I am pretty sure but another alternative makes me doubt*" whereas the RC selected less times in both experimental conditions is the fourth option, "*I am not sure because I doubt between various alternatives*". Additionally, a Student t test was computed to test for mean differences in the number of times each RC was selected by condition with non-significant results ($p > .05$). All these results enable us to state that there are no differences in the distribution of RC by experimental condition.

Thus, we continue by analyzing the differences in Mean *Time Reading Feedback* messages depending on the RC item selected. Results to a 2x2 Mixed ANOVA analysis with dependent variable *Mean Feedback Reading Time, when error,* as between-subjects variable the experimental condition and as within- subject variable the RC selected, show that there are no significant differences in feedback reading times depending on the RC item selected [$F(3, 23) = .588$, $p = .625$]. Now, we proceed to analyze the interaction of these variables to see the impact of RC on *Time Reading Feedback* depending on the experimental condition. Results show there is no significant differences between conditions [$F(3, 23) = .061$, $p = .807$], as predicted. On Table 1 you can find the mean times reading feedback depending on the experimental condition and the response certitude selected.

Table 1. *Mean times and standard deviations of time reading feedback depending on the RC selected.*

RC	**Condition**	**M**	**SD**
RC1	1	5.127	5.363
	2	4.257	2.531
RC2	1	4.357	3.306
	2	3.007	1.802
RC3	1	4.539	2.962
	2	3.405	3.617
RC4	1	4.682	4.911
	2	3.114	2.893

Our sixth hypothesis, we were interested in analyzing the success rates after reading the different feedback messages depending on the condition in which the students were enrolled.. Thus, the *Success Rate in Second Attempt* was introduced as a dependent variable in a new t test for independent samples. We did not find a significant difference between groups, in this case contrary to our prediction both groups had a similar performance; $EF_{Correct}$ group (M = .221, SD = .117) and $EF_{Combimed}$ (M = .203, SD = .120).

Finally, to give response to our final hypothesis, we analyzed and compared the performance of both experimental groups in *Score on Transfer Test*. Results indicate that there were statistically significant differences between $EF_{Correct}$ and $EF_{Combined}$ on the number of scores on the posttest [*t*(91) = 2.332, *p*. = 022, *d* = 0.352.] Those students in *combined* condition score significantly higher (*M* = 9.01, *SD* = 4.41) than those in $EF_{Correct}$ condition (*M* = 7.72, *SD* = 2.70).

4. DISCUSSION AND CONCLUSIONS

The present investigation had two main objectives. First, to replicate the findings on Study 3 to confirm that $EF_{Combined}$ is the most effective type of feedback, as suggested in previous findings. Our second objective was to study in more detail the amount of time devoted by students to read the feedback messages provided to them after each answer analyzing their response certitude.

To shed light on these objectives, we elaborated six hypotheses to give response to the major aspects that we consider most relevant to be able to determine what type of feedback is more efficient when learning declarative knowledge from textual information following an online instructional design. These are: the *usefulness* and the *coherence* effect as well as their interaction; the amount of time dedicated to read the elaborated feedback message; the influence of searching on feedback; the impact of question type on feedback reading times, and the effect of different types of feedback messages on the final performance outcome on a transfer test after a short delay.

Feedback reading times tend vary vastly within and between students, thus we wanted to explore if this variation in the amount of time spent processing the feedback message had any relation with their response certitude. Consequently, we examined the relation between the amount of time students spend reading the feedback messages after each question and if they were sure of answering correctly the question or between how many alternatives they were doubting.

The time students devote to reading feedback varies depending on the experimental condition and therefore, the content of the elaborated feedback message provided to the students. As we hypothesized, those students in $EF_{Correct}$ condition spend significantly more time reading the feedback message than those in the $EF_{Combined}$

condition, confirming our prediction. With respect to the effect of success and failure on *Feedback Reading Times,* we predicted that students will invest more time in processing the feedback message after failure, and this will also depend on the type of feedback message received depending on the condition the student was enrolled in. We found no significant differences between conditions after success, this is positive as both conditions received the same feedback message after giving a successful answer. Additionally, as we predicted, after committing an error, students in the $EF_{Correct}$ condition spent significantly more time reading the feedback message than those in the $EF_{Combined}$ condition. Contrary to our expectations, the interaction between usefulness and *coherence* effect was not shown to be significant. Although we consider that when we give an erroneous answer, it is easier to process the information if it fits with our mental model built to answer the question, than when the information guides us towards the correct answer but does not adjust with our initial thought as it is not coherent with it, we have to consider the specific characteristics of this experiment that might have interfered. The difference between both conditions was very reduced and thus, very difficult to observe differences and measure the impact of the feedback message. Both conditions contained an elaborated feedback message with a similar length in number of words. Additionally, both conditions contained the same message after success so the only difference between them was after failure what makes it harder to observe differences in the interaction between the usefulness and coherence effect.

Regarding the effect of searching, we have seen on the previous study and on this present one that there are few searches made by students. Gil et al. (2015) gives evidence of the impact of searching decisions in performance, suggesting that it is as important the number of accurate search decisions as the number of accurate no-search decisions to explain scores in question answering tasks. This goes in line with our initial thought explained in our prediction, as the feedback message containing elaborated information already gives information about content related to the question. What this shows is that independently of the effectiveness of the type of EF, the perception of the student is that the EF message is useful and reliable, as they decide not to search for additional information, and they answer the second attempt with only the additional information facilitated to them through the EF. Thus, although it takes more time for a student to read and process an elaborated feedback message which is incoherent with their mental representation, it is still perceived as useful, as no other available aid has been used.

Concerning the impact of question type on feedback reading times, we hypothesized that students will invest more time to process the feedback messages after failure for SM questions than for TB questions and that this differences will be notable across experimental conditions. According to our expectations, student spent more time reading feedback for SM questions. This may be because SM questions involve higher order processes in comprehension such as integrating different textual ideas with one’s prior knowledge, or being able to transfer the knowledge acquired from one situation to another different one. Thus, students spend more time processing the information in this type of questions rather than in TB questions where information needed to answer the question appeared on the text. Likewise, RC might have made students more aware of their reasoning and thus encourage them to spend more time processing more complex information.

To examine in more depth the feedback reading times and to explore if student’s spend more time reading the feedback messages when they have more doubts about giving a correct answer we decided to include on this final study the metacognitive variable *Response Certitude*. The Oxford English Dictionary (1933) defines certitudes as, "Subjective certainty; the state of being certain or sure of anything," and this is what we have observed when analyzing the results. We did not find any clear pattern that related the certitude of a student when answering with the time they dedicate to reading and understanding the feedback message. We consider that this is the case due to the inherent subjectivity of this metacognitive variable which has shown to have some underlying problems.

The aim of this variables is to understand the student’s metacognitive process of predicting his or her performance on a task and how this affects the reading of feedback messages, considered a beneficial aid to enhance learning and performance. In research, response certitude has also been said to be the "feeling of knowing" (Butterfield, Nelson & Peck, 1988; Metcalfe, 1986; Nelson, 1988; Nelson, Leonesio, Landwehr, & Narens, 1986). Although this variable has been found to have some benefits, researchers have found that the stability of an individual's feeling-of-knowing accuracy fluctuates significantly (Nelson, 1988).

Additionally, there are many inconsistencies in the results of response certitude studies. Thus, although our intention was to shed some light, our results gave few information. On this line, Metcalfe (1986) found that in memory tasks, students were able to predict accurately their feeling-of-knowing but, however, overestimated their response certitude on high level questions or problem-solving tasks. In the contrary, other studies (Dempsey & Driscoll, 1996) have found contrary results, finding that students tend to underestimate their feeling-of-knowing. So, this shows that although literature and research has focused during many years on studying the benefits and limitations of response certitude, there has not been consensus on the findings. This show that the subjective perception of the students varies widely within and between tasks and learners find challenging to assess accurately their abilities. This might be why we have not found recent studies which focused on analyzing the impact on RC on feedback reading times.

Finally, transfer test performance results demonstrate that those students in the $EF_{Combined}$ condition scored significantly more right answers than those in the $EF_{Correct}$ condition, a difference of 1.29 correct answer between experimental conditions. Thus, we can state that EF made of explanations is more beneficial than KCR to transfer knowledge (Butler, Godbole, and Marsh, 2013), and moreover, if the explanation given in the EF fits with the students' mental representation, by reinforcing their idea after a correct answer and explaining their error made after an erroneous answer, the beneficial effects on the post test will be increased.

All of the above supports the idea that $EF_{Combined}$ is an effective feedback tool, as when the elaborated message gives information to the student fits with the mental representation, he builds to answer the question on the task, this one is easier to process and helps students to acquire the correct ideas for future successful transfer of knowledge to other situations.

It is interesting to note that although we have not been able to accept all of our initial predictions, results in time reading feedback after failure and results in performance in the post-test seem to indicate the need of students to refute the error before giving a new response, as the processing is faster and performances after a short delay is better. The studies with refutation texts (Diakidoy & Kendeou, 2001; Diakidoy et al., 2003; Kendeou & O'Brien, 2014; Kendeou et al., 2014; van den Broek & Kendeou, 2008), show us the consecutive phases to repair errors in the mental model, and following the KReC framework, we consider that $EF_{Combined}$ is the most adjusted type of EF to deal with misconceptions when students fail an answer and also reinforce students’ response after a correct answer. We consider this an interesting and important result both for research and the educational communities as it is a new way of dealing with misconceptions and enhancing performance.

A limitation of this study is the sample size, as although it is acceptable for the number of experimental conditions inherent in this study, increasing the sample size would make results stronger. Another limitation of the study is the increase of task demands and the consequence this has had on student’s engagement. We already saw a decrease in feedback reading times in Study 3 which we can now suggest it might be due to the increased duration of the task as it has double the texts and increased largely the number of questions. After studying the result on the present study, we can suggest that this was the case also now, and it has been enhanced due to the task demands. The fact of incorporating RC to the experimental procedure has increased by two the demands for the student’s as before answering the question they had to think not only of the correct answer but to create a mental dialogue where they had to select a response certitude statement as an outcome. This has likely been perceived as a demanding task and engagement has been negatively affected by this.

Finally, future research should examine the long term effect of RC and formative feedback, ideally on natural settings and with only one text We consider that RC is a variable that has to be studied and result on the present study are useful to create different guidelines and procedures for next studies. Additionally, it would be very interesting to compare the results we obtained with our RC to those with a numerical RC which might be easier to process and understand for students.

CHAPTER 7

GENERAL DISCUSSION

DISCUSSION

Feedback on learning outcomes has been studied during decades, although there is not so many research focused on the effects of formative feedback in learning from text with question-answering task, which however, is one of the most common instructional strategies in educational setting. Formative feedback gives information to the learner on his current performance to facilitate learning by making possible cognitive or behavioral changes (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mathan & Koedinger, 2005; Narciss et al., 2014; Shute, 2008) as it is considered the variable with one of the strongest impact on performance. Nevertheless, its effectiveness has been widely studied in digital environments, using different types of task and different types of feedback messages. According to the meta analysis conducted by van der Kleij et al., 2015, EF messages are the most effective ones in knowledge acquisition, followed by KCR and lastly KR messages. The authors also conclude that EF is more effective when the questions involve higher order processes such as situation model questions which involve connecting different ideas exposed on the text or transferring the textual ideas to new situations or contexts (e.g., Butler et al., 2013; Hattie y Timperley, 2007; Kluger y DeNisi, 1996; Narciss, 2004; Shute, 2008).

Additionally, the role of feedback has also been shown to have a beneficial effect when students rate their response certitude while answering questions from a text. Response certitude is a metacognitive variable that predicts comprehension Kulhavy and Stock's model (1989), When students receive elaborative feedback after answering a question, they make more accurate predictions of their response certitude reducing the distance between their perceived response certitude and the correctness of the response (Butler, 2008).

EF may contain very varied and different types of information. In the studies presented in this dissertation, EF was given automatically to the students after they answered each question using the *Read&Learn* software. These messages contained KR information plus an elaboration which varied depending on the experimental condition. On the first two studies EF could be oriented to the correct response ($EF_{Correct}$) or oriented to correct the misconception ($EF_{Incorrect}$). The first type gave information related to the idea involved in the correct response, whereas the second type explained why the chosen statement in the MCQ was incorrect. This second type meets the principles in the KReC framework (Kendeou & O'Brien, 2014) which state that 5 principles of this framework have to happen in order to assure the reparation of misconceptions. First, the *principle of coding*, suggests that information encoded and stored in the LTM cannot be eliminated but we can only reduce its level of activation by the mechanism of propagation of information as the *principle of passive activation* proposes. Next, the principle of co-activation, notes that both the correct and the incorrect information must be activated at the same time in order for revision to happen. Subsequent, the principle of integration states that the explanation of the new correct information should be active at the same time as the erroneous one to be able to meet the *principle of competition* were both the new correct information and the erroneous one will compete to be the most active information. Thus, contemplating this information, we can consider $\text{EF}_{\text{Incorrect}}$ type of feedback message a form of refutation as we can find that the incorrect idea, an explanation refuting the incorrect idea and the correct idea are activated at the same time when a student gives an incorrect answer. An additional type of EF included in the last two studies was $\text{EF}_{\text{Combined}}$, which showed students $\text{EF}_{\text{Correct}}$ messages after giving a correct answer and $\text{EF}_{\text{Incorrect}}$ messages when their selected choice was incorrect. This way, the EF the student received was always consistent with the mental representation created from the text. If the student is correct, $\text{EF}_{\text{Correct}}$ will reinforce his idea and give more information related to this correct idea, whereas if the student gives an erroneous answer, $\text{EF}_{\text{Incorrec}}$t will explicitly refute the specific content of the chosen statement which is active in the students' mental representation, but it includes incorrect knowledge.

