This is a pre-print version of
Martín-del Pozo M.Á., Rascón-Estébanez D. (2021) Thinking Skills in Exam Models for CLIL Primary Subjects: Some Reflections for Teachers. In: Carrió-Pastor M.L., Bellés Fortuño B. (eds) *Teaching Language and Content in Multicultural and Multilingual Classrooms*. Palgrave Macmillan, Cham.
https://doi.org/10.1007/978-3-030-56615-9_13

THINKING SKILLS IN EXAM MODELS FOR CLIL PRIMARY SUBJECTS: SOME REFLECTIONS FOR TEACHERS

MARÍA ÁNGELES MARTÍN DEL POZO, Departamento de Didáctica de la lengua y la literatura, Universidad de Valladolid
maryange@dlyl.uva.es

DÉBORA RASCÓN ESTÉBANEZ, Departamento de Didáctica de la lengua y la literatura, Universidad de Valladolid
deborascon@fing.uva.es

ABSTRACT:

Although CLIL methodology is very relevant in education, it still confronts numerous challenges. One of them is the maximization of content learning, language learning and content competences acquisition. All factors in the teaching process are relevant: teachers, methods, classroom interaction, materials and assessment. This chapter centers on how summative assessment is presented in the exam models provided by textbooks publishers. Specifically, this analysis aims to show how several thinking skills are considered when content is being assessed.

A corpus of 30 exam models was created, including the subjects of Social Science and Natural Science, different academic years and assorted terms. The questions and tasks were classified following the categories in Bloom's Revised Taxonomy (BRT) (Anderson & Krathwohl, 2001). This taxonomy is a model for classifying thinking according to six levels of complexity: remember, understand, apply, analyze, evaluate, create. The analysis grid includes information about the verbs used in the questions and tasks (define, compare, classify, evaluate inter alia.) These verbs are closely related to the thinking skills.

Results show that the most assessed thinking skills in the analyzed exams correspond to the remember and understand categories in BRT. The analyzed exams assess content knowledge through different approaches, which do not interfere with language knowledge.

The chapter concludes with suggestions of some possible ways of reconsidering CLIL methodology in the light of assessment tasks. Also, a recommendation of the importance of implementing HOTS (high order thinking skills) and how the teaching of low order thinking skills (LOTS) may be expanded to the teaching of HOTS.

KEY WORDS: CLIL, thinking skills, Primary education, summative assessment.

1. INTRODUCTION

Content and Language Integrated Learning (CLIL) is advancing into its third decade. Practice and research have developed from the early phases. The majority of research aimed to understand language learning outcomes and language use (De Graff, 2017: xv) and was carried out by applied linguists. In comparison, studies from the vantage point of subject pedagogy and learning outcomes are still scarce. "CLIL invites investigation which draws on a much wider field of research than is associated with language learning per se" (Coyle, Hood and Marsh, 2010: 165).

This chapter aims to shed light on the pedagogical process of teaching non linguistic content in CLIL contexts, specifically on the most frequent final moment of the teaching process: summative assessment. This is one of the identified future paths for CLIL research: assessment how it may account for both content and language concerns (Dalton-Puffer & Nikula, 2014). The chapter analyses how summative assessment is presented in the exam models provided by Primary Education textbooks publishers, specifically the thinking skills identified in Bloom´s Revised Taxonomy.

2. THEORETICAL FRAMEWORK

This section of the chapter begins providing a discussion of issues in the difficult question of assessment in CLIL and potential approaches to answer it. Special attention is paid to the issue which is the object of study in this chapter: the assessment of thinking skills. The section concludes with some considerations about the context in which these thinking skills are analyzed: commercialized CLIL materials.

2.1. Assessment in CLIL

At the opening of the section, it becomes necessary to make explicit how the terms assessment and evaluation differ in CLIL contexts. Coyle, Hood and Marsh (2010: 112) reserve the term evaluation for "program evaluation" with connotations of judging effectiveness of the program. In contrast, the term assessment is used to refer to classroom context and student´s gains.

Assessment is a repeated concern for those working within the CLIL approach. Thus it is often a major area of teacher uncertainty in CLIL contexts. The following paragraphs will provide an overview of the main points in which assessment in CLIL presents challenges.

A first issue is the types of assessment processes and the classical major division of formative and summative. The former connects with assessment for learning (Assessment Reform Group, 2002). The latter focuses on the assessment of learning, since the main purpose of summative assessment is to make a judgment on a certain capability of the learner at a particular point in time. This final result usually serves as information to others, apart from the learner (parents, school managers, for instance). This chapter is to focus on summative assessment.

A second relevant issue connects to the two key linked questions which are asked a by the majority of CLIL teachers (Coyle et al., 2010: 114).

1) The *what* question: what to assess, language, content or both.
2) The *how* question: what methods can provide assessment information which is reliable because none of the elements (language and content) are impeding the other.

The *what* and *how* questions can be itemized in more specific inquiries, as Coyle et al. (2010: 114) report teachers tend to.

1. What do we assess - CONTENT or LANGUAGE?
2. What language do we assess?
3. Can students answer in their L1?
4. What tools can we use for assessment?
5. How can we assess previous knowledge and progression?
6. How can I deal with learning difficulties?
7. Provided we assess in English, how can we minimize the effect of the language in the content assessment?
8. How can we evaluate the skills/processes? Example: planning an investigation / designing a work of art / reaching conclusions.
9. How can/should we assess group work?

The nine questions in this list make evident the already mentioned uncertainty surrounding assessment in CLIL. Classroom –based research could shed light on each one of the issues. This chapter aims to make a contribution with a focus on questions 1 and 8: the assessment of content and skills.

Content should not be understood only as "knowledge acquisition" but rather as "the knowledge, skills and understanding we wish our learners to access" (Coyle et al., 2010: 53). Consequently, the delineation of which aspects of content are being assessed becomes necessary. The following aspects could be considered (Ibidem, p. 114).

- Factual recall (detail)
- General understanding (major points)
- Ability to manipulate the content, using higher-level thinking skills such as interpretation, analysis, synthesis or application.
- Ability to research more independently and extend the topic knowledge beyond what has been presented by the teacher.

An attempt to respond to the challenge in the context of CLIL programmes in Germany and Switzerland was the model by Massler, Stotz, and Queisser (2014). According to this model assessment should comprise the dimensions of 'subject specific themes' (i.e. content knowledge), 'subject-specific skills and competencies' (i.e. cognitive skills such as observing, describing, and explaining) and 'foreign language communicative competencies.

This connects to another dimension of the 4Cs CLIL Framework, cognition, which relates to learning and thinking processes. "In order to be effective, CLIL must challenge learners to create new knowledge and develop new skills through reflection and engagement in higher order as well as lower order thinking" (Coyle et al. 2010: 54). Cognition, apart from knowledge, involves the development of intellectual or thinking skills. Next section details these categories of cognitive processes.

2.2 Thinking skills in Bloom's Taxonomy

According to Banegas (2014), regarding the balance between linguistic and cognitive complexity in CLIL, authors employ Cummins' BICS (Basic Interpersonal Communication Skills) and CALP (Cognitive Academic Language Proficiency), and

Bloom's revised taxonomy (Anderson and Krathwohl, 2001). This chapter opts for the latter, explained in depth in this section.