The general objective that has guided this work has been to analyze the effectiveness of different types of formative feedback messages to enhance learning of declarative knowledge from a question-answering task using the *Read&Learn* software. An important novelty of the present dissertation was using online measures to examine feedback processing in terms of reading times. To do this, we carried out 4 experimental studies with specific objectives. The first two studies used one scientific text to analyze the differences between KR feedback and two types of EF messages ($\text{EF}_{\text{Correct}}$ and $\text{EF}_{\text{Incorrect}}$) when students were able to search on the text after failing an answer and before answering on the second attempt (study 1) and without being able to search in the text for additional information at any point of the study (study 2). The last two studies included two different experimental texts to analyze the differences between 3 different types of EF messages. The first two types were the same ones used in studies 1 and 2, but the third type was a combination of both. Students in this third experimental condition ($EF_{Combined}$) received $EF_{Correct}$ after giving a correct answer and $EF_{Incorrect}$ after giving an erroneous answer, receiving the most adjusted information to the students' mental representation in that exact moment. Additionally, the last study also included the variable response certitude. All actions and reading times were registered in the *Read&Learn* software to be later analyzed.

The objective of study 1 was to analyze the differences in effectiveness between KR feedback and the two types of EF messages on undergraduate students' performance on a task-oriented reading activity. Specifically, in addition to exploring the effectiveness of EF, we wanted to study how students process these feedback messages depending on three different context variables: the correctness of their answers, the text availability and the type of question. Results led us to ask ourselves what would happen if we replicated the experiment without giving the students the option to search on the text for additional information before the second attempt after failing the question, thus being the feedback message the only instructional aid after reading initial reading of the text. This was the objective of study 2. Once we isolated the effect of instructional feedback, results indicated that the effectiveness of EF in the learning phase is stronger compared to KR when students are not able to search of the text, as those students in the KR cannot get additional information from any source, although these findings were not significant in delayed learning according to the results in transfer test.

At this point, we decided to revise the materials used, shorten the text we were using and include an additional science text to our investigation. We also made decisions regarding the experimental conditions and we decided to focus only on EF, maintaining both EF conditions used in study 1 and 2, and including $EF_{Combined}$ condition.

The objective of study 3 was to investigate what type of EF ($EF_{Correct}$, $EF_{Incorrect}$ or $EF_{Combined}$) was more efficient to help students learn in question-answering tasks both the learning phase and after a short delay, and the impact on the correctness of their answers, the text availability and the type of question. Finally, in study 4, we used the same materials and procedure to replicate the findings on study 3 and to investigate why some feedback reading times were so short and if this depended on if the students had a second option in mind when they were told that their answer was incorrect. To try to give answer to this issue, we included a response certitude variable.

The set of results obtained in the four studies allows us to formulate several general conclusions related to both effectiveness of different types of EF in answering comprehension questions and learning in digital environments and the decisions they make when processing these EF messages.

Firstly, we would like to point out that in instructional environments, the use of feedback given by a teacher or by a digital tool becomes a mean of communication between the learner and the instructor with the aim of cooperation. Grice (1975) in its classic general principle of cooperative communication affirms that the sender of information (in this case, the instructor in classroom or virtual setting) must adapt his conversational contribution in the stage where it takes place, the requirements that mark the purpose or direction of the exchange that is sustained. The four assumptions underlying this general principle are: (1) Maximum quantity, (2) maximum quality, (3) maximum relevance and (4) maximum mode or manner. If we apply these principles to formative feedback, the cooperative communication presented to the students as a feedback message should have (1) the adequate length, not too concise nor with too much information, (2) true information, (3) necessary and useful information and (4) clearly expressed and non-ambiguous information.

When students receive formative feedback, they use this information to verify or refute their task model, and to update it if they consider necessary as we saw in the Question-Answering Model (Vidal-Abarca et al., 2019). Of course, depending on the type of feedback received by the student the cooperative communication principles will be covered or not. If the student only receives KR, he will use feedback to verify or refute his answer although it will be more difficult for him to know which one is the correct answer or why his answer is wrong. On the other hand, when additional information is given to the students in form of an EF message, it helps the student to update the mental representation of the idea tested in the question and to reduce the knowledge gap between his answer and the correct answer (e.g., Hattie & Timperley, 2007). Metacognitive strategies such as thinking on how sure they are about their new answer being correct, or evaluating their own comprehension help the student update the mental model after receiving feedback.

The totality of comprehension theories state that in order comprehension to be successful, the mental representation formed by the learner will result in a coherent mental representation were the textual elements such as concepts and ideas will be integrated with the reader’s prior knowledge through semantic relationships (McNamara & Magliano, 2009). Thus, coherence is a key element to assure comprehension. Likewise, when we speak about the feedback messages, it is also important that they are coherent with the text and the student’s mental representation to assure its helpfulness in improving performance and maintaining correct knowledge.

Therefore, an effective EF should be adapted and useful, helping students to construct knowledge, especially when dealing with complex concepts or when students may have erroneous previous ideas (Gadgil, Nokes-Malach, & Chi, 2012; Kendeou, Walsh, Smith, & O'Brien, 2014; Posner, Strike, Hewson, & Gertzog, 1982). According to Roper (1977), feedback is ideal to correct the learners' errors.

A feedback message is coherent when it addresses the textual idea and the idea we have outlined about a certain topic (our mental representation), independently of the correctness of this last one. Thus, the effect of *coherence* in feedback suggests that when the information given to the learner is coherent with the student's response, it will be easier to be processed as it will be easily integrated in his mental representation through the semantic relationship of the feedback message and the learner's idea.

In the present dissertation we have explored how variations in coherence affect the processing of different feedback messages in our four studies after giving a correct or an incorrect answer. The use of feedback has been shown to de different depending on the correctness of the students' answer. It is important to note that in studies 1 and 2 after a successful answer, the students who are in the $EF_{Correct}$ condition receive information that supports their response and that is cohernt with their mental representations build to answer the question. However, those students in the $EF_{Incorrect}$ group received information than can be somehow confusing, as the message they receive after a success gives them an explanation about why the other statements are not correct which is not coherent with his mental representation. The reverse situation occurs when the students answer is wrong: the students in the $EF_{Correct}$ group receive a directed to the correct answer by giving some kind of hint or guidance, but this information does not allow them to know the reasons why their selected answer is incorrect, causing incoherence between their selected answer and the idea the message holds and, therefore, it does not help the student to update the task model and build a correct mental representation. On the contrary, the students in the $EF_{Incorrect}$ group, after committing an error receive information that helps them to understand the error made, so the information will be coherent with their response, making it easier for them to update their task model and refute their last response. To overcome this limitation, in studies 3 and 4 we added $EF_{Combined}$ condition. Those students enrolled in $EF_{Combined}$, were always given a coherent feedback both after success and after failure. This was the case because the aim of this type of feedback was to provide always coherent feedback, so the message was identical to $EF_{Correct}$ after success and to $EF_{Inorrect}$ after failure.

This coherence effect was only partially verified in the results of studies 1 and 2, as the *coherence* effect was only shown in study 1, showing greater reading times of feedback for those students in $EF_{Correct}$ condition than for those in $EF_{Incorrect}$ condition. On the contrary, when the text was not available after an error as in study 2, these differences were reduced. Given that the questions and the messages were identical, in both studies, we consider that this difference between studies, must not be related to coherence must be due to the variable text availability, which we will discuss at a later point. In studies 3 and 4, those students in $EF_{Combined}$ always had a coherent message. Results support the coherence effect as feedback reading times were greater in $EF_{Correct}$ than in the other two experimental conditions.

This is linked with the usefulness effect of feedback that suggests that the time students spend reading the formative feedback is related to the degree to which the information is relevant and useful to perform the task (Stobart, 2008). Our studies confirm this fact as students who received some kind of explanation ($EF_{Correct}$, $EF_{Incorrect}$ and

$EF_{Combined}$) spent more time reading the feedback after an error than those students who only received KR, when the information could be used to give a new answer. Nevertheless, we also found that students in $EF_{Incorrect}$ and $EF_{Combined}$ spent less time reading the feedback message after an erroneous answer than $EF_{Correct}$ condition, although their performance was better in a second attempt and in a transfer test. Thus, this is related with the coherence effect.

Regarding text availability, previous research in reading states that an available text is an important context variable that has an influence on learning by facilitating or hindering comprehension (Ozuru et al., 2007; Schroeder, 2011). The studies in this dissertation have shown that engaging in the feedback messages shown after each question is a metacognitive strategy dependent on other variables such as text availability. If we compare studies 1 and 2 we observed that more searches on the text took place when students lacked information (i.e., they were enrolled in KR condition) in study 1, and that performance for this homologous condition in study 2 was diminished as the text was not available. These results show that when we lack an efficient feedback message, more searches on the text take place to counterbalance the lack of information. Additionally, if we look at the number of searches in studies 3 and 4 were all the conditions were exposed to EF we can see that the number of searches was drastically reduced. This can be considered as an index of perceived *usefulness* of feedback by the student, as not searching seems to indicate that the student considered the EF message is enough informative to update the task model and give a new response. Considering the lack of feedback studies with the objective of analyzing how an available text affects reading behavior and question-answering performance we consider this is a notorious finding.

Furthermore, in a similar study by Mañez (2019) were students had to make de decision of accessing or not the elaborative feedback message, it was found that access to the EF was more frequent when they had not been able to consult the text before answering the question. Therefore, as we outlined before, it seems that EF and text availability are two similar types of aids and that when the student receives one of them, the possibility of using the other is reduced.

We can assess the understanding of text-base and situation model discourse representation by making surface level and deep level questions from a text. On the one hand, text-base questions (or surface level) are related to an idea that explicitly appears on the text and on the other hand, situation model questions (or deep level) make the learner form inferences between ideas and previous background knowledge. EF messages have been found to be more effective for situation model questions as found by Butler et al., (2013) during transfer phase, as they give more information to the learner so he can form a correct mental model to answer the question correctly. In our studies we have found similar results, as students spent more time processing the feedback message for SM questions although differences did not reach significance level throughout the studies in this dissertation. If we link this to the *coherence* effect, we can state that providing EF on situation model questions which imply high order cognitive processes benefits performance (i.e., Hattie y Timperley, 2007; Kluger y DeNisi, 1996; Mason y Bruning, 2001; Mory, 2004; Narciss, 2004; Shute, 2008 van der Kleij et al., 2015) and support the evidence showing the effectiveness of using SM questions to improve learning over TB questions (Cerdán et al., 2009; Hamilton, 1985; Ozgungor & Guthrie, 2004; Paris, Lipson, & Wixson, 1983; Sagerman & Mayer, 1987).

Regarding response certitude, we did not find significant differences between conditions on RC or on the time reading feedback depending on the RC, but we did find significant differences between $EF_{Correct}$ and $EF_{Combined}$ in the time spend answering RC. Previous research suggests that the information presented in the feedback message has an important instructional function for the rectification of incorrect responses and the conservation of correct responses over time (Kulhavy & Stock, 1989; Hancock, Stock, & Kulhavy, 1992). Although students in our study 4 first reported their RC and after they validated their answer and saw the EF message, it is well known that results in one question may affect the subsequent questions.

We believe that the most plausible explanation of this fact is related to the process of refutation, as set out in the KReC model (Kendeou & O'Brien, 2014), given that before assimilating a correct idea the student must repair the error by combining the information of the error with its refutation by means of an explanation. This model could be used to understand how the student processes the EF after answering questions. This assertion is also supported by the analyses of time reading the EF, since students in $EF_{Correct}$ groups need more time after errors to process the feedback message that allows them to confirm their knowledge or to correct possible metacognitive inaccuracies (Butler et al., 2008), and this increase in time is not associated to an improvement in performance in second attempt or in a posttest after a small delay. In this sense, the $EFI_{ncorrect}$ more adjusted and easier to process since it has at the same time slot the necessary information to carry out the refutation process, challenging the students' cognition and guiding them to the use of more cognitive resources to understand the error (Maier et al., 2016). Our results seem to indicate that the student processes the EF in less time and additionally retain better this conceptual change improving their performance in a transfer test after a short delay.

However, the aids guiding the student towards to the correct answer after an error, could be hindering the refutation process, since they provide information that is not necessary at that moment, which could lead to an overload of the WM. Thus, although the students' objective when answering questions from a text is to give correct answers and obtain good results (Farr, Pritchard, & Smitten, 1990; Rouet et al., 2017; Rupp, Ferne, & Choi, 2006; Vidal-Abarca et al., 2010), our results show that learning outcomes are improved after receiving $\text{EF}_{\text{Incorrect}}$.

Additionally, it is very important to focus on the specific characteristics of $\text{EF}_{\text{Combined}}$ and the outcomes observed in this type of EF. As we noted before, the studies in this dissertation prove the beneficial effect of conceptual change through refutational feedback messages, but study 4 sheds light on an important result regarding $\text{EF}_{\text{Combined}}$ messages. This last condition shared EF messages with $\text{EF}_{\text{Correct}}$ after success and with $\text{EF}_{\text{Incorrect}}$ after failure. Even though EF messages in $\text{EF}_{\text{Incorrect}}$ and $\text{EF}_{\text{Combined}}$ after failure were exactly the same, we only found significant differences bertween $\text{EF}_{\text{Correct}}$ and $\text{EF}_{\text{Combined}}$. This very important result suggests that although giving refutational feedback to favor conceptual change is a key element to deal with misconceptions, it is not the only element that we should focus on. It is essential for students to obtain EF messages which are adjusted to their mental representation also after giving a correct answer. This paraphrasis of the correct idea helps the student consolidate and maintain the correct knowledge and be able to transfer it to homologous situations in a delayed learning test. Thus, we consider that the key element of feedback is the coherence between the message given and the last active idea of the student's mental model.

To sum up, the results of the studies in the present dissertation show that formative EF messages are useful to help student in task-oriented reading activities with the objective of learning. Additionally, this dissertation sheds light on the benefits and limitations of different types of EF, concluding that the most beneficial type of feedback is the one which is more adjusted to the students' mental representation, as it is easier to understand by the student and facilitates reinforcement of correct ideas and refutation of misconception. Thus, the most beneficial type of EF for the students who participated on our dissertation was $EF_{Combined}$. Our results seem agree with other researchers results, student's knowledge or information cannot be replaced by new information, thus students need to refute their previous ideas so new information becomes more active due to understanding (Dunbar, Fugelsang, & Stein, 2007; Kendeou & O'Brien, 2014). This is especially relevant if we consider electronic environments as they give us the possibility to give immediate automatic feedback when the students' knowledge is still active and this way have a stronger impact in delayed performance and knowledge acquisition. It is also important to point out that when students are given EF, the number of searches to the text diminish drastically compared to when they only receive KR feedback, as they are perceived as complementary aids and text availability can compensate for the lack of feedback. Hence, the four studies have help us to step forwards in the knowledge of use, processing and characteristics of effective formative feedback messages in digital environments.

Limitations

The present work is not exempt from limitations that should be addressed and considered for future research. In the first place, although we tried to include this task as a classroom activity of the subject "teaching science" for students enrolled in teaching studies or in "instructional and educational psychology" for those students coursing psychology studies, it was not always possible. Students showed their interest in collaborating with us, although the motivation and interest decreased once the experimental procedure was explained. Participating in the experiment, in most cases, was awarded with a bonus point in the subject, which increased participation but not necessarily engagement in the task and feedback processing is likeky to be affected (golke at al. 2015). Additionally, we have observed a decrease on time reading EF messages in studies 3 and 4 when we compare the *Mean Times Reading Feedback after failure* between the four studies. In the first study, those students enrolled in $EF_{Correct}$ read the feedback message for a mean time of 14.021 (SD = 9.145) seconds and the $EF_{Incorrect}$ for a mean time of 7.915 (SD = 3.981) seconds. These times were similar to those obtained in Study 2, as students enrolled in $EF_{Correct}$ read the feedback message for a mean time of 10.243 (SD = 8.248) seconds and the $EF_{Incorrect}$ for a mean time of 7.165 (SD = 5.046) seconds. A decrease in *Mean Times Reading Feedback* is observed in study 3 were those students enrolled in $EF_{Correct}$ read the feedback message for a mean time of 4.879 (SD = 3.468) seconds, the $EF_{Incorrect}$ for a mean time of 3.623 (SD = 1.715) seconds and the $EF_{Combined}$ for a mean time of 3.367 (SD = 2.243) seconds. This was similar to the feedback reading times in study 4 as those students enrolled in $EF_{Correct}$ read the feedback message for a mean time of 5.450 (SD = 3.086) seconds, the $EF_{Combined}$ for a mean time of 4.017 (SD = 3.086) seconds. This decrease shows that in longer tasks, with more texts and more questions, students reduce the time they devote to reading the feedback message. Probably, other variables such as engagement or motivation might have been affected by diminishing feedback reading times.

In second place, although we arranged the experimental procedure considering the time limitations we had, it was a handicap for the transfer test as between the learning phase and the transfer test there was only a 15 minutes' gap. Longer times between the learning phase and the transfer phase seem to have a stronger impact on memory and retention (cepeda et al 2009 and shute 2008) There is many controversies in among researchers in how much time is considered enough delay to be able to categorize the findings as transfer learning and not only remembering, or if time difference is not so important whereas the content of the posttest is a determinant factor.

Finally, a last limitation refers to the population who participated in our research. We decided to use university students because we had easy access to the sample and because we saw that many research related to feedback is carried out in secondary education. This limits the generalizability of our findings to younger educational phases, although the materials and content of the texts were associated to the level used in 3rd grade of secondary education.

Future directions

The findings on this work contribute in the study of the use and processing of formative feedback in digital environments to enhance reading comprehension and learning from a text, and also opens up new scope of questions that could be addressed in subsequent studies derived from the main conclusions and limitations of this work. Firstly, it would be essential for the educational community to develop research addressing the different educational stages. I personally think that it is fundamental to ensure that younger children in primary education are exposed to elaborative feedback messages to promote learning outcomes. This will enable them to be more efficient in the use and processing of this EF tools in future educational stages and will shed up light in the current state in this early years.

In second place, as in every instructional tool and aid the student can use in educational settings, ability to use the tool is a determinant factor. In this research, we found that some students did not know what a feedback message was. Therefore, if you don’t know what something is, it is very difficult that you can make a correct use of it and maximize the beneficial outcomes of the tool. Thus, elaborating a feedback tutoring system to help students know what is feedback, how to use the information given and give them the chance to practice will benefit students and exploit the benefits of this instructional tool.

Thirdly, another issue that could be addresses in following studies could be the question format and the type of text format. It would be very enriching for the educational community to evaluate the effectiveness of EF in open ended questions. In our research we only used open ended questions in the transfer test, though it would be very beneficial to use them during the training phase and be able to give the students feedback on the elaborations of their answers. Moreover, using different texts such as dialogues or argumentative texts will enrich the empirical research in educational area. Although the traditional way of presenting a topic to students is using an expository text, each time we use in education a wide range of different means to present a topic to students, and studying what type of text work better with different types of feedback would be very interesting.

In fourth place, it would be very interesting to examine in more depth other variables related to feedback such as the mode of feedback presentation which may be oral, written or using a video among others. This would help us to obtain more practical information for the type of feedback which is more beneficial for classroom settings were students do not use electronic or digital tools.

Finally, it would be very relevant to examine the role of individual differences in the use and efficiency of EF messages. There is not many research on this topic, thus it is a field yet to be exploited. It would be necessary to gather information on the effectiveness of different types of feedback messages depending on the students' previous knowledge, reading fluency, comprehension skills, etc. Additionally, it would be very interesting to investigate the relation between EF and other personal variables such as motivation and engagement.

Practical implications for education

A series of practical implication about the use and effectiveness of formative feedback for the educational community can be derived from the conclusions of the present work. As we have previously outlined, feedback is a powerful instructional tool used in education to favor and consolidate learning. Along the recent history of research in this topic, a wide range of studied have studied feedback in different ways, reaching varied results regarding the content feedback should include in its message and its effectiveness of different types of feedback (e.g., shute 2008). Still, there is a common agreement between researchers which suggest that those messages which contain additions information or an explanation, in other words, elaborative feedback messages, are more effective than those containing only information about the correctness of the students' response or those explicitly stating the correct answer (van del Kleij et al., 2015).

Regarding the practical implications for education, we consider that the findings on our studies suggest that it is fundamental to create conditions that can simplify feedback provision by teachers and augment its use by students. For this to happen, we consider that there are two fundamental issues that should be addressed: orienting the students on the purpose of feedback, and creating a disposition on the students for seeking feedback.

It is essential that students are oriented on the purpose of feedback because they have to be conscious of the usefulness of this instructional tool which will only favor knowledge acquisition and misconceptions correction and will never have a negative impact on learning. Showing student’s the usefulness of educational strategies or tools, such as feedback, has not yet received the attention it deserves (Bjork, Dunlosky, & Kornell, 2013; Ness, 2011). Feedback should not be seen by students as a message given by the teacher, and the teacher should not give feedback with the only purpose of writing a message to the student. Feedback should be considered as a system of promoting learning for active learners.

Consequently, consciousness is necessary to create on the learner a positive disposition that encourages him to seek for feedback during and after the realization of instructional tasks. How can we create this positive disposition towards seeking feedback? By designing instructional tasks that make students understand and feel the advantages of feedback and the outcomes on performance. Teaching these strategies should facilitate metacognitive knowledge concerning when and how to use it (Paris, Cross y Lipson, 1984; Pressley, Snyder y Cariglia-Bull, 1987).

Valuable information for the educational community regarding feedback has been determined in these set of studies. The fundamental conclusions about feedback that should be considered for practical implications include results on feedback effectiveness and use. According to previous research, as we outlined before, EF messages has been shown to me more effective than KR and KCR feedback in instructional settings, even though there is not a consensus on the information EF should include to maximize its effectiveness. Besides, our contribution suggests that the EF message the student receives should be coherent with the mental representation he has created to give response to the task. This means that the same feedback message cannot be given independently of the students' success or failure when answering a question. Different issues should be addressed in each case. When a student gives a correct answer, the feedback message should reinforce the idea implicit in the answer by the use of techniques such as paraphrases. Accordingly, when a student gives a wrong answer the feedback should be addressed directly to the wrong information in the answer. This means that each wrong answer, or each of the wrong statements in a MC question, should have its specific feedback message which is coherent with the student mental representation and explicitly states why the idea in the answer is incorrect by giving an explanation. This will help students to process the feedback message in less time and understand it better, reducing the students need to look for additional instructional aids such as searching on the text, and enhancing knowledge acquisition and conceptual change in future tasks.

A final educational implication refers to the beneficial effect of the findings in the present dissertation for the design and development of electronic books to foster learning or online learning websites. There are many digital books which its only difference with the paper version of the book is the medium of use. We consider that digital books should be much more interactive and should facilitate the student with additional information that paper books cannot facilitate, such as feedback dependent on the students' answers. Additionally, this tools should base their advances on empirical research and scientific findings such as those exposed in the present section to maximize learning outcomes.

"If I had to reduce all of educational psychology to just one principle, I would say this: the most important single factor influencing learning is what the learner already knows. Ascertain this and teach him (or her) accordingly"

Ausubel (1968, p. vi)

REFERENCES

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., & Roediger, H. (2008). *The Influence of Retrieval Practice on Memory Precision View project Collective Memory View project*. doi: 10.1002/acp.1391

Alvermann, D. E., & Hague, S. A. (1989). Comprehension of Counterintuitive Science Text: Effects of Prior Knowledge and Text Structure. *Journal of Educational Research, 82*(4), 197–202. doi: 10.1080/00220671.1989.10885893

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science*. doi: 10.1016/0364-0213(89)90021-9

Anderson, R. C. (1978). Schema-Directed Processes in Language Comprehension. In *Cognitive Psychology and Instruction*. doi: 10.1007/978-1-4684-2535-2_8

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology*. doi: 10.1037/h0030766

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1972). Conditions under which feedback facilitates learning from programmed lessons. *Journal of Educational Psychology*. doi: 10.1037/h0032653

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In *Handbook of Reading Research*. doi: 10.1017/cbo9781139524513.007

Andre, T., & Thieman, A. (1988). Level of adjunct question, type of feedback, and learning concepts by reading. *Contemporary Educational Psychology*. doi: 10.1016/0361-476X(88)90028-8

Ariasi, N., & Mason, L. (2011). Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional Science*, *39*(5), 581–601. doi: 10.1007/s11251-010-9142-5

Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education*, *16*(3), 363–383. doi: 10.1007/BF03173188

Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, *99*(3), 626–639. doi: 10.1037/0022-0663.99.3.626

Bangert-Drowns, R. L., Kulik, C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213–238. doi: 10.3102/00346543061002213

Baumann, J. F., & Bergeron, B. S. (1993). Story map instruction using children's literature: Effects on first graders' comprehension of central narrative elements. *Journal of Literacy Research*. doi: 10.1080/10862969309547828

Beker, K., Kim, J., Van Boekel, M., van den Broek, P., & Kendeou, P. (2019). Refutation texts enhance spontaneous transfer of knowledge. *Contemporary Educational Psychology*, *56*, 67–78. doi: 10.1016/j.cedpsych.2018.11.004

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417-444. doi:10.1146/annurev-psych-113011-143823

Block, C. C. (1993). Strategy Instruction in a Literature-Based Reading Program. *The Elementary School Journal*, *94*(2), 139–151. doi: 10.1086/461756

Braasch, J. L., Goldman, S. R., & Wiley, J. (2013). The influences of text and reader characteristics on learning from refutations in science texts. *Journal of Educational Psychology*, *105*(3), 561–578. doi: 10.1037/a0032627