The original taxonomy was created in 1956, by Bloom. It was revised in 2001 by Anderson and Krathwohl. The basis of the taxonomy is a double entry table, which has been maintained by the latter authors. The horizontal part focuses on the cognitive process dimension whereas the vertical column focuses on the knowledge dimension (which is the dimension used in this research). There are two reasons for this update: it was necessary "to refocus educators' attention on the value of the original" document, and because there was a "need to incorporate new knowledge and thought into the framework" (Anderson & Krathwohl 2001, XXI-XXII), as research in education had led to new data that could improve and complete the taxonomy. One of the most visible changes were the names of each category: Anderson & Krathwohl used verbs instead of nouns (See Figure 1). The reason for this change is that the Taxonomy classifies objectives, and "they contain a verb and a noun. The verb generally describes the intended cognitive process" (Anderson & Krathwohl, 2001: 4). "The continuum underlying the cognitive process dimension is assumed to be cognitive complexity" (Anderson & Krathwohl, 2001: 5). Thus, the lower categories are less complex than the higher: it is easier for a student to remember data than to analyze it, for instance. This will be reflected later in the difference between LOTS (Lower order thinking) and HOTS (Higher order thinking) (see Table 1). Whereas the goal of instruction in the category "remember" is promoting the retention of the presented knowledge, the rest of the categories focus on promoting transfer of this knowledge.

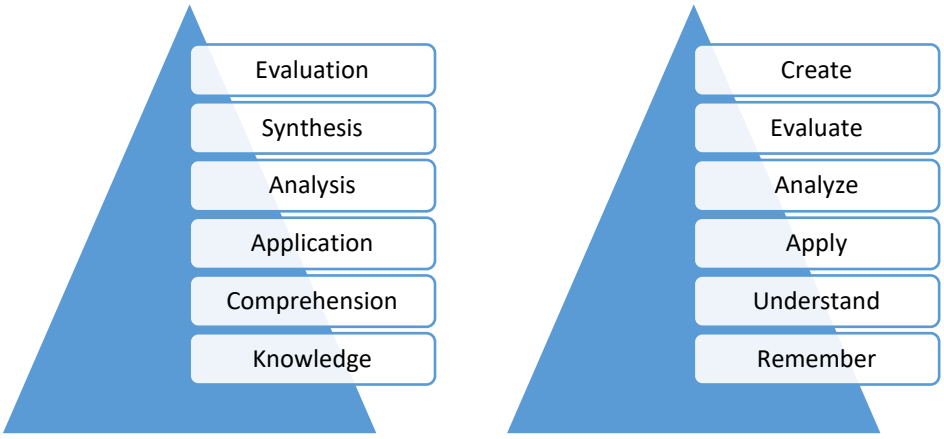


Figure 1. Comparison between the Original Taxonomy and its revision.

Table 1

HOTS and LOTS in Bloom's Taxonomy

TAXONOMY	HOTS & LOTS
REMEMBER	LOTS
UNDERSTAND	
APPLY	
ANALYZE	HOTS
EVALUATE	
CREATE	

Next paragraphs explain the new categories related to the cognitive process following Anderson & Krathwohl (2001).

a. Remember.

It happens "[W]hen the objective of instruction is to promote retention of the material in much the same way as it was taught […] it involves retrieving relevant knowledge from long term memory" (Anderson & Krathwohl 2001: 66). The student is only required to retrieve information previously studied and memorized. As Dale and Tanner (2012: 32) exposes, the question can be summarized as "Can learners remember?".

Assessment of this category is uncomplicated, because "the student is given a recognition or recall task under conditions very similar to those in which he or she learned the material" (Anderson & Krathwohl 2001: 66). When planning the exam, the teacher is going to prepare questions like: What is…?, When did….?, etc. The student can write down or say literal words from the textbook.

Even though this is the most basic category, it is significant because all meaningful learning needs to be based on the previous knowledge, what they remember.

Two cognitive processes are included in this category (see Table 2). First, recognizing or identifying -students retrieve "long-term memory in order to compare" (Ibidem: 69). Secondly, recalling or retrieving -knowledge is retrieved "from long-term memory when given a prompt to do so" (Idem), such as a question.

Table 2

Category: Remember

REMEMBER	Alternative names	Possible assessment
Recognize	identifying	True-false, match, multiple choice
Recall	retrieving	questions

b. Understand.

It refers to building "connections between the `new´ knowledge to be gained and prior knowledge" (Ibidem: 70). There are different cognitive processes in this category, summarized in Table 3 but they all can be classified as "Can learners explain?" (Dale and Tanner, 2012: 32). Examples of these verbs are:

- Interpret: change the form of representation (example: paraphrasing, or changing other codes like numbers, etc.).
- Exemplify: the student gives an "example or instance of a general concept or principle" (Anderson & Krathwohl, 2001: 71).
- Classifying: it "occurs when a student recognizes that something (e.g., a particular instance or example) belongs to a certain category (e.g., concept or principle)" (Ibidem: 72). A similar term is exemplifying, which involves the opposite (give the example and the student has to find the principle or concept).
- Summarizing: It "involves constructing a representation of the information, and abstracting a summary from it" (Ibidem: 73). Other ways of referring to it are generalizing and abstracting.
- Inferring: It "involves finding a pattern within a series of examples or instances" (Ibidem: 74). It "occurs when a student is able to abstract a concept or principle that accounts for a set of examples or instances by encoding the relevant features of each instance and, most important, by noting relationships among them" (Idem). Other cognitive tasks are: extrapolating, interpolating, predicting or concluding.
- Comparing: "involves detecting similarities and differences between two or more objects, events, ideas, problems, or situations" (Ibidem: 78). Alternative terms are: contrasting, matching or mapping.

– Explaining: “occurs when a student is able to construct and use a cause-and-effect model of a system” (Ibidem: 75). Another term is constructing.

Table 3

Category: Understand

UNDERSTAND	Alternative names	Possible assessment
interpreting	--	Paraphrase, Constructed response: supply an answer. Selected response: choose an answer. (But some information is *new*)
Exemplifying	Illustrating, instantiating	Constructed response: create an example. Selected response: select an example.
Classifying	Categorizing, subsuming	Classify, group, circle the elements that belong to one category. Constructed response: given the instance, produce the concept or principle. Selected response: given the instance, select the principle.
Summarizing	Generalizing, Abstracting	Summarize. Constructed response: read a passage and write a title. Selected response: read a passage and select the best title.
Inferring	Extrapolating Interpolating Predicting Concluding	Completion tasks, analogy tasks, oddity tasks (circle the odd one)
Comparing	Contrasting Matching mapping	Mapping, compare
Explaining	Constructing models	Explain, construct a cause-and-effect chain of events, reason why, troubleshooting (diagnose something that is wrong), redesigning, predicting.

c. Apply.

"It involves using procedures to perform exercises or solve problems" (Anderson & Krathwohl, 2001: 77). Two cognitive processes are included in this category: executing (the task is a familiar exercise) and implementing (the task is a problem; not familiar). Alternative names are carrying out or using (See Table 4). Dale and Tanner (2012: 32) expose the question that can summarize all the possible questions as "Can learners use the information in another situation?": the knowledge students have need to be transformed a little bit.

Table 4

Category: Apply

APPLY	Alternative names	Possible assessment
Executing	Carrying out	Solve a problem using a formula.
Implementing	Using	Solve an unfamiliar problem, specifying the process.

d. Analyze.

It "involves breaking material into its constituent parts and determining how the parts are related to one another and to an overall structure (Anderson & Krathwohl, 2001: 79). Dale and Tanner (2012: 32) summarizes all the possible questions in: "can learners break the information into parts and see the relationships?". Alternative names (see Table 5) are differentiating -"distinguishing the parts of a whole structure in terms of their relevance or importance" (Anderson & Krathwohl, 2001: 80)-, organizing -"identify the elements of a communication or situation and recognize how they fit together into a coherent structure" (Ibidem: 81)-, and attributing (being able to determine the underlying message of the author).