Britt, M., Rouet, J., & Durik, A. (2018). Literacy beyond text comprehension: A theory of purposeful reading. In J. P. M. & G. S. M. T. McCrudden (Ed.), *Text relevance and learning from text* (pp. 19–52). Greenwich, CT: Information Age Publishing. Retrieved from https://content.taylorfrancis.com/books/download?dac=C2015-0-58380-5&isbn=9781317404453&format=googlePreviewPdf. doi: 10.4324/9781315682860-3

Broughton, S. H., Sinatra, G. M., & Reynolds, R. E. (2010). The nature of the refutation text effect: An investigation of attention allocation. *Journal of Educational Research*, *103*(6), 407–423. doi: 10.1080/00220670903383101

Butler., A., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*(2), 290–298. doi:10.1037/a0031026

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*. doi: 10.1037/0278-7393.34.4.918

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. doi:10.3102/00346543065003245

Butterfield, B., & Metcalfe, J. (2001). Errors Committed with High Confidence Are Hypercorrected. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(6), 1491–1494. doi: 10.1037/0278-7393.27.6.1491

Calin-Jageman, R. J., & Ratner, H. H. (2005). The role of encoding in the self-explanation effect. *Cognition and Instruction*, *23*(4), 523–543. doi: 10.1207/s1532690xci2304_4

Caravita, S., & Halldén, O. (1994). Re-framing the problem of conceptual change. *Learning and Instruction*, *4*(1), 89–111. doi: 10.1016/0959-4752(94)90020-5

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press. Retrieved from https://global.oup.com/academic/product/the-origin-of-concepts-9780199838806?cc=es&lang=en&. doi: 10.1093/acprof:oso/9780195367638.001. 0001

Carey, S. (2000). Science Education as Conceptual Change. *Journal of Applied Developmental Psychology*, *21*(1), 13–19. doi: 10.1016/S0193-3973(99)00046-5

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. doi: 10.1016/j.lindif.2010.11.007

Cerdán, R., Pérez, A., Vidal-Abarca, E., & Rouet, J. F. (2019). To answer questions from text, one has to understand what the question is asking: differential effects of question aids as a function of comprehension skill. *Reading and Writing*. doi: 10.1007/s11145-019-09943-w

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of educational psychology, 100*(1), 209. doi: 10.1037/0022-0663.100.1.209

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*(1), 13–27. doi: 10.1016/j.learninstruc.2007.12.003

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences*. doi: 10.1207/s15327809jls1402_1

Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science*. doi: 10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, *4*(1), 27–43. doi: 10.1016/0959-4752(94)90017-5

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching*. doi:

10.1002/tea.20171

Clariana, R B. (1990). A comparison of answer until correct feedback and knowledge of correct response feedback under two conditions of contextualization. *Journal of Computer-Based Instruction, 17*(4), 125–129. Retrieved from https://psycnet.apa.org/record/1991-19725-001

Clariana, R B. (1993). A review of multiple-try feedback in traditional and computer-based instruction. *Journal of Computer-Based Instruction*. Retrieved from https://eric.ed.gov/?id=EJ476365

Clariana, Roy B., & Koul, R. (2005). Multiple-Try Feedback and Higher-Order Learning Outcomes. *International Journal of Instructional Media*. Retrieved from https://search.proquest.com/openview/880c9f9927911433457cbd84492bcf59/1?pq-origsite=gscholar&cbl=30932

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30*(10), 1241–1257. doi: 10.1002/tea.3660301007

Cohen, V. B. (1985). A Reexamination of Feedback in Computer-Based Instruction: Implications for Instructional Design. *Educational Technology, 25*(1), 33-37. Retrieved from http://www.jstor.org/stable/44424353

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American educator, 15*(3), 6-11. Retrieved from https://www.aft.org/ae/winter1991/collins_brown_holum

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/365024.365111

Dempsey, J. V., & Driscoll, M. P. (1996). Error and feedback: Relation between content analysis and confidence of response. *Psychological reports, 78*(3), 1079-1089. doi: 10.2466/pr0.1996.78.3c.1079

Diakidoy, I. A. N., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: A comparison of the effectiveness of two instructional approaches. *Learning &*

Instruction, 11(1), 1–20. doi: 10.1016/s0959-4752(00)00011-6

Diakidoy, I. A. N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology, 28*(3), 335–356. doi: 10.1016/s0361-476x(02)00039-5

Diakidoy, I. A. N., Mouskounti, T., Fella, A., & Ioannides, C. (2016). Comprehension processes and outcomes with refutation and expository texts and their contribution to learning. *Learning and Instruction*, *41*, 60–69. doi: 10.1016/j.learninstruc.2015.10.002

Dictionary, O. E. (1933). Oxford: Oxford University Press, and OED online.

Dilber, R. (2011). Students' understandings of air pressure. In A. Demirbas (Ed.), *ENERGY EDUCATION SCIENCE AND TECHNOLOGY, PART B* (pp. 751–760). SILASCIENCE, UniversiteMah. Retrieved from https://www.researchgate.net/profile/Refik_Dilber/publication/290316419_Students'_understandings_of_air_pressure/links/58e8933e458515e30dcb0492/Students-understandings-of-air-pressure.pdf

Dole, J. A., Brown, K. J., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly*, *31*(1), 62–88. doi: 10.1598/rrq.31.1.4

Dole, J., & Smith, E. (1989). Prior knowledge and learning from science text: An instructional study. In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction* (pp. 345–352). Chicago, IL: NRC. Retrieved from https://eric.ed.gov/?id=ED313664

Dunbar, K. N., Fugelsang, J. A., & Stein, C. (2007). Do naïve theories ever go away? Using brain and behavior to understand changes in concepts. In M. C. Lovett & P. Shah (Eds.), *Carnegie Mellon symposia on cognition. Thinking with data.* (pp. 193–205). Mahwah, NJ: Lawrence Erlbaum Associates Publishers. Retrieved from https://psycnet.apa.org/record/2007-16448-008

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *Psychological Record*, *52*(2), 187–201. doi: 10.1007/BF03395423

Farr, R., Pritchard, R. y 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:10.1111/j.1745-3984.1990.tb00744.x

Fazio, L., & Marsh, E. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, *16*(1), 88–92. doi: 10.3758/PBR.16.1.88

Fenesi, B., Sana, F., & Kim, J. A. (2014). Evaluating the Effectiveness of Combining the Use of Corrective Feedback and High-Level Practice Questions. *Teaching of Psychology*, *41*(2), 135–143. doi: 10.1177/0098628314530344

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. doi: 10.1016/j.cedpsych.2017.10.002

Fiorella, L., & Mayer, R. (2016). Eight Strategies to Encourage Generative Learning | Student Success. Retrieved from http://s3.amazonaws.com/academia.edu.documents/40251120/Fiorella___Mayer_2015_G

Frède, V. (2008). Teaching astronomy for pre-service elementary teachers: A comparison of methods. *Advances in Space Research*, *42*(11), 1819–1830. doi: 10.1016/j.asr.2007.12.001

Fox, J.-P., Klein Entink, R. y Timmers, C. (2014). The joint multivariate modeling of multiple mixed response sources: Relating student performances with feedback behavior. *Multivariate Behavioral Research, 49*(1), 54-66. doi:10.1080/00273171.2013.843441

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*(1), 47–61. doi: 10.1016/J.LEARNINSTRUC.2011.06.002

Gerrig, R. J., & O’Brien, E. J. (2005). The Scope of Memory-Based Processing.

Discourse Processes, *39*(2–3), 225–242. doi: 10.1080/0163853x.2005.9651681

Gibbs, G., & Simpson, C. (2005). Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education.*, *1*, 3–31. doi:10.1007/978-3-8348-9837-1

Gil, L., Martinez, T., & Vidal-Abarca, E. (2015). Online assessment of strategic reading literacy skills. *Computers and Education*, *82*, 50–59. doi: 10.1016/j.compedu.2014.10.026

Gilman, David A. (1969). Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology*. doi: 10.1037/h0028501

Gilman, David Alan. (1969). The effect of feedback on learners' certainty of response and attitude toward instruction in a computer—assisted instruction program for teaching science concepts. *Journal of Research in Science Teaching*. doi: 10.1002/tea.3660060212

Golke, S., Dörfler, T., & Artelt, C. (2015). The impact of elaborated feedback on text comprehension within a computer-based assessment. *Learning and Instruction*. doi: 10.1016/j.learninstruc.2015.05.009

Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback Specificity, Exploration, and Learning. *Journal of Applied Psychology*. doi: 10.1037/0021-9010.89.2.248

Graesser, A., Léon, J., & Otero, J. (2002). Introduction to the psychology of scientific text. In *The psychology of science text comprehension*. Mahwah: New Jersey: LEA. Retrieved from https://dialnet.unirioja.es/servlet/libro?codigo=510673

Graesser, A C. (2007). An introduction to strategic reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (p. 3–26). Lawrence Erlbaum Associates Publishers. Retrieved from https://psycnet.apa.org/record/2007-15600-001

Graesser, Arthur C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371–395. doi: 10.1037/0033-295x.101.3.371

Guzzetti, B., Snyder, T., & Glass, G. (1992). Promoting Conceptual Change in Science: Can Texts be Used Effectively? *Journal of Reading, 35*(8), 642-649. Retrieved from http://www.jstor.org/stable/40032156

Guzzetti, B. J., Snyder, T. E., Glass, G. V, & Gamas, W. S. (1993). Promoting Conceptual Change in Science: A Comparative Meta-Analysis of Instructional Interventions from Reading Education and Science Education. *Reading Research Quarterly, 28*(2), 116. doi: 10.2307/747886

Hamaker, C. (1986). The Effects of Adjunct Questions on Prose Learning. *Review of Educational Research*. doi: 10.3102/00346543056002212

Hamilton, R. J. (1985). A Framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research*. doi: 10.3102/00346543055001047

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. In *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. doi: 10.4324/9780203887332

Hattie, J., & Gan, M. (2011). Instruction Based on Feedback. In *Handbook of Research on Learning and Instruction*. doi: 10.4324/9780203839089.ch13

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

Higgs, K., Magliano, J. P., Vidal-Abarca, E., Martínez, T., & Mcnamara, D. S. (2016). *Discourse Processes Bridging Skill and Task-Oriented Reading*. doi: 10.1080/0163853X.2015.1100572

Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*. doi: 10.1016/j.jml.2013.03.002

Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattray, C. (2000). Hypothesis testing in science: group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction*, *10*(4), 361–391. doi: 10.1016/S0959-4752(00)00004-9

Hynd, C. R., Alvermann, D., & Qian, G. (1997). Preservice elementary school teachers' conceptual change about projectile motion: Refutation text, demonstration, affective factors, and relevance. *Science Education*, *81*(1), 1–27. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1098-237X(199701)81:1%3C1::AID-SCE1%3E3.0.CO;2-M

Hynd, C R, & Guzzetti, B. (1998). What do we mean by knowledge and learning. *Learning from text across conceptual domains, 15-44.* Retrieved from https://books.google.es/books?hl=es&lr=&id=QgJeAgAAQBAJ&oi=fnd&pg=PA15&dq=Hynd,+C+R,+%26+Guzzetti,+B.+(1998).+Learning+from+text+across+conceptual+domains.+(C+R+Hynd,+Ed.)&ots=h_xnjPYp_i&sig=aqgPQizy1cgZPh7IaY5GbI-cCAc#v=onepage&q&f=false

Hynd, C. R. (2001). Refutational texts and the change process. *International Journal of Educational Research*, *35*(7–8), 699–714. doi: 10.1016/S0883-0355(02)00010-1

Hynd, C. R., McWhorter, J. Y., Phares, V. L., & Suttles, C. W. (1994). The role of instructional variables in conceptual change in high school physics topics. *Journal of Research in Science Teaching*, *31*(9), 933–946. doi: 10.1002/tea.3660310908

Jaehnig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *Psychological Record, 57*(2), 219–232. doi: 10.1007/BF03395573

Kaestle, C. F., Campbell, A., Finn, J. D., Johnson, S. T., & Mikulecky, L. J. (2001). Adult Literacy and Four Studies Based on the National Adult Literacy Survey. *Educational Research*. Retrieved from https://nces.ed.gov/pubs2001/2001534.pdf

Kealy, W., & Ritzhaupt, A. (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

Keil, F. C. (2011). Science starts early. *Science*, Vol. 331, pp. 1022–1023. American Association for the Advancement of Science. doi: 10.1126/science.1195221

Kendeou, P., & O'Brien, E. (2014). The Knowledge Revision Components (KReC) framework: Processes and mechanisms. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 353–377). MIT Press. Retrieved from https://psycnet.apa.org/record/2014-41945-016

Kendeou, P., Braasch, J. L. G., & Bråten, I. (2016). Optimizing Conditions for Learning: Situating Refutations in Epistemic Cognition. *Journal of Experimental Education*, *84*(2), 245–263. doi: 10.1080/00220973.2015.1027806

Kendeou, P., Butterfuss, R., Kim, J., & Van Boekel, M. (2019). Knowledge revision through the lenses of the three-pronged approach. *Memory and Cognition*, *47*(1), 33–46. doi: 10.3758/s13421-018-0848-y

Kendeou, P., Butterfuss, R., Van Boekel, M., & O'Brien, E. J. (2017, March 1). Integrating Relational Reasoning and Knowledge Revision During Reading. *Educational Psychology Review*, Vol. 29, pp. 27–39. Springer New York LLC. doi: 10.1007/s10648-016-9381-3

Kendeou, P., & O'Brien, E. J. (2014). The knowledge revision components (KReC) framework: Processes and mechanisms. In D. N. . Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (p. 353-378). Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9737.003.0022

Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(3), 854–865. doi: 10.1037/a0029468

Kendeou, P., & Van Den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory and Cognition*, *35*(7), 1567–1577. doi: 10.3758/BF03193491

Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes*, *51*(5–6), 374–397. doi: 10.1080/0163853x.2014.913961

Kintsch, W, & Franzke, M. (1995). *The role of background knowledge in the recall of a news story*. Retrieved from https://psycnet.apa.org/record/1995-97200-017

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review*. doi: 10.1037/0033-295X.95.2.163

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

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*. doi: 10.1037/0033-295X.85.5.363

Kintsch, W., & van Dijk, T. A. (1978). Toward a Model of Text Comprehension and Production. In *Psychological Review 85*(5), 363–394. doi:10.1037/0033-295x.85.5.363

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*. doi: 10.1037/0033-2909.119.2.254

Kulhavy, R. W. (1977). Feedback in Written Instruction. *Review of Educational Research*. doi: 10.3102/00346543047002211

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

Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology*. doi: 10.1016/0361-476X(85)90025-6

Lefevre, D., & Cox, B. (2016). Feedback in technology-based instruction: Learner preferences. *British Journal of Educational Technology*, *47*(2), 248–256. doi: 10.1111/bjet.12227

Legare, C., & Gelman, S. (2008). Bewitchment, Biology, or Both: The Co-Existence of

Natural and Supernatural Explanatory Frameworks Across Development. *Cognitive Science, 32*(4), 607–642. doi: 10.1080/03640210802066766

Legare, C. H., Gelman, S. A., Wellman, H. M., & Kushnir, T. (2008). The Function of Causal Explanatory Reasoning in Children. *Poster Presented at the Annual Meeting of the Cognitive Science Society*, 1006–1010. Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2008/pdfs/p1105.pdf

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, *11*(4–5), 357–380. doi: 10.1016/S0959-4752(00)00037-2

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*. doi: 10.1111/jcal.12134

Llorens, A. C., Cerdán, R. y 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

Llorens, A. C., Vidal-Abarca, E., Cerdán, R., & Ávila, V. (2015). Does formative feedback on search behaviour help students in answering comprehension questions from an available text? / ¿Ayuda la retroalimentación formativa sobre el comportamiento de búsqueda a contestar preguntas de comprensión en lecturas con texto d. *Infancia y Aprendizaje*. doi: 10.1080/02103702.2015.1076269

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

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 and Education*. doi: 10.1016/j.compedu.2015.12.002

Maña, 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*(4), 553–565. doi: 10.1174/021037009789610412

Mandernach, J. (2005). Relative Effectiveness of Computer-Based and Human Feedback for Enhancing Student Learning. *The Journal of Educators Online*. doi: 10.9743/jeo.2005.1.5

Mañez, I. (2019). *Processing and effectiveness of formative feedback to increase comprehension and learning of conceptual knowledge in digital environments*. Universitat de València. Retrieved from https://roderic.uv.es/handle/10550/71034

Máñez, I., Vidal-Abarca, E., Kendeou, P., & Martínez, T. (2019). How do students process complex formative feedback in question-answering tasks? A think-aloud study. *Metacognition and Learning, 14*(1), 65-87. doi: 10.1007/s11409-019-09192-w

Máñez, I., Vidal-Abarca, E., & Martínez, T. (2019). Does computer-based elaborated feedback influence the students' question-answering process? *Electronic Journal of Research in Educational Psychology*. doi: 10.25115/ejrep.v17i47.2156

Mannes, S. M., & Kintsch, W. (1987). Knowledge Organization and Text Organization. *Cognition and Instruction*, *4*(2), 91–115. doi: 10.1207/s1532690xci0402_2

Maria, K. (2000). Conceptual change instruction: A social constructivist perspective. *Reading and Writing Quarterly*. doi: 10.1080/105735600278042

Martínez, T., Vidal-Abarca, E., Gil, L., & Gilabert, R. (2009). On-line assessment of comprehension processes. *Spanish Journal of Psychology, 12*(1), 308–319. doi: 10.1017/S1138741600001700

Mason, L., & Gava, M. (2007). Effects of epistemological beliefs and learning text structure on conceptual change. In S. Vosniadou, A. Baltas, & X. Vamvakoussi (Eds.), *Advances in learning and instruction series. Reframing the conceptual change approach in learning and instruction* (pp. 165–196). Elsevier Science. Retrieved from https://psycnet.apa.org/record/2007-04764-013

Mason, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: The interplay of text, epistemological beliefs, and topic interest. *Journal of Educational Psychology, 100*(2), 291–309. doi: 10.1037/0022-0663.100.2.291

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the Intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*. doi: 10.1207/s15326985ep4004_7

Mccrudden, M., & Schraw, G. (2007). *Relevance and Goal-Focusing in Text Processing Handbook of Multiple Source Use View project*. doi: 10.1007/s10648-006-9010-7

McCrudden, M. T., & Kendeou, P. (2014). Exploring the link between cognitive processes and learning from refutational text. *Journal of Research in Reading, 37*(S1), S116–S140. doi: 10.1111/j.1467-9817.2011.01527.x

McKeown, M. G., Beck, I. L., Sinatra, G. M., & Loxterman, J. A. (1992). The Contribution of Prior Knowledge and Coherent Text to Comprehension. *Reading Research Quarterly, 27*(1), 78–93. doi: 10.2307/747834

McNamara, D. S., & Magliano, J. (2009). Towards a comprehensive model of comprehension. In *The Psychology of Learning and Motivation*: *Advances in research and theory (vol. 51*, 297–384). San Diego: Elsevier. doi: 10.1016/S0079-7421(09)51009-2

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(2), 288–294. doi: 10.1037/0278-7393.12.2.288

Mikkilä-Erdmann, M. (2001). Improving conceptual change concerning photosynthesis through text design. *Learning and Instruction*, *11*(3), 241–257. doi: 10.1016/s0959-4752(00)00041-4

Mikkilä-Erdmann, M., Penttinen, M., Anto, E., & Olkinuora, E. (2008). Constructing mental models during learning from science text. In *Understanding Models for Learning and Instruction* (pp. 63–79). Springer US. doi: 10.1007/978-0-387-76898-4_4

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*(1–2), 99–113. doi: 10.1023/b:truc.0000021811.66966.1d

Mory, E. H. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Technology Research and Development, 40*(3), 5–20.

doi:10.1007/bf02296839

Mory, E. H. (2004). Feedback research revisited. In Jonassen, D. (eds.), *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Erlbaum Associates. Retrieved from https://members.aect.org/edtech/29.pdf

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merril, J. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125 144). New York, NY: Erlbaum

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*. doi: 10.1344/der.2013.23.7-26

Narciss, S, & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. Instructional design for multimedia learning, (181-195). Retrieved from https://studierplatz2000.tu-dresden.de/toolkit/presentations/CD/Literatur/Publikationen/ID_nahu.pdf

Narciss, S., Sosnovsky, S., & Andres, E. (2014). Adapting tutoring feedback strategies to motivation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/978-3-319-11200-8_22

Nelson, T. O. (1988) Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory*, New York: John Wiley & Sons, 190– 196. Retrieved from https://www.semanticscholar.org/paper/Predictive-accuracy-of-the-feeling-of-knowing-tasks-Nelson/1d921a91b632d1f29b85598134a5b372ea73df69

Nelson, T. O. (1988). Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practice aspects of memory (Vol. 1*, pp. 190–196). New York: Wiley.

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

Noonan, J. V. (1984). Feedback Procedures in Computer-Assisted Instruction, *Journal of Research on Computing in Education,* vol 20, 10-20, doi: 10.1080/08886504.1986.11008431

Novak, J. D. (1988). Learning science and the science of learning. *Studies in Science Education*, *15*, 77–101. doi: 10.1080/03057268808559949

Nussbaum, E. M., Cordova, J. R., & Rehmat, A. P. (2017). Refutation Texts for Effective Climate Change Education. *Journal of Geoscience Education*, *65*, 23–34. doi: 10.5408/15

O'Brien, E. J., Cook, A. E., & Guéraud, S. (2010). Accessibility of Outdated Information. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(4), 979–991. doi: 10.1037/a0019763

OECD. (2018). PISA for Development Reading Framework. In *PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science* (pp. 25–48). OECD Publishing. doi: 10.1787/9789264305274-4-en

Ohlsson, S. (2009). Resubsumption: A Possible Mechanism for Conceptual Change and Belief Revision. *Educational Psychologist*, *44*(1), 20–40. doi: 10.1080/00461520802616267

Özdemir, G., & Clark, D. B. (2007). An Overview of Conceptual Change Theories. *Eurasia Journal of Mathematics, Science and Technology Education*, *3*(4), 351–361. doi: 10.12973/ejmste/75414

Ozgungor, S., & Guthrie, J. T. (2004). *Interactions Among Elaborative Interrogation, Knowledge, and Interest in the Process of Constructing Knowledge From Text*. doi: 10.1037/0022-0663.96.3.437

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*. doi: 10.1080/07370000701632371

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. *Cognition and Instruction*. doi: 10.1207/s1532690xci0102_1

Palmer, D. H. (2003). Investigating the relationship between refutational text and conceptual change. *Science Education*, *87*(5), 663–684. doi: doi.org/10.1002/sce.1056

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*. doi: 10.1016/0361-476X(83)90018-8

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning Memory and Cognition*. doi: 10.1037/0278-7393.31.1.3

Phye, G. D. (1979). The processing of informative feedback about multiple-choice test performance. *Contemporary Educational Psychology*. doi: 10.1016/0361-476X(79)90057-2

Phye, G. D., & Bender, T. (1989). Feedback complexity and practice: Response pattern analysis in retention and transfer. *Contemporary Educational Psychology*. doi: 10.1016/0361-476X(89)90028-3

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. doi: 10.1002/sce.3730660207

Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: a reaction time study with different levels of interference. *International Journal of Science and Mathematics Education*, *13*(1), 21–43. doi: 10.1007/s10763-014-9520-6

Pressley, M., Snyder, B. L., & Cariglia-Bull, T. (1987). How can good strategy use be taught to children? Evaluation of six alternative approaches. In *Transfer of learning* (pp. 81-120). Academic Press.

Pressley, M., Wharton-McDonald, R., Mistretta-Hampston, J., & Echevarria, M. (1998). Literacy Instruction in 10 Fourth-Grade Classrooms in Upstate New York. *Scientific*

Studies of Reading. doi: 10.1207/s1532799xssr0202_4

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging*. doi: 10.1037/0882-7974.16.1.145

Ramsburg, J. T., & Ohlsson, S. (2016). Category change in the absence of cognitive conflict. *Journal of Educational Psychology*, *108*(1), 98–113. doi: 10.1037/edu0000050

Rapp, David N., & van den Broek, P. (2005). Dynamic Text Comprehension. *Current Directions in Psychological Science*, *14*(5), 276–279. doi: 10.1111/j.0963-7214.2005.00380.x

Ratcliff, R., & McKoon, G. (1988). A Retrieval Theory of Priming in Memory. *Psychological Review*, *95*(3), 385–408. doi: 10.1037/0033-295X.95.3.385

Redfield, D. L., & Rousseau, E. W. (1981). A Meta-Analysis of Experimental Research on Teacher Questioning Behavior. *Review of Educational Research*. doi: 10.3102/00346543051002237

Renkl, A. (2014). Toward an Instructionally Oriented Theory of Example-Based Learning. *Cognitive Science*, *38*(1), 1–37. doi: 10.1111/cogs.12086

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

Roper, W. J. (1977). Feedback in Computer Assisted Instruction. *Programmed Learning and Educational Technology*, *14*(1), 43–49. doi: 10.1080/1355800770140107

Rouet, J.-F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. *Text Relevance and Learning from Text*, (June), 19–52. Retrieved from https://www.researchgate.net/profile/M_Anne_Britt/publication/268011216_Relevance_processes_in_multiple_document_comprehension/links/54eb45e60cf2082851bd7272.pdf

Rouet, J. F. (2006). The skills of document use: From text comprehension to web-based learning. In *The Skills of Document Use: From Text Comprehension to Web-Based*

Learning. doi: 10.4324/9780203820094

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

Rubio, A. (2018). Técnicas de aprendizaje de conocimiento científico a partir de textos. *Investigaciones Sobre Lectura*, *10*, 1–29. doi: 10.37132/isl.v0i10.255

Rupp, A. A., Ferne, T. y Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, *23*(4), 441-474. doi:10.1191/0265532206lt337oa

Sagerman, N., & Mayer, R. E. (1987). Forward Transfer of Different Reading Strategies Evoked by Adjunct Questions in Science Text. *Journal of Educational Psychology*. doi: 10.1037/0022-0663.79.2.189

Schimmel, B. J. (1988). Providing meaningful feedback in courseware. In *Instruction Design for Microcomputing Courseware* (pp. 183–194). Taylor and Francis. doi: 10.4324/9780203062692

Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two Wrongs May Make a Right ... If They Argue Together! *Cognition and Instruction*, *18*(4), 461–494. doi: 10.1207/S1532690XCI1804_2

Schworm, S., & Renkl, A. (2002). Learning by Solved Example Problems: Instructional Explanations Reduce Self-Explanation Activity. *Proceedings of the Annual Meeting of the Cognitive Science Society*. doi: 10.4324/9781315782379-175

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209–2015. doi: 10.1016/j.cognition.2012.04.005

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

Siegler, R. S. (1995). How does change occur: a microgenetic study of number conservation. *Cognitive Psychology*, *28*(3), 225–273. doi: 10.1006/cogp.1995.1006

Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science*, *11*(4), 433–448. doi: 10.1111/j.1467-7687.2008.00689.x

Sinatra, G. M., Broughton, S. H., Diakidoy, I.-A. N., Kendeou, P., & Van Den Broek, P. (2011). *Bridging Reading Comprehension and Conceptual Change in Science Education: The Promise of Refutation Text*. doi: 10.1002/RRQ.005

Stobart, G. (2008). *Testing Times. The Uses and Abuses of Assessment*. London: Routledge. doi: 10.4324/9780203930502

Stock, W. A., Winston, K. S., Behrens, J. T., & Harper·marinick, M. (1989). The effects of performance expectation and question difficulty on text study time, response certitude, and correct responding. In *Bulletin of the Psychonomic Society 27*(6), 567–569. doi:10.3758/bf03334671

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*. doi: 10.1007/s11165-016-9602-2

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285. doi: 10.1126/SCIENCE.1192788

Timms, M., DeVelle, S., & Lay, D. (2016). Towards a model of how learners process feedback: A deeper look at learning. *Australian Journal of Education, 60*(2), 128-145. doi:10.1177/0004944116652912

Timmers, C. F. y Veldkamp, B. P. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education*, *56*(3), 923-930. doi:10.1016/j.compedu.2010.11.007

Tippett, C. D. (2010). Refutation Text in science education: a review of two decades of research. *International Journal of Science & Mathematics*, *8*(6), 951–970. doi:10.1007/s10763-010-9203-x

Trumper, R. (2001). A cross-age study of senior high school students' conceptions of basic astronomy concepts. *Research in Science and Technological Education*, *19*, 97–109. doi: 10.1080/02635140120046259

Tzeng, Y., Van Den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the landscape model: Modeling inferential processes and memory representations of text comprehension. *Behavior Research Methods*, *37*(2), 277–286. Springer New York LLC. doi: 10.3758/BF03192695

Van Boekel, M., Lassonde, K. A., O'Brien, E. J., & Kendeou, P. (2017). Source credibility and the processing of refutation texts. *Memory and Cognition*, *45*(1), 168–181. doi: 10.3758/s13421-016-0649-0

Van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: the role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, *22*(3), 335–351. doi: 10.1002/acp.1418

Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. *The construction of mental representations during reading*, 71-98. Retrieved from https://www.researchgate.net/profile/Michael_Young7/publication/308309047_Van_den_Broek_Young_Tzeng_Linderholm_1999_The_construction_of_mental_representations_during_reading/links/57e023ac08aece48e9e1ee60.pdf

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: A Meta-Analysis. *Review of Educational Research*. doi: 10.3102/0034654314564881

Van Der Kleij, F. M., Timmers, C. F., & Eggen, T. J. H. M. (2011). The effectiveness of methods for providing written feedback through a computer-based assessment for learning: A systematic review. *Cadmo*. doi: 10.3280/CAD2011-001004

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.473.5491&rep=rep1&type=pdf

Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual Differences for Self-Regulating Task-Oriented Reading Activities. *Journal of Educational Psychology*, *102*(4), 817–826. doi: 10.1037/a0020062

Vidal-Abarca, E., Martínez, T., Gil, L., García, A., & Máñez, I. (2019). Learning tasks in electronic environments: Advances towards interactive etextbooks. In *Deep Comprehension: Multi-Disciplinary Approaches to Understanding, Enhancing, and Measuring Comprehension*. doi: 10.4324/9781315109503

Vosniadou, Stella. (2013). *International Handbook of Research on Conceptual Change*. Routledge. doi: 10.4324/9780203154472

Vosniadou, Stella, & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585. doi: 10.1016/0010-0285(92)90018-W

Vygotsky, L. S. (1978). Mind and Society: The Development of Higher Psychological Processes. In *Harvard University Press*. Retrieved from https://books.google.es/books?hl=es&lr=&id=Irq913lEZ1QC&oi=fnd&pg=PR13&dq=ygotsky,+L.+S.+(1978).+Mind+and+Society:+The+Development+of+Higher+Psychological+Processes.+In+Harvard+University+Press.+doi:&ots=HbAsF5xlse&sig=D-_e-3Wj1w924ERaRGSy2pAKmhc#v=onepage&q&f=false

Whyte, M. M., Karolick, D. M., Nielsen, M. C., Elder, G. D., & Hawley, W. T. (1995). Cognitive Styles and Feedback in Computer-Assisted Instruction. *Journal of Educational Computing Research*. doi: 10.2190/m2av-gehe-cm9g-j9p7

Wittrock, M. C. (2010). Learning as a generative process. *Educational Psychologist, 45*(1), 40–45. doi: 10.1080/00461520903433554

Yazbec, A., Borovsky, A., & Kaschak, M. P. (2019). Examining the impact of text style and epistemic beliefs on conceptual change. *PLOS ONE*, *14*(9), e0220766. doi: 10.1371/journal.pone.0220766

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of Situation Model Construction in Narrative Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386–397. doi: 10.1037/0278-7393.21.2.386

APPENDIX

APPENDIX A:PREREST, TEXT AND QUESTIONS IN STUDIES 1 AND 2

PRETEST. PREVIOUS KNOWLEDGE QUESTIONNAIRE

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

LEARNING PHASE: TEXT

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.

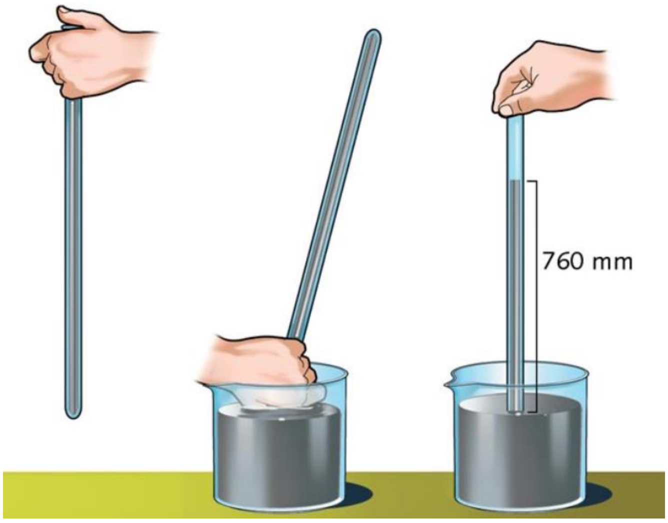
Todos los materiales y los seres vivos que poblamos este planeta estamos sometidos a la presión atmosférica. Si no somos conscientes de este peso del aire es porque ya estamos adaptados porque se ejerce por igual en todas direcciones y nuestros líquidos internos están a la misma presión.

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.

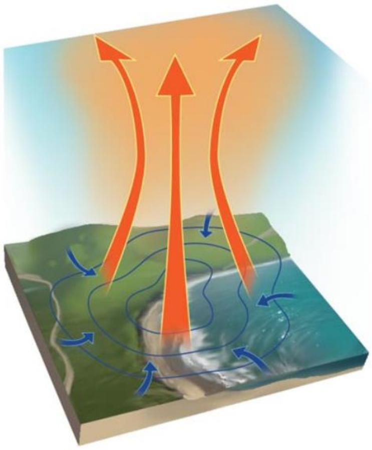
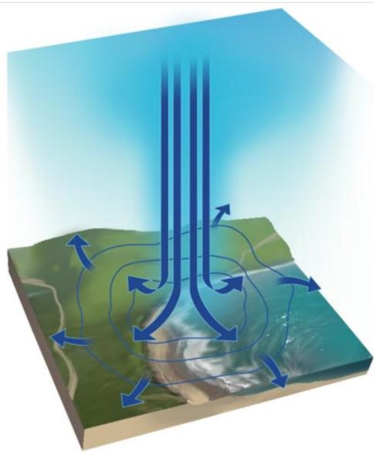


Imagen 3 Formación de una zona de alta presión o anticiclón. El aire frío desciende y comprime el aire inferior que se dispersa al llegar a la superficie.

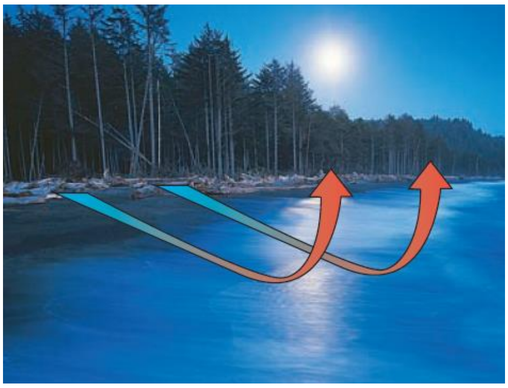


Si combinamos estos dos fenómenos podemos entender cómo funciona la dinámica de la atmósfera y cómo se produce el viento. El viento es el movimiento de grandes masas de aire a través de la troposfera (la capa inferior de la atmósfera). La existencia de zonas de baja y alta presión ocasiona movimientos 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.



LEARNING PHASE: MC QUESTIONS

TEXT-BASE QUESTIONS

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 atmosfera.

    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. ¿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.

4. ¿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.

5. ¿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.

6. ¿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.

SITUATION MODEL QUESTIONS

1. 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.

2. 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.

3. 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.

4. 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?

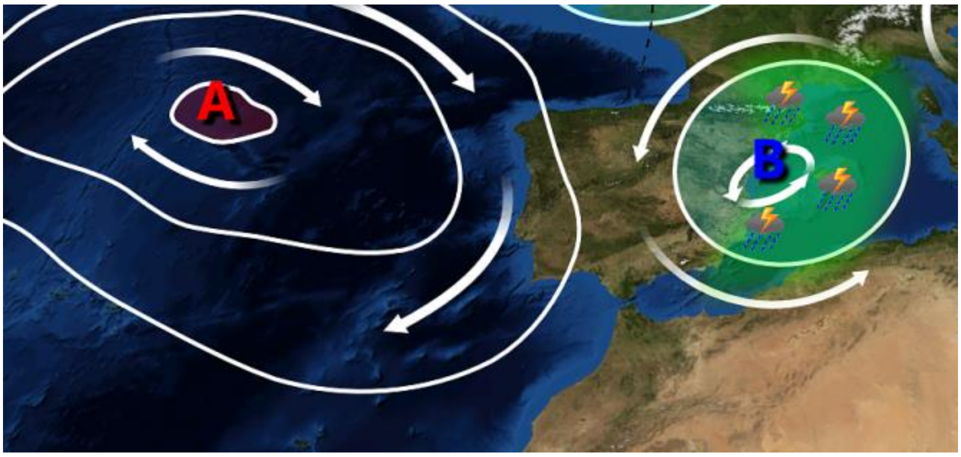
a) Es igual a la del aire exterior porque el globo está abierto.

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.

5. ¿Qué ocurriría en España si hubiera una borrasca en el Mediterráneo y un anticiclón en Portugal? (mapa)



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.

6. 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.

POST-TEST: MC AND OE QUESTIONS

INSTRUCCIONES

Esta es una prueba de aprendizaje sobre la unidad de Ciencias que estudiaste el otro día. Contiene algunas preguntas parecidas a las que respondiste y otras diferentes. Ten en cuenta lo siguiente:

1) Responde únicamente a lo que cada pregunta te pide.
2) Contesta **todas** las preguntas con la información que recuerdes.
3) En las preguntas de test, redondea la alternativa correcta. En las preguntas abiertas, responde de forma clara y breve.

1. ¿Qué da lugar a la presión atmosférica?

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

3. Imagina que Torricelli hubiera dispuesto de oro líquido, el cual es bastante más denso que el mercurio, y lo hubiera introducido en el tubo. ¿Qué crees que habría pasado?

a) El oro habría quedado a la misma altura, 760 mm.

b) El oro habría quedado más abajo de los 760 mm.

c) Se habría salido prácticamente todo el oro.

d) No habría salido casi oro quedando cerca de los 1000 mm.

4. 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?

5. ¿La fuerza que ejerce la atmósfera es igual en todos los puntos de la tierra?

a) Sí, porque la composición de las partículas del aire es uniforme en todos los puntos.

b) Sí, porque esta fuerza se ejerce por igual en todas las direcciones.

c) No, porque depende del peso del aire de las capas inferiores.

d) No, porque la densidad del aire depende de la altura de ese punto.

6. ¿Cómo varía la densidad del aire de la atmósfera en lo alto de una montaña?

a) Es menor porque la columna de aire que tenemos por encima es menor.

b) Es mayor porque al estar más altos estamos más cerca de las capas de aire.

c) Es mayor porque el aire tiene menos humedad.

d) Es menor porque la temperatura es más baja.