Table 5

Category: Analyze

ANALYZE	Alternative names	Possible assessment
Differentiating	Discriminating Selecting Distinguishing Focusing	Circle or select specific items, read a text about a process and divide the main steps.

		Constructed response: given some material, show the most important parts. Selected task: given some material, choose the most relevant parts.
Organizing	Structuring Integrating Finding coherence Outlining Parsing	For and against reasons. Constructed response: write an outline from a passage. Selection task: given a passage, select the best graphic diagram.
Attributing	Deconstructing	Determine the point of view or purpose of the author.

e. Evaluate.

It "is defined as making judgments based on criteria and standards" (Ibidem: 83). It involves checking -"testing for internal inconsistencies or fallacies in an operation or a product" (Idem)-, and critiquing -"judging a product or operation based on externally imposed criteria and standards" (p. 84) (see Table 6). The key questions can be: "Can learners justify a position?" (Dale & Tanner, 2012: 32).

Table 6

Category: Evaluate

EVALUATE	Alternative names	Possible assessment
Checking	Testing Detecting Monitoring Coordinating	Detect inconsistencies in a message; check whether a conclusion follows from the observed data.
Critiquing	Judging	Critique his or her own (or others') hypothesis or creations based on different criteria.

f. Create.

It "involves putting elements together to form a coherent or functional whole" (Anderson & Krathwohl, 2001: 84). In contrast with the rest of cognitive processes, it "involves the construction of an original product" (Ibidem: 85), that is, something new is going to be produced with the knowledge the person has. Alternative terms (see Table 7) are

generating - “representing the problem and arriving at alternatives or hypothesis that meet certain criteria” (Ibidem: 86)-, planning - “developing a plan for solving a problem” (Ibidem: 87) - and producing -it “involves carrying out a plan for solving a given problem that meets certain specifications” (Idem)-.

Table 7

Category: Create

CREATE	Alternative names	Possible assessment
Generating	Hypothesizing	Produce alternatives or hypothesis.
Planning	Designing	Submit the outline of a paper, develop worked-out solution plans or select solution plans for a given problem.
Producing	Constructing	A design task: create a product that corresponds to certain specifications.

2.2.1 Bloom’s taxonomy used to evaluate assessment in CLIL

Bloom’s taxonomy has served as analytical framework for studies related to the present investigation of assessment in CLIL. Special attention is paid to two of them. Both studies come from school contexts in Hong Kong during the implementation of the medium-of-instruction (MOI) initiative, which could be considered equivalent to the CLIL approach.

Chan (2016) evaluated all the test items in the examination papers in the dimensions of cognitive and knowledge categories using the revised Bloom’s taxonomy. In the investigated corpus, assessment questions with lower-level cognitive processes (e.g., the categories of “remember,” “understand,” “apply”) were more prominent than those with a higher level (categories of “analyze,” “evaluate,” “create”). Students presented a decrease in score proportions of questions assessing higher-level cognitive processes.

A second study analyzed over 4900 questions in Science/Biology textbooks, workbooks and examination papers in Hong Kong (Lo & Fung 2018). This study focused both on language demands and content demands, including cognition. The framework used (Lo & Lin 2014) was based on BRT. Results revealed that low level cognitive (i.e. recall of

knowledge) and linguistic (i.e. no production or word-level production) demands had the highest frequency in junior secondary assessment. However, higher-order thinking skills (i.e. application and analysis of knowledge) were required in senior secondary assessments. This results evidence a big leap in both cognitive and linguistic demands from junior to senior secondary education in the investigated context.

2.3 CLIL commercialized materials

Literature has repeatedly advocated the relevance of materials as a factor for successful CLIL implementation (Clegg 2007; Dalton-Puffer 2007; Coyle *et al.* 2010; Kelly 2014; Czura 2017; Ball 2018, *inter alia*). Frequently, European teachers have reported the difficulty of the scarcity of suitable CLIL materials (Morton 2013; Ball 2018), even in one of the leading countries in CLIL education, Finland (Mäkiranta 2014).

Commercially produced CLIL course books are quite a recent phenomenon and many countries still lack them or are only beginning to emerge (Morton 2013; Kelly 2014). This is not the case of Spain, where the implementation of bilingual sections and bilingual programs by concerted top-down policy generated a demand of specifically designed materials. This need has resulted on a wide range of commercially available textbooks and materials. Publishers have found a niche in the market. A glossary of Marketing terms defines niche market as: «A small, specialist area of the market. A niche market is a specific, focused, portion of a market. A segment of the market that has different preferences or needs from the mainstream audiences». This expansion does not necessarily correlate with quality, as marketed materials present assets, pitfalls and challenges (Ball 2018). In study conducted in the region of Castile and León (central Spain) teachers acknowledge a continuous improvement in published materials, though criticism on behalf of the users (these same teachers) is still extensive and harsh (Durán-Martínez & Beltrán-Llavador 2017).

Publishers' commercial websites will evince how thoroughly these editorial professionals are targeting the niche market. These textbooks and materials intend to provide numerous resources to teachers so that their effort is minimized. The supplementary materials accompanying textbooks include teachers' guide, posters, visuals, flashcards, digital

resources, online resources, extra activities, revision and reinforcement activities and exam models.

These textbooks and materials are becoming foci of research in Spain. Some questioned the real implementation of the dual focus of CLIL (Martín & Rascón 2015) or the contribution of multimedia supplementary resources to language learning (Martín & Rascón 2017). More in connection to this study, the analysis of four international UK-produced series with a CLIL component marketed in Argentina (Banegas 2014) revealed that the type of activities according to the procedures and cognitive skills only involved lower-order thinking skills, 'remember' in particular in the books targeted to students with the least level of English. Similar findings were obtained in the study of 193 activities in Natural Science textbooks in the six years of Primary Education (Criado & Carrasco 2019).

The analysis by Santo-Tomás González (2011) sheds light on the presence of HOTS and LOTS in Science books for second grade primary Spanish students. The detailed analysis of 53 activities and their correspondent text information revealed that the skills activated by the analyzed textbooks fall mainly into the category of Remember and Understand. Her suggestions for publishers to improve a dynamic pedagogy should not be disregarded either. An interesting finding of this study was that the only book which somehow promoted HOTS was a translation from the Spanish book. Her suggestions for publishers to improve a dynamic pedagogy should not be disregarded either.

Another investigation of six Natural Science textbooks for sixth year of Primary Education designed by different publishing houses revealed that 66% of the analyzed materials do not promote the necessary Higher Order Thinking Skills for the proper implementation of the CLIL approach (Cerezo & Romeu 2019).

These studies are an example of attempts to response to the challenge that "coursebook evaluation needs to be a systematic activity, […] an activity based on informed and supported views that make room for inter-subjectivity rather than personal feelings or random perceptions" (Banegas 2018: 22).

To continue with this suggested research line of "[S]pecially written CLIL course books with a European audience in mind" (Banegas 2018: 24) and undertaking the challenge of

making textbook evaluation a systematic activity, this paper is to focus on a component of the textbooks still not approached, to our knowledge: the exam models included as supplementary material. These exam models may serve the teacher for summative assessment at the end of a unit or at the end of term.

3 RESEARCH QUESTIONS, CORPUS AND METHODOLOGY

3.1 Research Questions,

The previous sections presented an overview of issues related to assessment in CLIL, some considerations and studies of marketed materials and of the classification of thinking skills which is to be used as analytical framework for this research. This paper connects these three aspects by investigating thinking skills in exam models provided with textbooks.

Specifically, the main research questions are the following:

R.Q.1. How is summative assessment presented in the exam models provided by textbook publishers?

R.Q.2. How are thinking skills considered when content is being assessed?