7. Si quisiéramos subir una gran montaña, como el Everest, veríamos que nos cuesta mucho respirar. ¿Por qué crees que sucede esto?

a) Porque el aire al estar muy frio es muy pesado y cuesta respirarlo.

b) Porque la presión del aire es mayor e impide que podamos respirar bien.

c) Porque hay menos aire, ya que este es menos denso a esa altura.

d) Porque a esa altura el aire es más denso y empieza condensarse.

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

9. ¿Qué hace que el aire caliente ascienda?

a) Que haya una zona de baja presión atmosférica.

b) Que tiende a expandirse para ocupar todo el espacio.

c) Que tiene una densidad menor que el aire frío.

d) Que las partículas de aire se vuelvan más ligeras.

10. 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?

11. ¿Qué pasará con un globo elástico hinchado y cerrado herméticamente si sube la temperatura y se calienta el aire de su interior?

a) Al calentarse aumentará su volumen.

b) Al calentarse disminuirá su volumen.

c) Al calentarse pesará menos y ascenderá.

d) Nada, ya que el número de partículas dentro del globo permanece constante.

12. ¿Cómo se explica el viento?

a) Por los movimientos del aire que tienden a enfriar la temperatura del planeta.

b) Por el desplazamiento del aire desde zonas de alta a las de baja presión.

c) Por el movimiento de rotación de la troposfera alrededor de la tierra.

d) Por el proceso de mezcla de los gases que componen la atmosfera.

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

14. Una borrasca se origina porque el aire que hay en la superficie....

15. En las zonas costeras, por la noche, podemos notar que la temperatura no es tan baja como en las zonas de interior ¿a qué se debe este fenómeno?

a) A que la humedad existente en el ambiente en las zonas costeras produce que la sensación térmica sea más cálida y existan menos variaciones de temperatura.

b) A que la presión atmosférica en las zonas de costa es más estable que en las zonas de interior por lo que existen menos movimientos de masas de aire que las enfríen.

c) A que en el interior hay una mayor altitud por lo que son alcanzadas antes por las masas de aire frio que bajan desde las capas altas de la atmosfera.

d) A que el mar se enfría más despacio que la tierra; así, por la noche el aire caliente sobre el mar asciende dirigiéndose hacia el interior de la costa, mientras que el aire más frio de la costa se dirige por la capa inferior hacia el mar calentándose.

16. En la orilla del mar, durante el verano a mediodía, podemos notar cierta brisa ¿A qué se debe?

APPENDIX B: PRETEST, TEXT AND QUESTIONS IN STUDIES 3 AND 4

PRETEST. PREVIOUS KNOWLEDGE QUESTIONNAIRE

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

LEARNING PHASE: TEXT 1

Lee con atención el siguiente texto intentando comprenderlo lo mejor posible. Luego encontrarás una serie de preguntas. ***NO** podrás consultar el texto* *para responderlas.*

LA PRESIÓN ATMOSFÉRICA

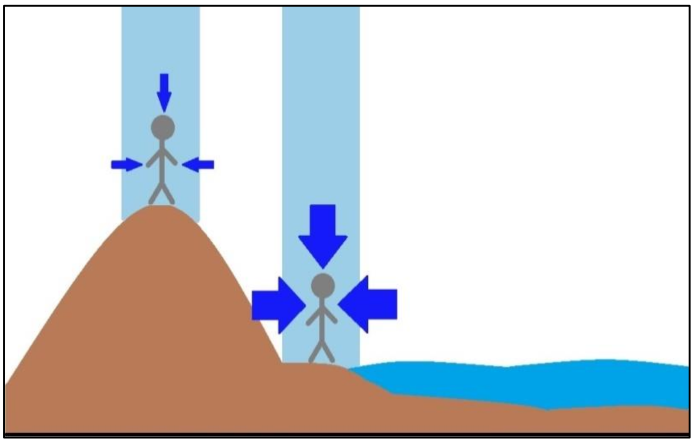
La Tierra está rodeada por una capa de gases que la separa del espacio vacío que constituye, en su mayor parte, el Universo. Esta capa recibe el nombre de atmósfera y está formada por una mezcla de gases que llamamos "aire".

La presión atmosférica

Los gases de la atmósfera, como toda materia, están compuestos 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 (Imagen 1). Todos los materiales y seres vivos estamos sometidos a la presión atmosférica. Si no somos conscientes de este peso del aire es porque ya estamos habituados, porque se ejerce por igual en todas direcciones y porque nuestros líquidos y gases internos están a esa misma presión.

Imagen 1. Concepto de presión atmosférica

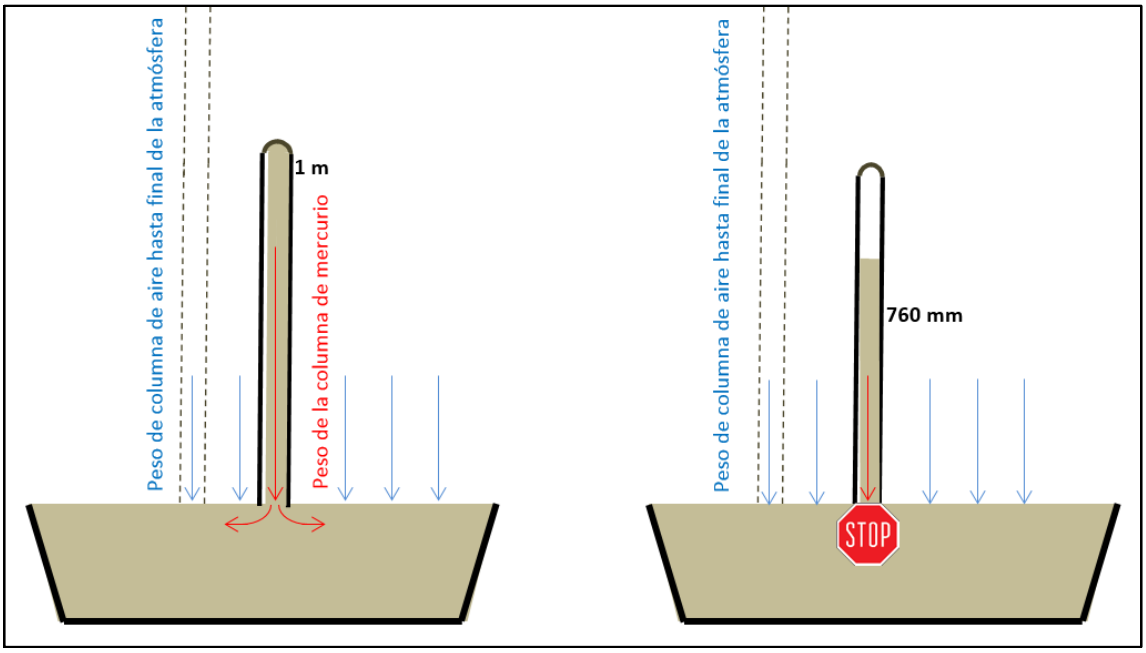


El descubrimiento de la presión atmosférica

La fuerza del peso del aire fue descubierta por el científico italiano Evangelista Torricelli en el año 1643. Torricelli realizó un experimento para demostrar que el aire ejercía presión. De este modo estableció cómo medir la presión atmosférica.

Para su experimento, Torricelli llenó con mercurio un tubo de un metro de longitud cerrado por un extremo y tapó el extremo abierto con el dedo. Posteriormente, lo introdujo invertido en una cubeta llena de mercurio y, finalmente, retiró el dedo con cuidado para que no entrara aire. En ese momento, el mercurio del tubo comenzó a salir hacia la cubeta (Imagen 2a), pero cuando bajó hasta una altura de 760 mm dejó de salir mercurio (Imagen 2b). ¿Por qué comenzó saliendo el mercurio del tubo y luego se paró? Porque en un principio, el peso del mercurio del tubo era mayor que el peso de la columna de aire sobre el mercurio de la cubeta, pero llegó un momento en que el peso del mercurio del tubo era igual al peso de la columna de aire. La conclusión de Torricelli fue clara: el peso de la columna de aire (presión atmosférica) era igual al peso de una columna de 760 mm de mercurio.

Imagen 2. Experimento de Torricelli



a) *El mercurio sale del tubo porque el peso del mercurio es mayor que el peso de la columna de aire.*

b) *El mercurio deja de salir porque el peso de la columna de 760 mm de mercurio es igual que el peso de la columna de aire.*

Torricelli diseñó el barómetro, un aparato 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 (760 mmHg = 1 atm).

La presión atmosférica varía

Las partículas de los gases se mueven libremente ocupando todo el volumen, aunque la distribución de esas partículas no es uniforme, ya que la atmósfera es un espacio muy extenso y las condiciones de cada lugar son diferentes. Así, la densidad de las capas de aire varía según la altura, siendo el aire de las capas inferiores más denso porque soportan el peso del aire de las capas que están por encima. Entonces, en las capas inferiores las partículas están más comprimidas, es decir, 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 tanto, la presión en la cima de una montaña es menor que en una playa. Se ha establecido, por tanto, que la presión normal, la equivalente a 1 atm (1013 hPa), es la que se da al nivel del mar.

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. Al separarse reducen su densidad, lo que hace que disminuya su presión. Lógicamente, lo contrario ocurre cuando los gases se enfrían.

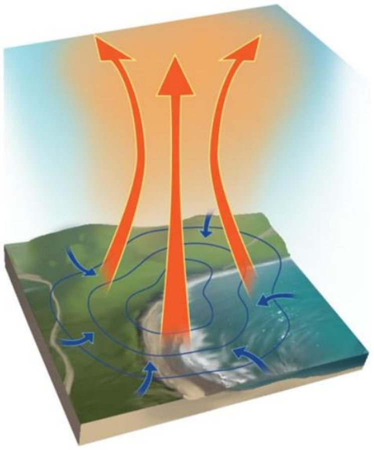
La presión atmosférica y el viento

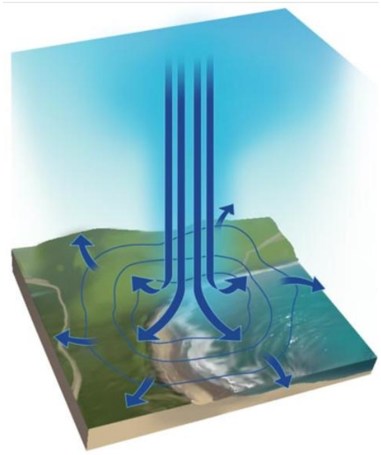
Las diferencias de densidad y presión atmosférica en las distintas partes del planeta son las responsables de los vientos y de otros fenómenos meteorológicos.

Imagen 3. *Formación de una zona de baja presión o borrasca*. Las zonas de baja

Imagen 4. *Formación de una zona de alta presión o anticiclón*. En las zonas de alta presión atmosférica, llamadas ciclón o borrasca, se forman por masas de aire caliente que cuando ascienden dejan tras de sí un área de baja densidad que es rellenada por masas de aire vecinas.

presión, llamadas anticiclón, 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 y su dispersión 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.

LEARNING PHASE: MC QUESTIONS FROM TEXT 1

TEXT BASE QUESTIONS

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

a) Porque a nivel del mar la presión se ve más afectada por el viento.

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

c) Porque la presión aumenta con la altura al hacer más frio.

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

2. ¿Por qué se origina una borrasca?

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

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.

3. ¿Qué sucede con las corrientes de aire una vez que se iguala la presión de dos regiones vecinas?

a) Se forma un anticiclón.

b) Se forma una borrasca.

c) Deja de haber corrientes de aire.

d) Deja de haber borrascas, pero no anticiclones.

4. ¿Qué quería demostrar Torricelli con su experimento?

a) La relación entre presión y temperatura.

b) Que el aire ejercía presión.

c) Cómo usar el barómetro.

d) El valor equivalente a una atmósfera.

SITUATION MODEL QUESTIONS

1. Imagina que hinchamos un globo en la orilla de la playa. Después tomamos un vuelo en helicóptero y subimos a 5000 metros de altura. ¿Qué sucederá con el volumen del globo cuando ascendamos? ¿Por qué?

a) El globo disminuirá su volumen por efecto de la presión del exterior.

b) El globo se aplastará, ya que el aire del interior tiene una presión baja.

c) El globo mantendrá su volumen ya que el número de partículas de aire no varía.

d) El globo se hinchará, ya que disminuirá la presión exterior.

2. Imagina el siguiente experimento: llenamos dos globos elásticos idénticos con la misma cantidad de aire y los cerramos herméticamente. Si subimos 10º la temperatura de uno de ellos (A), pero bajamos 10º la del otro (B), ¿variará el tamaño de ambos?

a) Sí, B tendrá un tamaño superior a A.

b) No, ambos globos mantendrán su tamaño.

c) No necesariamente, depende de la altura a la que realices el experimento.

d) Sí, A tendrá más tamaño que B.

3. Si Torricelli hubiera llenado el tubo con una sustancia líquida más densa, como por ejemplo el oro en estado líquido (Au), ¿a qué altura habría quedado?

a) Habría quedado por debajo de 760 mm.

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

c) Habría quedado por encima de 760 mm.

d) Habría salido todo el oro del tubo.