3.2 Corpus of study

The corpus of analysis comprises 42 exams of three courses: 4th, 5th and 6th of Primary Education, the last courses of this educational stage. The exams are from two different publishers, and the subjects are Natural Science and Social Science. These subjects were chosen because of their high status in Spanish bilingual sections. Table 8 provides the distribution of the exams across subjects and years. These exams come with the textbook as supplementary material in digital format.

Table 8

Corpus description

PUBLISHER	**COURSE**	**SUBJECT**	**Number of EXAMS**
Publisher 1	5th	Natural Science	9
		Social Science	9
	6th	Social Science	9
Publisher 2	4th	Natural Science	15

The three final years of Primary Education were selected because students are expected to be quite familiarized with the English language, both in the oral and written forms. Moreover, students´ cognitive development is moving towards more abstract capacities. On the other hand, the contents are somehow familiar as they already know about most of the topics. Subject progression entails learning more complex aspects about them. In the 42 exam papers, a total of 1953 verbs were found and classified, according to the methodology described in the following section.

3.3 Methodology of analysis

The methodology used in the analysis is mixed. In the first stage, the qualitative methodology was the most appropriate: based on the six categories created by Bloom, all the exercises were analyzed and revised, in order to place them in the suitable category. The key element in the exercises statement was the verbs. Different verbs can relate to the same category. For example, in the case of the category Remember, Figure 2 and Figure 3 show two different possibilities:

In the Natural Science exam on plants (4th grade), one of the exercises is (see Figure 2):

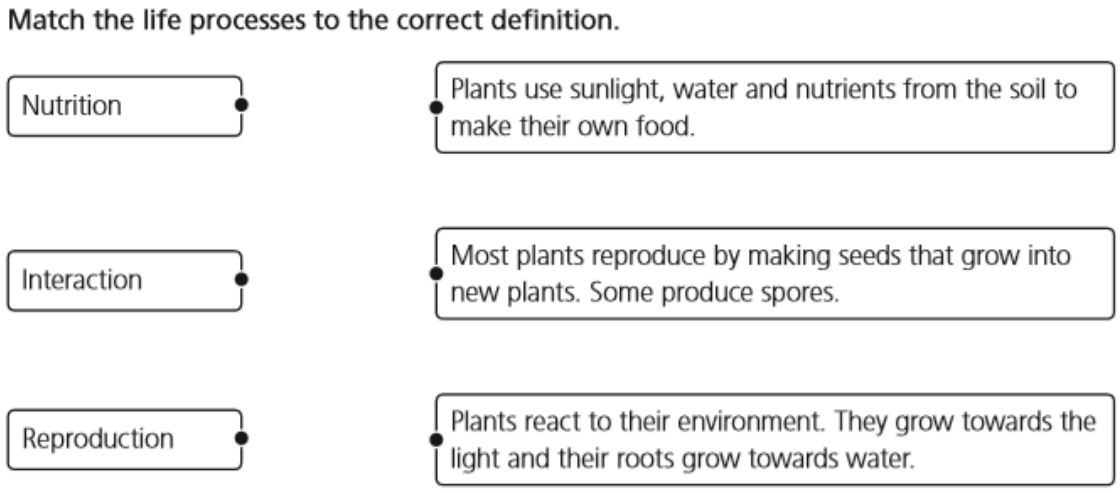
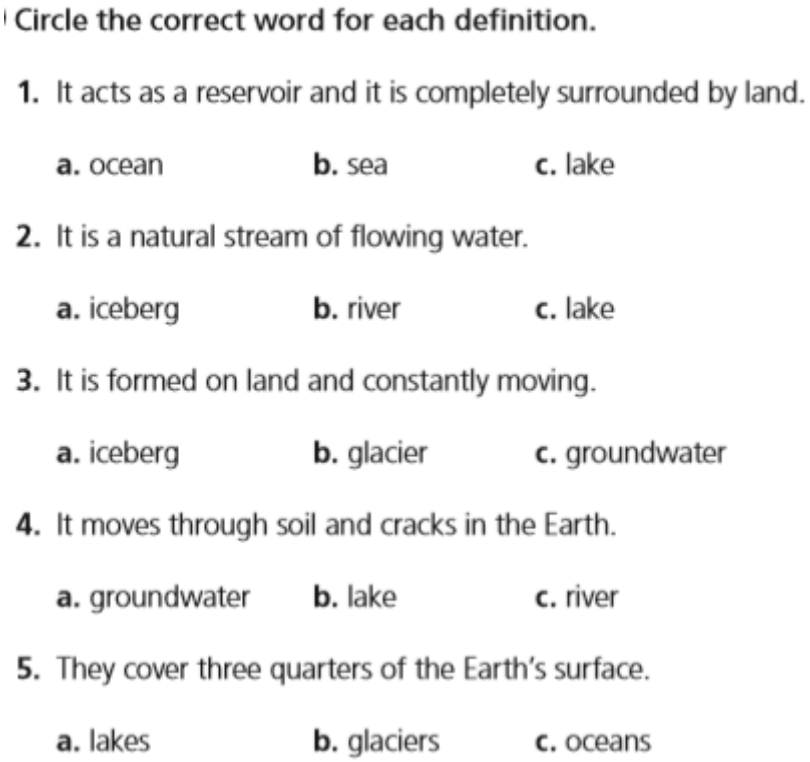


Figure 2. Example of a Natural Science exercise.

The verb “Match” corresponds to the Remember category: students only have to recognize the definition and the word defined.

Figure 3 presents an exercise for the 6th grade, in Social Science, to test the Geosphere. Students also have to recognize the definition and the word defined, but in this case, they have to remember the exact word, because there are three different possibilities for each definition.



Circle the correct word for each definition.

1. It acts as a reservoir and it is completely surrounded by land.

   a. ocean **b.** sea **c.** lake

2. It is a natural stream of flowing water.

   a. iceberg **b.** river **c.** lake

3. It is formed on land and constantly moving.

   a. iceberg **b.** glacier **c.** groundwater

4. It moves through soil and cracks in the Earth.

   a. groundwater **b.** lake **c.** river

5. They cover three quarters of the Earth's surface.

   a. lakes **b.** glaciers **c.** oceans

Figure 3. Example of a Social Science exercise.

Atlas.ti has been the tool used in this stage to create and organize the classification. In a second stage, the quantitative methodology was necessary in order to measure the different categories. To organize the data and create the figures the researchers have used the *SPSS* program.

A mixed methodology was considered convenient to target the dual aim of the research: observing the manners in which the categories occurred as well as their frequencies.

4 RESULTS

Results are similar in all courses and publishers, as can be seen in Figure 4. The four following graphics (one per textbook) show that most of the verbs correspond to the

Remember category (1495 occurrences in the corpus). Understand is the second in frequency, (233 ccurrences in the corpus), followed by Apply (175 occurrences in the corpus). Either one or the other appear in the majority of topics, although it seems that if the category Apply appears, the same will happen with category Understand, except for three topics. In a few cases other categories are found, such as analyze (three books) and create (only one book). Evaluate is the only category totally absent from the corpus.

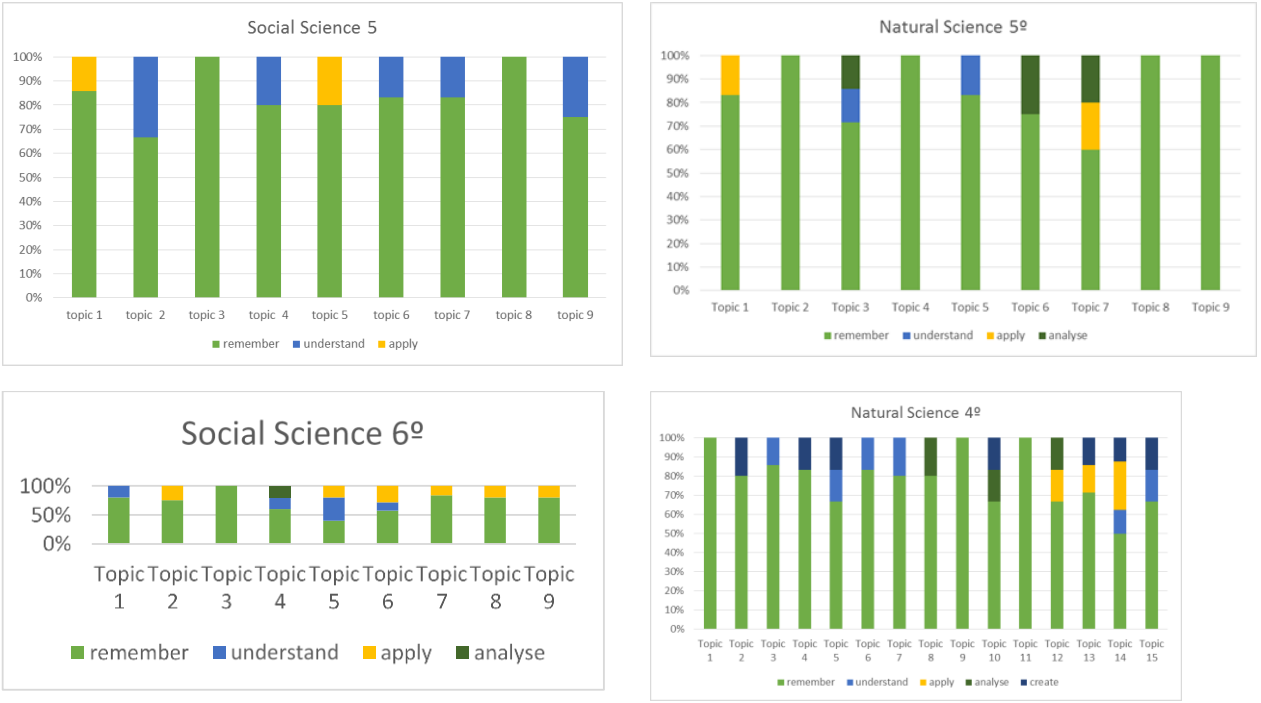


Figure 4. Total results of the different categories in each textbook.

Figure 5 shows the frequency of verbs that occur in each of the textbooks. Some types of exercises are found in all the books (circle the correct word, match, complete, label, classify, differentiate, tick or True/False sentences). They all belong to the Remember category.

It can also be observed that, on the contrary, there are some types of exercises that are present only in one book: Fill in, give examples or mark on the map are found in the Social Science book (6º), and writing, free writing and word games appear in the Natural Science book (4º).

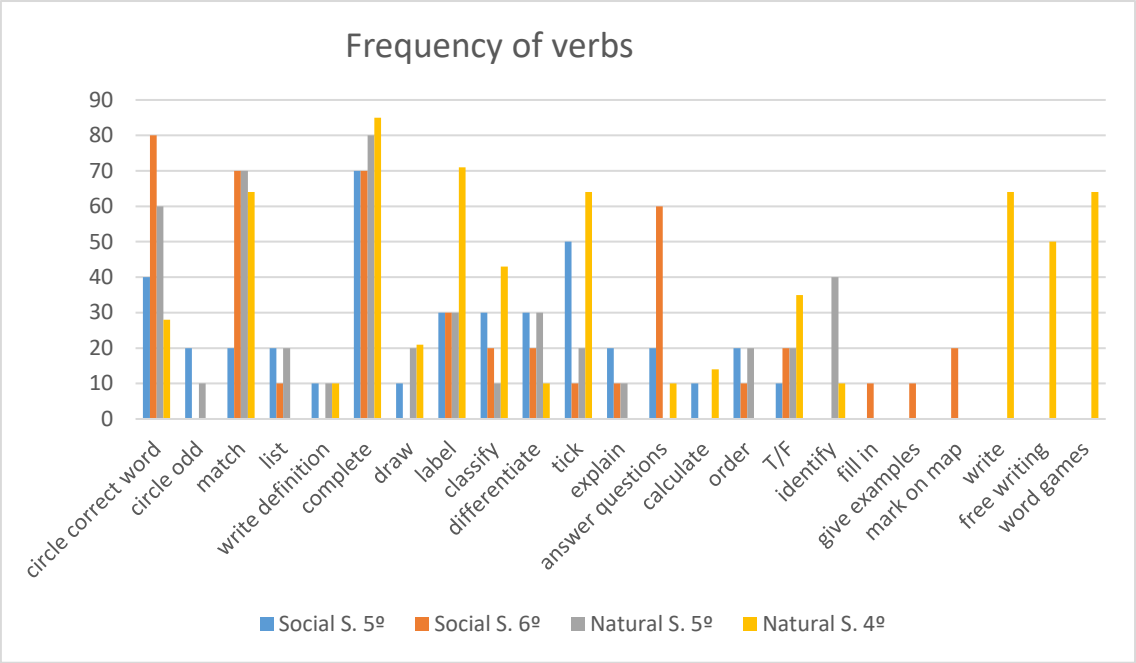


Figure 5. Total frequency of verbs.

Next, categories with the highest frequencies (remember, understand, apply) will be analyzed in detail. Figures 6, 9 and 12 show all the verbs connected to the categories in all the courses.

Category Remember.

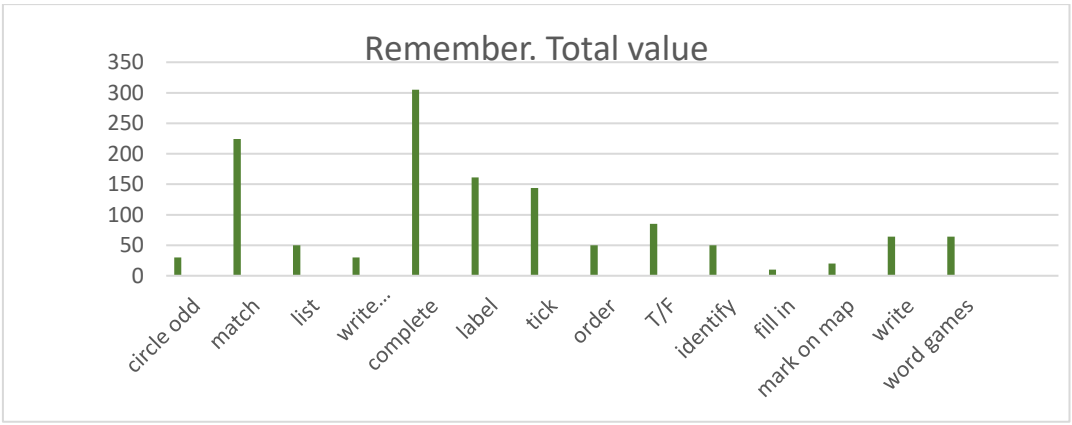


Figure 6. Results of the category Remember.

The type of activities involved in this thinking skill relates to repeating the information exposed in the unit. There is a great variety of exercises, all dealing with identifying the correct information, for example: circle the odd element (208 exercises, the most frequent one), order events, mark on a map the places mentioned in the text or identify the true or false sentences. This latter exercise can also add a second part in which students have to

correct the false sentence, so they actually prove their knowledge and do not choose randomly. In all these cases, these exercises merely require the recognition and retrieval of the information presented in class/ unit.

Some other types of exercises in the exam models require:

- Matching different words: implies that the students remember different associations, such as countries and continents.
- Listing requires repeating all the words that have something in common, such as the elements that are representative of the Romanesque style.
- Identify: it can be with an illustration (map of a city, picture of a plant/animal, photos of different styles of medieval art, etc).
- Word games: unscramble letters in order to obtain the main terms studied in the topic, for example.
- Label: it is based on an image, and students have to add the correct words in the exact place (See Figure 7).
-

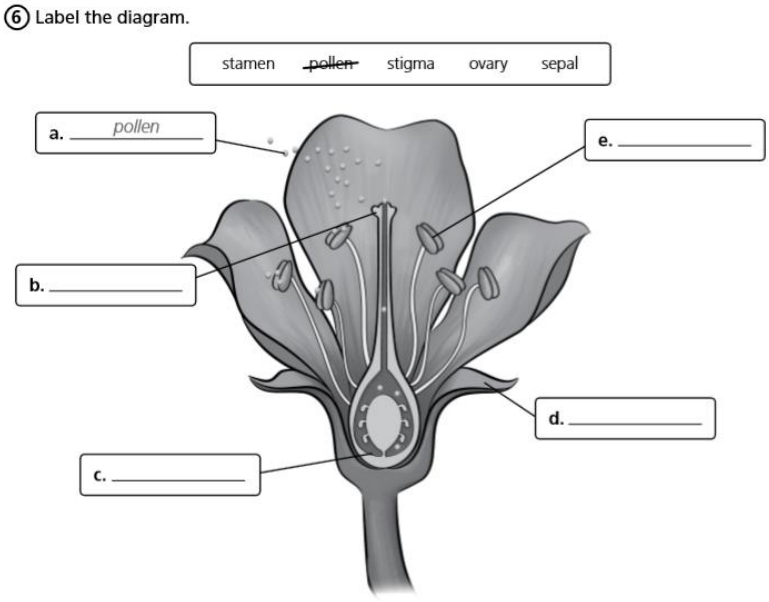


Figure 7. Example of a Label exercise.

- Complete: This type of exercise may present a variety of formats:

1. A table (it has some information, but other elements are missing and students have to add them).

2. A text: in this case, they may have to add the defined word, or add different words to a paragraph (Figure 8).

3. A diagram.

4. A timeline (the terms may appear in a table and they have to place them in the appropriate moment, or may not appear, and students have to remember them).

Depending on the course and the level of difficulty, the words to be added could be provided in a dialogue box, or (in this case it is more difficult), students have to remember them. Another more complex possibility is offering extra words that are not directly related to what appears in the exercise.

Complete the sentences with the processes involved in nutrition.

a. During _____, oxygen and nutrients are carried in the blood.

b. During _____, you absorb the nutrients your body needs from food.

c. During _____, you breathe in air and obtain the oxygen you need.

d. During _____, you expel the waste materials that your body does not use.

Figure 8. Example of a complete exercise.

– Identify countries and the continent they belong to.
– Order: it can refer to chronological order (in the case of historical events or the steps followed in a natural process, for example).
– Writing a definition implies that the students know by heart some data (even though they may not understand anything, but at least they can repeat all the words).

Category Understand

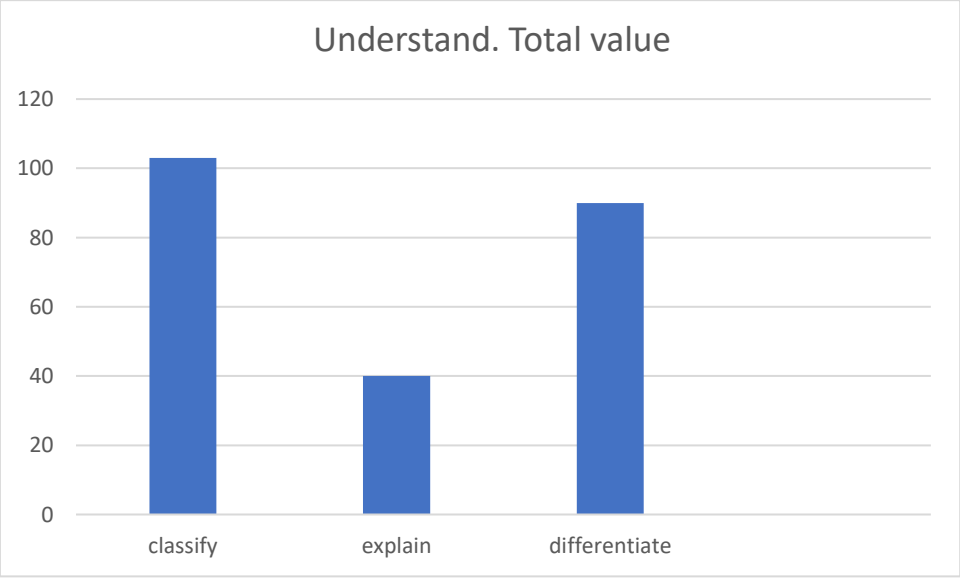


Figure 9. Results of the category Understand.

The amount of verbs in this category is rather inferior. This is the case of classify (103, the most frequent exercise), explain or differentiate (Figure 9). Students need to make connections with the information they know. The most frequent verb is Classify. The different types of exercises are described next.

- Explain (for example, explain what two different climates have in common, or a phenomenon such as lunar eclipse). This exercise can add some extra data (a graph or a map) and students have to explain what they see, according to what they have studied previously
- Classify: photos according to the period they belong to (in history), words in a pyramid to explain Feudalism; characteristics according to the style, etc. Figure 10 is an example of an activity requiring this skill.

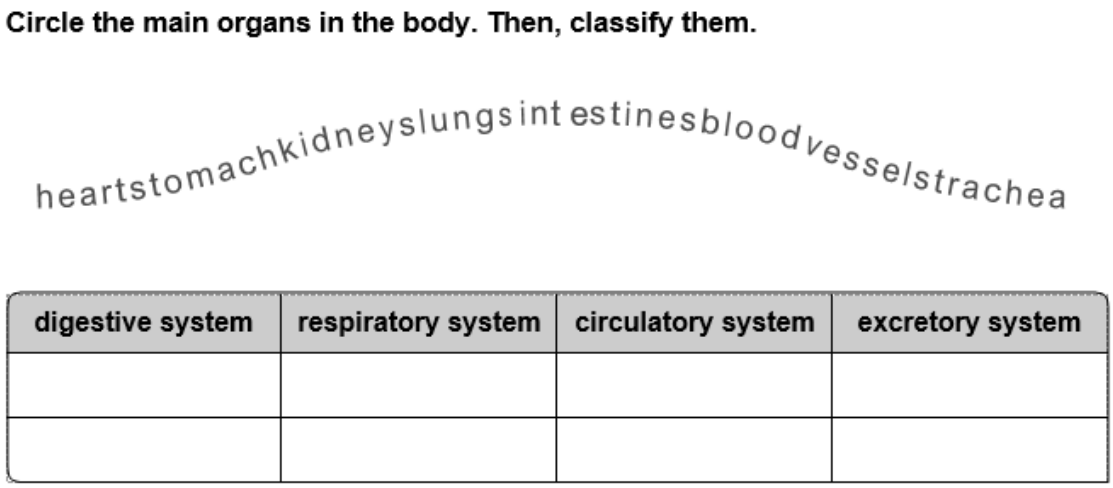


Figure 10. Example of a Classify exercise.

Category Apply

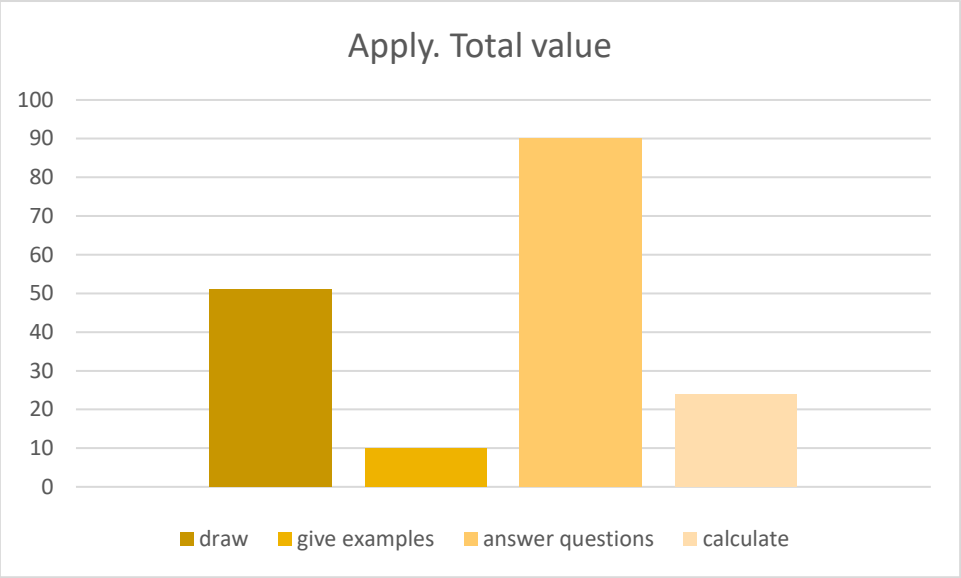


Figure 11. Results of the category Apply.

In this category there are three different types of exercises (see Figure 11): draw, give examples, calculate and answer questions. The highest frequency is found in the category "answer questions" (90 exercises). This does not have to do with exercises in which students have to answer a question related to information literally presented in the book (for example: When was America discovered?, as in the category remember). These questions require elaboration of the information. For example: What are the similarities and differences between a medieval city and a city of Al-Ándalus? In the case of Calculate, given some data, students have to solve a problem using a formula. See Figure 12 for an example.

Calculate the density of these objects. Then, answer the question.

object A mass = 60 g volume = 30 cm^3 _____

object B mass = 30 g volume = 60 cm^3 _____

- Which of these two objects floats in water? Explain. _____

_____

Figure 12. Example of an exercise requiring Calculate skill.

Category Analyze

Only four verbs related to this category have been found in the corpus: Differentiate, Write advantages, Solve problems and Calculate. The last ones refer to topics in which students have to use formulae to reach conclusions.

Category Create

Only one exam in publisher 2 includes an exercise that can be classified in this category. The exercise requires free writing (Figure 13). There are 50 of these exercises in total. Students have to write a short essay that reflects what they have learned. They have to make a text using their own words, showing what they know about a topic.

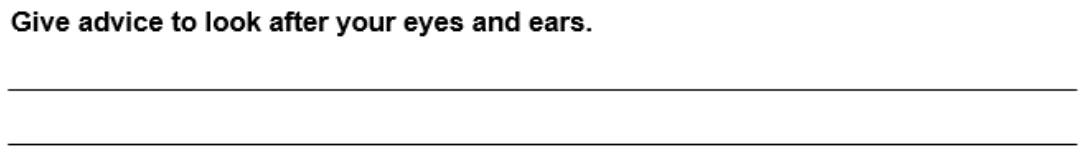


Figure 13. Example of a Create exercise.

5 **DISCUSSION**

As mentioned before, this chapter aimed to analyses the thinking skills in exam models. This section endeavors to answer the research questions posed as guidelines of the research according to the results obtained in the analysis.

The category with the highest frequency is clearly Remember. As described previously, it involves reproducing what has been explained, read or said. Remembering only includes practicing memory, but not comprehension. If students are only asked to remember, they will be failing to employ the language for uses such as creativity or knowledge construction. On the one hand, this practice is contrary to some CLIL principles which defend the need for more challenging tasks:

> In order to be effective, CLIL must challenge learners to create new knowledge and develop new skills through reflection and engagement in higher order as well as lower order thinking (Coyle et al. 2010: 54).

This and many other practices seem to indicate that, though CLIL theoretical framework is clear, practice and materials are still on an experimental phase.

On the other hand, CLIL classrooms are at risk of becoming classrooms where thinking is not promoted, as Ritchhart, Church and Morrison explain,

> Classrooms are too often places of ‘‘tell and practice.’’ The teacher tells the students what is important to know or do and then has them practice that skill or knowledge. In such classrooms, little thinking is happening […] Retention of information through rote practice isn't learning; it is training. (2011: 9).

Regarding the prevalence of categories of thinking skills, our results are in accord with all those previous studies on marketed textbooks (Santo-Tomás 2011; Banegas 2014; Cerezo & Romeu 2019; Criado & Carrasco 2019). The two latter and the first investigated Natural Science. Our study includes 18 units from Social Science and 24 Natural Science, and LOTS predominate in both. Santo-Tomás (2011) centered on second year. Our study and Cerezo & Romeu (2019) y Criado & Carrasco (2019) selected the final years in Primary. These studies refer to the tasks presented in the textbooks. Ours focuses on the suggested evaluation tasks, which, in theory should be in alignment with the content taught. Results of the four studies seem to indicate that commercialized textbooks and materials used in Spanish Primary Education tend to promote LOTS more than HOTS and, in consequence, summative assessment is aligned and also makes a choice for LOTS tasks. Though Banegas (2014) did not study CLIL textbooks but only the CLIL component in English as a Foreign Language books, his findings of the frequency of LOTS are also significant to our study because, as he concludes "in this respect, students with a low level of EFL are treated as students with low cognitive abilities" (252), statement which seems to be supported by the findings in all these studies together with ours. This fact is counter to what Genesee and Upshur (1996) forewarn; the content objectives assessed in contexts of second language education (such as CLIL) should be identical to those who receive instruction in their L1, and "lower standards of achievement should not be established for second language speakers" (47).

In a similar vein, our results coincide with studies which focused on examination questions (Chan 2016; Lo & Fung 2018). Questions seem to be placed low in the continuum in the cognitive domain. The remote contexts, and thus the much potential dissimilarity these contexts may present, raise an imperative question: if whenever and wherever an L2 is used as a medium of instruction the cognitive challenge of the content should decrease.

This discussion section closes with a quote from experienced and authorized teacher trainers. It seems to indicate the (undesirable) frequency of HOTS in CLIL classrooms and the corresponding requirement to move towards LOTS:

> When learners find the input first, firstly use questions which appeal to lower-order thinking skills (LOTS), such as remembering and understanding. However, since analysis of teachers' questioning in the classroom has shown that 70-80% of classroom questioning focuses on these skills of remembering and understanding (Wragg and Brown, 2001), it is important to challenge learners' thinking behaviors too. (Dale and Tanner, 2012: 32).

Though it was not one of the main research questions in this study, the analysis of tasks permitted to observe that their format require little language knowledge in most of the cases. Thus, content recall is not language challenging (grids, matching information, labels) as can be observed in the activities in Figures 2, 3, 7, 8, 10. The activities adhere to Coyle´s et al. recommendation for quality: "It is important to allow learners to express their responses to tasks in the most direct way possible so that language is not a barrier to demonstrating understanding o f content (2010: 123).

6 CONCLUSIONS

The chapter opened with reflection on some issues relate to assessment of content in CLIL contexts. BRT has been described and proposed as a convenient analytical framework to evaluate the cognitive dimension of content. The description of the different stages in the continuum of BRT could encourage teachers to reflect, as indicated in the title of the chapter.

Some considerations on commercialized textbooks and materials were presented. Studies reviewed indicated that tasks and questions in these materials tend to be low in the cognition continuum proposed by BRT. Findings in the present investigation are in concurrence with previous studies.

As a whole, studies like the present chapter and those in accord could serve to:

1) Raise CLIL teachers' awareness of their existing assessment practices.
2) Increase teachers' awareness of the potential influence of cognitive and linguistic demands on students' performance.

3) Support teachers in the recognition of strengths and weaknesses of textbooks.
4) Assist teachers in the development and fostering of HOTS and in consequence move beyond tasks which only require retention and understanding.

As final considerations, if assessing content is potentially very challenging, the complexity increases when students are concurrently learning language and learning content in all their dimensions, including thinking skills. The introduction to the chapter underlined the need for research that moves beyond examining CLIL students' L2 competence. We hope having contributed further to the understanding of how content and language are integrated in CLIL practices, specifically in the aspect of assessment and thinking skills.

REFERENCES

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

Assessment Reform Group. (2002). *Assessment for Learning: 10 Principles. Research-based principles to guide classroom practice*. Available at: http://www.qca.org.uk/libraryAssets/media/ 4031_afl_principles.pdf [Accessed 27/4/2009].

Ball, P. (2018). *Innovations and Challenges in CLIL Materials Design. Theory into Practice* 57/3, 222-231. DOI: 10.1080/00405841.2018.1484036

Banegas, D. L. (2014). An investigation into CLIL-related sections of EFL coursebooks: issues of CLIL inclusion in the publishing market. *International Journal of Bilingual Education and Bilingualism* 17/3, 345-359.

Banegas, D. L. (2018). Evaluating language and content in coursebooks, in M. Azarnoosh, M. Zeraatpishe, A. Faravani & H. R. Kargozari (eds.) *Issues in coursebook evaluation*. Leiden: Brill, 21-29.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co Inc.

Cerezo, J. & Romeu, C. (2019). *Thinking skills and CLIL: Two sides of the same coin?* Spanish Association of Applied Linguistics, Bilingualism: languages in contact. Valladolid, 27-29/4.

Chan, J. Y. H. (2016). The Fine-tuning Medium-of-Instruction Policy in Hong Kong: A Case Study of the Changing Schoolbased Test Papers in Science Subjects. *Education Journal* 44/1, 159–193.

Clegg, J. (2007). Analysing the Language Demands of Lessons Taught in a Second Language. *RESLA* 20/1, 113-128.

Coyle, D., Hood, P. & Marsh, D. (2010). *CLIL. Content and Language Integrated Learning*. Cambridge: Cambridge University Press.

Criado, R. & Carrasco, J.A. (2019). *Investigating CLIL textbooks in Spanish Primary Education*. Spanish Association of Applied Linguistics, Bilingualism: languages in contact. Valladolid, 27-29/4.

Czura, A. (2017). Translation is not Enough - the Need for Pedagogical Adaptation in CLIL Textbook Development, *Porta Linguarum* 27, 35-46.

Dale, L. & Tanner, R. (2012). *CLIL activities. A resource for subject and language teachers*. Cambridge: CUP.

Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated Learning (CLIL)*. Amsterdam: John Benjamins Publishing.

Dalton-Puffer. C & Nikula, T. (2014) Content and language integrated learning. *The Language Learning Journal* 42/2, 117-122, DOI: 10.1080/09571736.2014.891370

Decreto 26/2016, de 21 de julio, por el que se establece el currículo y se regula la implantación, evaluación y desarrollo de la Educación Primaria en la Comunidad de Castilla y León. file:///C:/Users/Admin/Downloads/Anexo%20I.B.pdf

Durán-Martínez, R. & Beltrán-Llavador, F. (2017). Key issues in teachers' assessment of primary education bilingual programs in Spain. *International Journal of Bilingual Education and Bilingualism*. DOI: 10.1080/13670050.2017.1345851

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2017). Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism* 22/6, 1-14.

Foran, D. (2008). Evaluación en CLIL. *Aula de Innovación Educativa* 15/168, 38-44.

Genesee, F. and Upshur, J. A. (1996). *Classroom-Based Evaluation in Second Language Education*. Cambridge: Cambridge University Press.

Kelly, K. (2014). *Ingredients for successful CLIL*. British Council, Teaching English. Available at: http://www.teachingenglish.org.uk/article/keith-kelly-ingredients-successful-clil-0 [Accessed 25/4/2019]

Lo, Y and Fung, D. (2018). Assessments in CLIL: the interplay between cognitive and linguistic demands and their progression in secondary education, *International Journal of Bilingual Education and Bilingualism*. DOI: 10.1080/13670050.2018.1436519

Lo, Y. Y., & A. M. Y. Lin. (2014). Designing Assessment Tasks with Language Awareness: Balancing Cognitive and Linguistic Demands. *Assessment and Learning* 3, 97–119.

Mäkiranta, P. (2014). *CLIL Teachers as Materials Designers*. (Master's thesis). Faculty of Education: University of Jyväskylä.

Marketing Association of Australia. *Glossary of Marketing Terms*. Available at http://www.marketing.org.au/ [Accessed 2/9/2014]

Martín del Pozo, M.A. & Rascón Estébanez, D. (2015). Textbooks for content and Language Integrated Learning: policy market and appropriate didactics? *Foro de educación* 1, 123-141.

Martín del Pozo, M.A. & Rascón Estébanez, D. (2017). *Recursos multimedia de los libros de texto CLIL (Natural Science): ¿qué aportan al aprendizaje de la lengua?*. Actas del III congreso de Educación Mediática y competencia digital. Facultad de Educación, Facultad de Ciencias Sociales, Jurídicas y de la Comunicación, UVA. 2688-2702.

Martín del Pozo, M.A. & Rascón Estébanez, D. (2019). *Middle Age through English, Middle Age through Spanish: spot the difference*. Spanish Association of Applied Linguistics, Bilingualism: languages in contact. Valladolid, 27-29/4.

Martínez López, J.A. & Cantero García, V. (2014). In search of an effective model for assessing learning in bilingual education: the authentic assessment. *Porta Linguarum* 21, 137-150.

Massler, U., D. Stotz, and C. Queisser. (2014). Assessment Instruments for Primary CLIL: The Conceptualisation and Evaluation of Test Tasks. *The Language Learning Journal* 42/2, 137–150.

Merriam-Webster Dicionary. Available at: https://www.merriam-webster.com/dictionary/think [Accessed 11/07/2019]

Morales, M.Y. y Restrepo, I. (2015). Hacer visible el pensamiento: alternativa en educación para el aprendizaje. *Infancias Imágenes*, 14/2, 89-100.

Morton, T. (2013). Critically evaluating materials for CLIL: Practitioners' practices and perspectives. In J. Gray (Ed.), *Critical perspectives on language teaching materials*. Basingstoke: Palgrave Macmillan, 111–136.

Ritchhart, R., Church, M., & Morrison, K. (2011). *Making thinking visible: How to promote engagement, understanding, and independence for all learners*. Jossey-Bass: New York.

Santo-Tomás González, M. (2011). *From low to high order thinking skills in CLIL Science Primary textbooks: a challenge for teachers and publishers*. (Tesis de Máster). Universidad Complutense de Madrid.

Wragg, E.C. & Brown, G. (2001). *Questioning in the Primary School*. Abingdon: Routledge Farmer.

This is a pre-print version of
Martín-del Pozo M.Á., Rascón-Estébanez D. (2021) Thinking Skills in Exam Models for CLIL Primary Subjects: Some Reflections for Teachers. In: Carrió-Pastor M.L., Bellés Fortuño B. (eds) *Teaching Language and Content in Multicultural and Multilingual Classrooms*. Palgrave Macmillan, Cham.
https://doi.org/10.1007/978-3-030-56615-9_13