4. Imagina un globo aerostático abierto por la base, completamente hinchado y ascendiendo al calentar el aire del interior con un quemador. ¿Cómo es la densidad del aire dentro del globo respecto a la del aire en el entorno fuera del globo? ¿Por qué?

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

b) Es mayor que la del aire exterior porque aumenta el tamaño de las partículas.

c) Es menor que la del aire exterior porque el aire está más caliente.

d) Es menor que la del aire exterior porque el volumen del globo disminuye.

LEARNING PHASE: TEXT 2

Lee con atención el texto que tienes a continuación intentando comprenderlo lo mejor posible. Luego encontrarás una serie de preguntas. ***NO*** *podrás consultar el texto para responderlas.*

LA TRANSMISION DEL CALOR

1. Calor, energía interna y temperatura

Aunque calor y temperatura son dos conceptos muy diferentes, frecuentemente se confunden. Mientras la temperatura es la energía promedio de todas las partículas que posee un cuerpo, el calor es la energía térmica que se transfiere de un objeto a otro debido a que están a diferente temperatura. Cuando dos cuerpos con diferente temperatura se ponen en contacto, la energía térmica (calor) se transfiere siempre del cuerpo de mayor temperatura al de menor temperatura hasta que ambos alcanzan la misma temperatura. En ese momento, cesa el tránsito de energía y su temperatura ya no cambia. La energía que va de un cuerpo a otro se mide en calorías, que es la unidad de medida del calor.

Para entender el tránsito de energía hay que saber que todos los cuerpos están formados por partículas en continuo movimiento, por lo que decimos que las partículas tienen energía cinética en función de la velocidad a la que se mueven. Cada partícula tiene una cantidad de energía cinética; la suma de las energías de todas esas partículas es la energía acumulada en un cuerpo, o energía interna. Cuando un cuerpo emite o recibe calor, cambia la energía interna acumulada en su interior. Por ejemplo, cuando echas leche muy caliente (60ºC) de un recipiente a una taza a 20ºC, la leche cede parte de su energía interna a la taza, aumentando la velocidad de las partículas de la taza. Ello ocurre porque las partículas de la leche y la taza chocan entre sí, por lo que las partículas del cuerpo de mayor temperatura ceden energía a las de menor temperatura (más lentas), haciendo que éstas se muevan más rápidamente. Algo similar ocurre en el juego del billar: la bola más rápida cede energía a las bolas lentas al chocar con ellas. Cuando las partículas de ambos cuerpos tienen la misma temperatura ya no hay cesiones apreciables de energía de un cuerpo a otro, lográndose el equilibrio térmico. Esto significa que cuando ponemos dos o más cuerpos en contacto directo durante suficiente tiempo, y no hay influencia de otros cuerpos, todos llegan a la misma temperatura final, sin importar su material y tamaño.

La temperatura de un cuerpo es el promedio (no el total) de la energía cinética que tienen sus partículas. Así, la temperatura nos dice cuán rápido se mueven las partículas de un cuerpo por término medio. Por ejemplo, las partículas de agua a 100ºC se mueven de promedio más rápidamente que cuando está a 0ºC. Sin embargo, para calcular la energía interna total de un cuerpo necesitamos saber no solo la velocidad promedio de sus partículas, sino también cuántas partículas tiene. Por tanto, a igual temperatura y material, un cuerpo grande tiene mayor energía interna que uno pequeño. Asimismo, a igual tamaño y material, mayor temperatura implica mayor energía interna.

2. La transmisión del calor por conducción

Materiales de diferente naturaleza se comportan de distinta forma frente al efecto del calor. Los metales, como por ejemplo el hierro, generalmente transmiten calor con facilidad, es decir, son buenos conductores térmicos. Otros materiales, como el aire, la madera y los materiales porosos son, en general, malos conductores del calor, es decir, son aislantes térmicos. En el caso de los buenos conductores térmicos, el calor que se comunica a una de sus partes se transmite rápidamente de esa parte al resto del cuerpo y al ambiente. Entonces se eleva la temperatura de todas las partes del cuerpo hasta llegar rápidamente a una temperatura interior homogénea y en equilibrio térmico con el ambiente. Además, los conductores también ceden fácilmente calor cuando están en contacto con otros cuerpos a una temperatura menor; es decir, se enfrían fácilmente de modo homogéneo y logran el equilibrio térmico en su interior y con el ambiente.

En un aislante térmico, una parte del cuerpo puede estar a una determinada temperatura durante largo rato mientras que otra parte, a cierta distancia, está a otra temperatura diferente. Asimismo, cuando un aislante térmico está en contacto con otro cuerpo de menor temperatura, el aislante dificulta la transmisión de calor. Esa es la razón por la que los aislantes se emplean para enlentecer la transmisión de calor. Sin embargo, no existe ningún aislante térmico capaz de impedir por completo que se transfiera el calor; de hecho, un aislante sólo reduce la velocidad de transferencia del calor, pero transcurrido el tiempo suficiente, se llegará al equilibrio térmico y ambos cuerpos tendrán la misma energía interna.

A veces se confunde la temperatura de los objetos con la sensación térmica que tenemos al tocarlos. Por ejemplo, si tocamos una superficie de mármol nos parece más fría que si tocamos una cuchara de madera que está encima del mármol. Esto sucede porque el mármol es mejor conductor térmico que la madera: el calor pasa fácilmente de nuestra mano al mármol, robándonos calor en una cantidad que nuestro sistema nervioso percibe. Sin embargo, la madera es mal conductor térmico, con lo cual no somos capaces de percibir la cesión de calor de nuestra mano a la madera, por lo que la notamos cálida. Por tanto, la sensación de frío o calor al tocar un objeto depende no sólo de nuestra temperatura y la del objeto, sino también de la conductividad del calor del material con el que están hechos los objetos.

LEARNING PHASE: MC QUESTIONS FROM TEXT 2

TEXT BASE QUESTIONS

1. ¿Qué medimos al obtener la temperatura de un cuerpo?

a) La energía cinética total de ese cuerpo en un momento dado.

b) El promedio de la energía cinética de sus partículas.

c) El tránsito de energía entre las partículas de ese cuerpo.

d) Las calorías de ese cuerpo en ese momento.

2. ¿Qué es el calor?

a) Un fenómeno físico provocado por el equilibrio térmico de los objetos.

b) La temperatura de un cuerpo medida en calorías.

c) Una propiedad interna de los cuerpos que se manifiesta en forma de energía.

d) La energía que se transfiere entre objetos de diferente temperatura.

3. ¿Cuándo se alcanza el equilibrio térmico entre dos cuerpos?

a) Cuando ambos cuerpos tienen la misma energía interna promedio de agitación.

b) Cuando ambos cuerpos tienen la misma energía interna total de agitación.

c) Cuando las partículas de un cuerpo ceden toda su energía a la del otro.

d) Cuando dejan de producirse choques entre las partículas de ambos cuerpos.

4. ¿Los objetos de igual temperatura y material siempre tienen la misma energía interna? ¿Por qué?

a) Sí, porque la temperatura determina la energía cinética de un cuerpo.

b) Sí, porque el calor que se les transmite desde el ambiente es el mismo.

c) No, porque también depende del tamaño de los objetos.

d) No, porque también depende de la velocidad de movimiento de las partículas.

SITUATION MODEL QUESTIONS

1. Si queremos mantener fría una lata de refresco para ir a la playa, ¿en qué deberemos envolverla para mantenerla más tiempo fría?

a) Papel de aluminio.

b) Una caja de metal.

c) Una bolsa de plástico.

d) Un trapo de lana.

2. Una casa tiene todo el suelo de parqué a excepción del suelo de la cocina, que es de cerámica. Uno de los inquilinos afirma que al ir descalzo está mucho más frío el suelo de la cocina que el del resto de la casa. ¿Su afirmación es correcta? ¿Por qué?

a) Sí, la cerámica es más fría que el parqué.

b) Sí, como los materiales son diferentes estarán a diferente temperatura.

c) No, el suelo está a la misma temperatura, pero la sensación térmica es diferente.

d) No, la cerámica nos parece más fría porque es mala conductora térmica.

3. ¿Por qué se instalan dobles ventanas en una casa cuando se quiere aislar del frío?

a) Porque el frío de fuera no puede traspasar el cristal exterior.

b) Porque el calor de dentro es absorbido por el primer cristal, y lo poco que queda es absorbido por el segundo.

c) Porque el aire entre los dos cristales dificulta que el interior ceda el calor al exterior.

d) Porque el aire entre los dos cristales impide que el frío se transfiera hacia dentro.

4. ¿Podría un cubito de hielo transferir calor a otro objeto?

a) No, porque las partículas en el hielo ya no tienen movimiento.

b) Sí, siempre que tenga un tamaño mucho mayor.

c) Sí, si entra en contacto con otro cuerpo más frío.

d) No, porque está muy frío y solo puede absorber energía.

POST-TEST: MC AND OE QUESTIONS

Esta es una prueba de aprendizaje sobre las unidades de Ciencias que has estudiado.

Ten en cuenta lo siguiente:

i. Responde únicamente a lo que cada pregunta te pide.
ii. Contesta todas las preguntas con la información que recuerdes.
iii. Debes completar la prueba en el tiempo restante de la clase, así que administra bien tu tiempo.

1. ¿Qué da lugar a la presión atmosférica?

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

3.Imagina que Torricelli hubiera dispuesto de oro líquido, el cual es bastante más denso que el mercurio, y lo hubiera introducido en el tubo. ¿Qué crees que habría pasado?

a) El oro habría quedado a la misma altura, 760 mm.

b) El oro habría quedado más abajo de los 760 mm.

c) Se habría salido prácticamente todo el oro.

d) No habría salido casi oro quedando cerca de los 1000 mm.

4. 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?

5. ¿La fuerza que ejerce la atmósfera es igual en todos los puntos de la tierra?

a) Sí, porque la composición de las partículas del aire es uniforme en todos los puntos.

b) Sí, porque esta fuerza se ejerce por igual en todas las direcciones.

c) No, porque depende del peso del aire de las capas inferiores.

d) No, porque la densidad del aire depende de la altura de ese punto.

6. ¿Cómo varía la densidad del aire de la atmósfera en lo alto de una montaña?

a) Es menor porque la columna de aire que tenemos por encima es menor.

b) Es mayor porque al estar más altos estamos más cerca de las capas de aire.

c) Es mayor porque el aire tiene menos humedad.

d) Es menor porque la temperatura es más baja.

7. Si quisiéramos subir una gran montaña, como el Everest, veríamos que nos cuesta mucho respirar. ¿Por qué crees que sucede esto?

a) Porque el aire al estar muy frio es muy pesado y cuesta respirarlo.

b) Porque la presión del aire es mayor e impide que podamos respirar bien.

c) Porque hay menos aire, ya que este es menos denso a esa altura.

d) Porque a esa altura el aire es más denso y empieza condensarse.

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

9. ¿Qué hace que el aire caliente ascienda?

a) Que haya una zona de baja presión atmosférica.

b) Que tiende a expandirse para ocupar todo el espacio.

c) Que tiene una densidad menor que el aire frío.

d) Que las partículas de aire se vuelvan más ligeras.

10. 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?

11. ¿Qué pasará con un globo elástico hinchado y cerrado herméticamente si sube la temperatura y se calienta el aire de su interior?

a) Al calentarse aumentará su volumen.

b) Al calentarse disminuirá su volumen.

c) Al calentarse pesará menos y ascenderá.

d) Nada, ya que el número de partículas dentro del globo permanece constante.

12. ¿Cómo se explica el viento?

a) Por los movimientos del aire que tienden a enfriar la temperatura del planeta.

b) Por el desplazamiento del aire desde zonas de alta a las de baja presión.

c) Por el movimiento de rotación de la troposfera alrededor de la tierra.

d) Por el proceso de mezcla de los gases que componen la atmosfera.

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

14. Una borrasca se origina porque el aire que hay en la superficie....

15. En las zonas costeras, por la noche, podemos notar que la temperatura no es tan baja como en las zonas de interior ¿a qué se debe este fenómeno?

a) A que la humedad existente en el ambiente en las zonas costeras produce que la sensación térmica sea más cálida y existan menos variaciones de temperatura.

b) A que la presión atmosférica en las zonas de costa es más estable que en las zonas de interior por lo que existen menos movimientos de masas de aire que las enfríen.

c) A que en el interior hay una mayor altitud por lo que son alcanzadas antes por las masas de aire frio que bajan desde las capas altas de la atmosfera.

d) A que el mar se enfría más despacio que la tierra; así, por la noche el aire caliente sobre el mar asciende dirigiéndose hacia el interior de la costa, mientras que el aire más frio de la costa se dirige por la capa inferior hacia el mar calentándose.

16. En la orilla del mar, durante el verano a mediodía, podemos notar cierta brisa ¿A qué se debe?

Esta tesis ha sido financiada por el Ministerio de Ciencia, Innovación y Universidades bajo el Programa de Formación del Personal Investigador (FPI), con el número de referencia BES-2015-075529. Asimismo, la autora de la presente tesis ha disfrutado de una estancia breve de 93 días en *Leibniz Institut für Bildungsverläufe*, Bamberg, Alemania becada por el mismo programa FPI.

GOBIERNO
DE ESPAÑA

MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES