

## **Which score is adequate: Approximation to the assessment rationale used in a Science through English CLIL written test**

*Miquel Àngel Fuentes, Universitat Autònoma de Barcelona  
Barcelona, Spain*

Article received 6 December 2012, accepted 24 June 2013, final version received 18 October 2013

---

Despite the surge of Content and Language Integrated Learning (CLIL) studies in many areas, assessment practices in CLIL contexts are considered to be somewhat unaccounted for (Poisel, 2007; Hönig, 2009; Maggi, 2012). This paper intends to contribute to this area and addresses the issue through a case study of a Science through English CLIL unit implemented at a state-funded secondary school in Barcelona, Spain<sup>1</sup>. The paper focuses on the written test of the unit and analyzes the students' responses and degree of satisfaction of the teacher after implementing and correcting the test, thus providing valuable insights into teacher development and offering a snapshot of CLIL assessment practices and their effects. The results allow a close look at the teacher's attitude towards content and language errors and points out the lights and shadows of assessment in CLIL while highlighting the need for divulgation of further assessment practices in similar contexts.

### **The relevance of language in the Science classroom**

Within Science teaching, language is a common matter of concern, especially when it comes to discussion of the opportunities (and obstacles) that assessment can grant to students, in particular those who are in the process of learning the language in which the content is taught, when the language is not their L1 (Lyon, Bunch & Shaw, 2012). One common belief among teachers is that students must first learn the language before learning the content (Lee, Maerten-Rievea, Buxton, Penfield & Secada, 2009). Similarly, Stobart (2008) states that the core problem is that many teachers perceive language as a technical issue to control while assessing, instead of as an inherent part of their own culture and the culture of assessing. Likewise, Lyon (2013) argues that teachers experience tensions between language assessment and scaffolding demands. Moreover, proficiency in the L1 has been directly linked with success in content-based assessment (Solano-Flores & Nelson-Barber, 2001).

However, several studies point out the benefits that learning Science in another language can entail for students if approached from a *talking science* perspective (see, for example, Bennett, Hogarth, Lubben, Campbell, & Robinson, 2010; Osborne, 2010; Smith, 2010 or Morton, 2012). Following Vygotskian ideas, Lemke (1990) proposes a model of teaching and learning science through talking science. This statement is not simply reduced to talking about science, but refers to a broader vision centered on the fact that language is not just vocabulary and grammar, but also a system of resources for making meanings, which allows reasoning and problem solving. Sprouting from the core idea that communication and teaching are always social processes, talking science involves the creation of a *community of practice* (Lave & Wenger, 1991) in which mastering the target knowledge allows members to become more central and valuable to the community. This stance is further linked to the sociocultural perspective of the lesson, through which “teachers should encourage students to discuss their own ideas on each topic, thus accepting alternative views which respect alternative religious and cultural positions while presenting the view of science and the reasons for that view” (Lemke, 1990:168). Drawing on the talking science model, Laplante (1997) emphasizes its potential for second language development among students and the necessity of teacher professional development in Science education so as to cater for minority students.

### **The weight of language and content in CLIL assessment**

Based on the traditional tension between content and language, assessment in CLIL has become a challenging issue due to its dual focus. This tension leads teachers to take decisions that can greatly affect the implementation of CLIL sessions (Llinares, Morton & Whittaker, 2012). Indeed, deciding aspects such as whether students are to be evaluated through language and content together or separately and the level of linguistic and non-linguistic output in the assessment tasks may condition the creation and implementation of materials.

From an ideal CLIL standpoint, as its nature of integration points out, assessment of goals in language and content should be performed simultaneously and should carry equal weight in the evaluation. For this purpose, several studies attempt to shed light on this area and provide some guidelines. For instance, Serragiotto (2007)

states that the weight of not only content goals but also the relevance of language in the assessment should be determined and shared with students in advance. Maggi (2012) takes into account the role of language even further, suggesting teachers should have a precise idea of their students' level of the target language before implementing CLIL projects by using tools such as the *Common European Framework of Reference* (CEFR) (Council of Europe, 2009) as a basis. Hönig (2009) highlights the dilemma among CLIL teachers whether or not to assess language use and the need to define the language awareness in order to achieve consistent results in the assessment. Coyle, Hood & Marsh (2010) state that language in CLIL must be approached from an instrumental perspective, thus remaining at the service of the development of the target content.

Hönig (2009) proposes the development and implementation of assessment instruments that incorporate both content and language-focused criteria as a measure which allows students to get solid credit for the knowledge and skills displayed. The creation of integrated assessment tools for CLIL is an area currently being researched. Recent contributions to this field include Poisel (2007) who proposes that a range of different assessment tasks is the most viable approach and encourages "self centered study for portfolio work combined with peer tutoring and the teacher as a facilitator" as a highly efficient assessment approach (p.46), and Barbero's (2012) set of holistic and analytic rubrics featuring a list of descriptors for content and language goals which draw on the CEFR.

## Context

### *The school*

The school where the sample was collected belongs to a state-funded secondary school in the metropolitan area of Barcelona. The school is implementing a Foreign Language Educative Project grounded in the use of a foreign language to work on the curriculum of content-based subjects. This project includes teaching Science through English. The teacher in this study received support from the collaborative research team CLIL-SI<sup>2</sup>, which was actively involved in the design of the materials and assisted the teacher through the implementation of the unit.

The sample was collected from a 1st year group made up of sixteen students between the ages of eleven and twelve. The school featured four groups, with a total of

60 students in the same course. However, the school decided to offer only 30 CLIL places. Afterwards, through parents' choice and an initial English level test, two small groups of 16 and 14 students were made. Consequently, 30 students were taught Science in English and the other 30 were taught Science in Catalan. All four groups shared the same Science teacher, however. Some of the students from the observed group had already gone through CLIL experiences in primary school and had experience working collaboratively in small groups or pairs.

The group responds to a middle class social status and has very few immigrant students, with the exception of one of the students who is an English native speaker. Besides this case, the students displayed an approximate level of competence in English of A1.2 according to the CEFR. The teacher held a certified B2 level in English competence with 25 years of experience teaching Science and one year of experience teaching Science through English.

#### *The teaching unit*

The teaching unit was implemented during the 2009-2010 course in the students' first year of secondary education and its materials were designed with the support of the collaborative research team CLIL-SI. The unit, titled *Life*, covers the topic of the Catalan 1<sup>st</sup> year of secondary education Science curriculum *Life in Action* (Departament d'Educació de la Generalitat de Catalunya, 2008:92). The unit spanned 12 sessions and featured the following assessment tasks:

- (a) a dossier which had to be handed in by the end of the unit, in which the teacher checked that students had followed and completed the tasks implemented during the unit.
- (b) an oral presentation given by students in pairs or groups of three in which they had to explain the features of certain endangered animal species.
- (c) a written test which examined the content covered throughout the implementation of the unit.
- (d) the participation of students and their overall attitude towards work and behavior in the classroom.

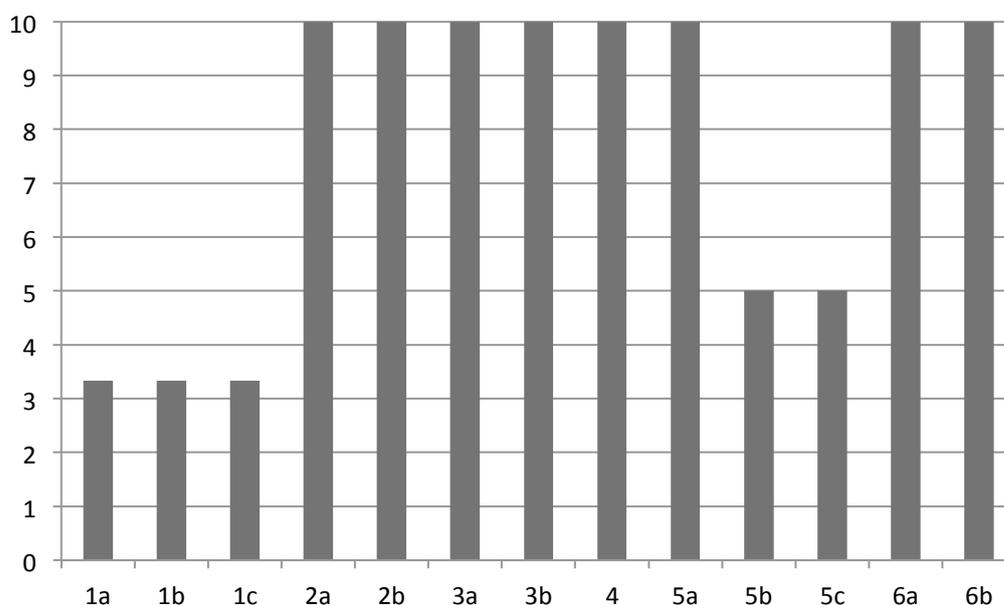
This study, which focuses on the written test, belongs to larger research project (Fuentes, 2011) and shares its data with a larger set of studies from the DALE-APECS project<sup>3</sup>. Some of the outcomes from this data corpus focused on the interactional co-

construction of dialogic explanations (Evnitskaya, 2012). Co-relative studies such as Evnitskaya and Morton (2011) allow for a broader scope on the teacher's conception of the classroom and, therefore, of the assessment.

### ***Distribution of the score among the test items***

The test is assessed over a total of 100 points in which, in order to pass, the student has to gather a minimum of 50 points out of a maximum score of 100 points. The figure below summarizes the distribution of the 100 points among the test items, in which the x axis represents the number of the test item and the y axis refers to the score of each question (over a maximum of 10 points):

Table 1. Distribution of the test score among test items



As can be seen in the table, most of the test items feature sub-questions and the score is unevenly distributed. The first item is divided in three parts which sum up to 10 points. The second item is formed by two 10-point sub-questions. The third item follows the same structure as the second. The fourth item only features one question which has a weight of ten points. The fifth item is split in three parts, the first one of ten points while the second and the third being of 5 points. Finally, the sixth and final item shares the same structure as the second and third items.

### ***Construction of the test items***

The information concerning the construction of the test items is ascertained through the field notes gathered by the researcher during the implementation of the unit, together with later written exchanges of feedback between the researcher and the teacher.

First item:

*Draw a picture of three different animal cells and name them. Explain what features they have in order to do their job properly.*

This item proposes three sub-tasks. The first one requests a non-linguistic output: drawing, followed by a second, linguistic output (naming) and finally demands a more complex task (explaining). The maximum score that students can get in this question are 10 points, with an equal distribution of the points among the three sub-questions (3.333 points each).

Second item:

*(1) Write down the different levels of specialization from cell to organism. (2) Classify the next words into the different levels of organization from the smallest to the biggest one: tree, plant cell, wood, palisade, chloroplast, leaves.*

This item proposes two tasks. The first one corresponds to a task in which students need to display the target knowledge, while the second demands a classification of the data provided in the first part. Each part represents 10 points of the score.

Third item:

*In two close islands a scientist found two different populations of very similar reptiles. What can he do to know if they belong to the same species? Can you say how species (sic.) are named by scientists? Write down an example.*

This item also presents two different parts. The first one has an open question format and requests students to apply the target knowledge. The second part corresponds to a more methodological area of the target knowledge of the unit which students need to defend through exemplification. Each part represents 10 points of the score.

Fourth item:

*Fill in the blanks with the appropriate word: Both are microorganisms. (1) Viruses are always ..... and bacteria are not. (2) Viruses cannot carry out the three ..... but bacteria do. (3) Bacteria and viruses can produce ..... (4) Bacteria have ..... and viruses ..... (5) Some bacteria are beneficial, for example those which produce .....*

This question comprises five statements in which students are expected to fill in the gaps with one word. Each statement has the value of 2 points, which add up to a maximum score of 10 points. These statements focus on the similarities and differences between the bacteria and the virus.

Fifth item:

*Identify the parts of the following cells. Say if they are prokaryotic or eukaryotic. Justify your answers.*

This item features two parts of a lower cognitive demand and a third part which demands a more elaborate answer (in this case, a justification). The item presents the aid of graphic support of stereotyped cell pictures to help students identify the parts of the cell. The first part (identify the parts of the cell) is worth 10 points, while the second (identify them as prokaryotic or eukaryotic) is worth 5 points. The justification has a further value of 5 more points.

Sixth item:

*Name some physical adaptations that help animals survive to: (1) blend in the snow: ... (2) breathe underwater: ... (3) reach leaves high up in the trees: ... (4) grab things and feed: ... (5) attract a mate: ... . Now, give examples of these adaptations.*

This item features two parts as well. The first part is formed by five sub-questions in which students are asked to name physical adaptations for a set of survival of the species objectives. Each sub-question has a value of 2 points, adding them up to 10 points of the score. In the second part, students are asked to supply examples of each of the five aforementioned physical adaptations, allowing students to score 2 points for each correct answer, up to a total of 10 points.

## Objectives of the study

This paper focuses on two objectives, the first one being to understand the rationale used by the teacher to grade language and content in the test and the second aiming to describe the students' performance in relation to the demands set by the teacher, which also involves contrasting this performance with the teacher's expectations as a sub-objective.

## Method

This study was carried out within an action-research project (Lewin, 1946; Stenhouse, 1975; Carr & Kemmis, 1986) which stemmed from a school-university partnership with teacher development purposes.

The university-based researcher became a participant observer in the CLIL classroom during a period of seven weeks, which corresponded to the later part of the twelve-week implementation of the CLIL teaching unit *Life on Science* through English. The data corpus for this paper is formed by the following elements:

1. The written test developed as one of the main assessment tools for the CLIL unit.
2. The outcomes of the test from the 16 students who belonged to the group, with their respective teacher's grading and comments.
3. A 24-minute, audio-recorded, focused semi-structured interview between the researcher and the teacher held during the correction of the tests, in which correction criteria and the teacher's feelings toward the test outcomes are addressed.

To describe the rationale used by the teacher to grade language and content, the 16 samples of the students' outcomes with the teacher comments were approached qualitatively in two stages, which correspond to Erickson's (2006) description of *molar* and *molecular* analysis. The first one (molar) scanned the 16 samples in order to find samples in which students had been given full mark on a certain item and other samples in which the students had been given half of the full mark for their outcome. The process was repeated for each of the test items. Afterwards, two samples of each category (full mark samples and half mark samples) were chosen randomly for each item. The second stage of the analysis (molecular) approaches the samples

comparatively and contrastively in order to reveal the criteria that led the teacher to give full mark and to deduct points for each of the items.

To understand the students' performance in relation to the demands set by the teacher, the students' scores were arranged quantitatively in order to ascertain the average grades of the students over the maximum score, using it as an spectrum (ranging from 0 to 100 points) which reveals the degree of satisfaction of the teacher regarding each item, student and on the whole. The results are then triangulated with the teacher through an interview, which has been audio-recorded, in order to check whether these preliminary findings match the teacher's impressions or not. Both perspectives are finally compared and contrasted qualitatively so as to shed light on the objective.

### The rationale used by the teacher to grade language and content

The first focus of this paper compares and contrasts couples of randomly selected samples from each item which were granted a full mark with other couples of randomly selected samples from each item which were granted half of the full mark.

#### Item 1

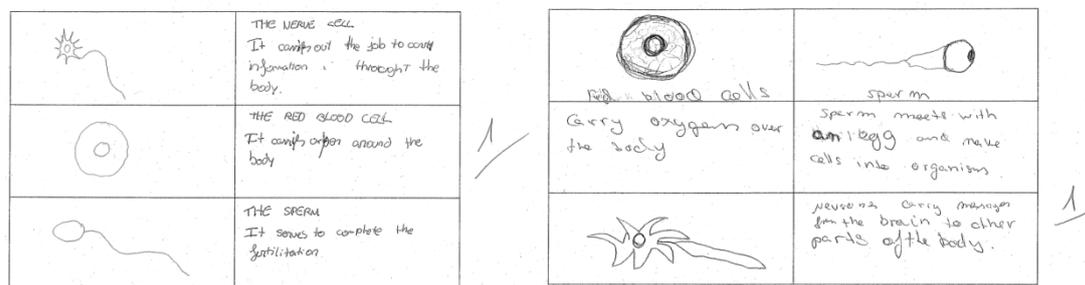


Figure 1. Two samples of quality students' outcomes which were granted full mark

In these samples, students got full score for the item (10 points) as they drew, named and explained the features of each cell. Students construct their short description (featuring approximately 10 words, distributed in one or two sentences) with the aid of the drawing. On the sample on the left, the teacher corrected the misspelled word "oxigen", reminiscent of its Catalan spelling, for "oxygen" and changed "carry" to "carries" to add coherence with the subject but did not subtract any score for that. Moreover, the teacher ignored the misspelling of the word "throught".

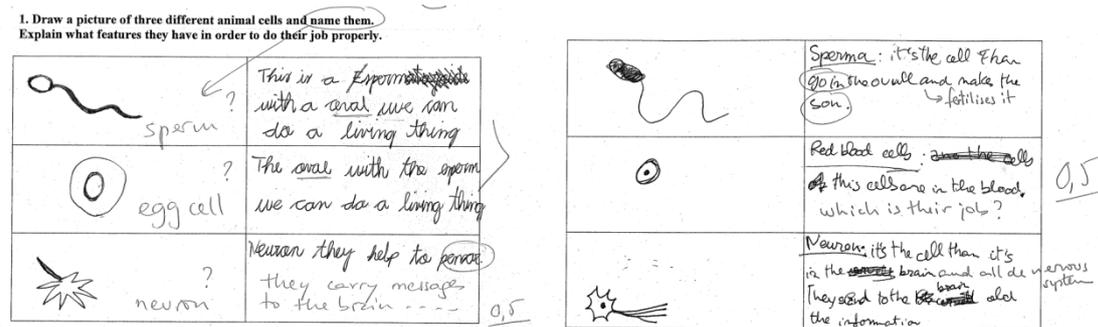


Figure 2. Two samples of students' outcomes which received half of the full mark

In these samples, both students got 5 points out of 10. In the sample on the left, the teacher remarks the fact that the drawings have not been named, which led to the teacher deducting 3 points. Besides, two more points are deducted for the misuse of the word “oval” for “egg cell”. Other mistakes are highlighted but no score is deducted. One is the crossing out of the “e” in the word “esperm”. Another circled word is “pensar” (Catalan for “think”), for which the teacher provides a more suitable model alternative (“they carry messages to the brain”).

In the sample on the right, points have been subtracted due to lack of completeness in the student's responses. In the first one, the teacher adds “fertilizes it” as a more proper alternative to “make the son”. In the second one, the teacher comments “which is their job?” to point out the missing information. Finally, in the third one, the teacher adds “and all the nervous system” to the student's response. Regarding the remarks which are highlighted but have not been deducted, the teacher circles “go in” and “son” but does not provide any comments or alternatives and changes “sand” to “send”. The teacher ignores mistakes such as the use of “ovull” for “egg cell” or the use of “old” as if it was “all”.

## Item 2

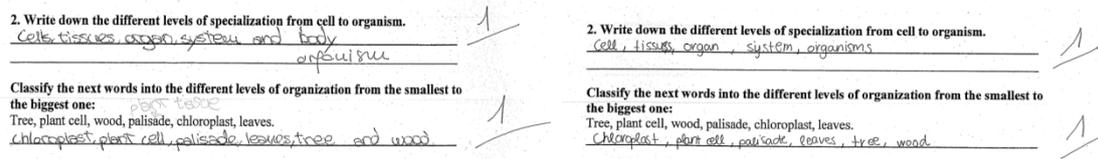


Figure 3. Two samples of quality students' outcomes which were granted full mark

In the samples in figure 3, the students have successfully completed the proposed tasks. In the sample on the left, the choice of the lexical item “body” has not led to the

deduction of points but the teacher has provided the more suitable “organism” below the word.

2. Write down the different levels of <sup>organization</sup> specialization from cell to organism.

Classify the next words into the different levels of organization from the smallest to the biggest one:  
Tree, plant cell, wood, palisade, chloroplast, leaves.

1) 2) 3) 4) 5) 6)

2. Write down the different levels of specialization from cell to organism.

Classify the next words into the different levels of organization from the smallest to the biggest one:  
Tree, plant cell, wood, palisade, chloroplast, leaves.

Figure 4. Two samples of students' outcomes which did not receive full mark

In figure 4, in the sample of the left, the student has not answered the first question, thus obtaining no score. Regarding the second part, the student is given half of the maximum mark by getting half of the items (4, 5 and 6) in the correct order. In the case of the sample on the right, the student got the first part with a perfect score but misplaced one of the items in the second, thus was subtracted 2 points from the total score, getting 8 points in total.

### Item 3

3. In two nearby islands a scientist found two different populations of very similar reptiles. What can he do to know if they belong to the same species?

They can look their DNA and compare it with another reptile. Also they can look the similarity at the others reptiles.

Can you say how species are named by scientists? Write down an example.

The scientist say the name species in latin. One example is "Ursus Arctos"

3. In two nearby islands a scientist found two different populations of very similar reptiles. What can he do to know if they belong to the same species?

They can put together two specimen which are from different sex. If they reproduce and get descendation they are from the same species.

Can you say how species are named by scientists? Write down an example.

The scientist use the latin to name the species. The pine's scientific name is pinus.

Figure 5. Two samples of quality students' outcomes which were granted full mark

In these samples (figure 5), the teacher considers that the knowledge of the target content is adequate and grants full mark to these students. Their explanation consists of two complete sentences. In the sample on the left, the teacher corrects the capitals in the Latin name “Ursus Arctos” but does not deduct from the student’s score. Also, the teacher ignores mistakes such as non subject-verb agreement or the misspelling of the word “spieces” (most likely caused by a typing error in the creation of the test which resulted in the misspelling in the wording).

In the sample on the right, the teacher adds brackets to the invented lexical item “descendation” and provides the more suitable “offspring”. The teacher also crosses out “the” before “latin”, corrects the misspelled “scientistic” for “scientific” and adds a capital “P” in “pinus”. However, the teacher does not subtract points for these mistakes.

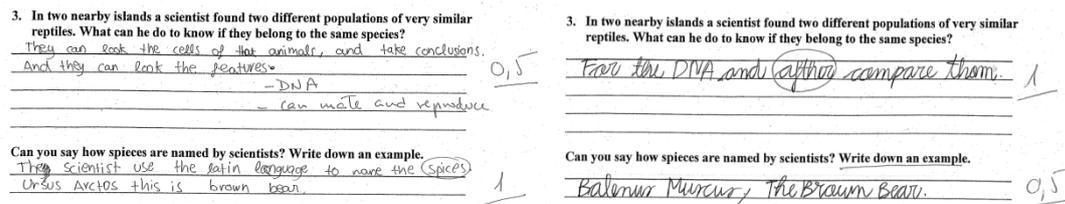


Figure 6. Two samples of students' outcomes which did not receive full mark

In the left sample in figure 6, the explanation provided in the first part is given half of the full score, with the teacher providing more suitable alternatives (“D.N.A.” and “can mate and reproduce”). Also, the teacher circles the misspelled “spices” but does not deduct points for it. The teacher ignores mistakes such as subject-verb agreement.

In the sample on the right, the student provides the requested example on the second part, but does not provide any explanation, which results in half of the total score. Moreover, the teacher circles the misspelled word “afther” but does not deduct anything for it.

#### Item 4

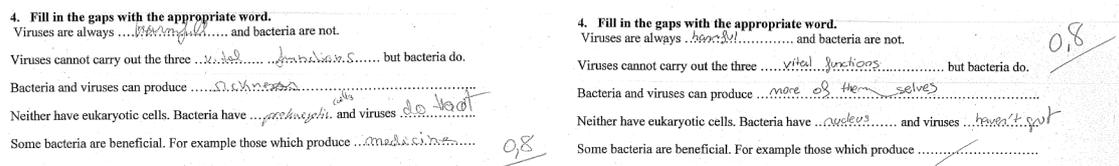


Figure 7. Two samples of students' outcomes which were granted the highest marks in the group

No students managed to get a perfect score in this item. In this case, two of the best samples have been randomly selected instead. In the left sample, the teacher completes the fourth item with “prokaryotic cells” and corrects “do too” for “do not”. These two errors result in deducted points. Moreover, the teacher ignores the misspelled “harmfull”. The sample of the left features an empty response, which meant deducted points. The teacher links “them” with “selves” but does not deduct from the total score. The teacher also ignores the ungrammatical “haven’t got”.

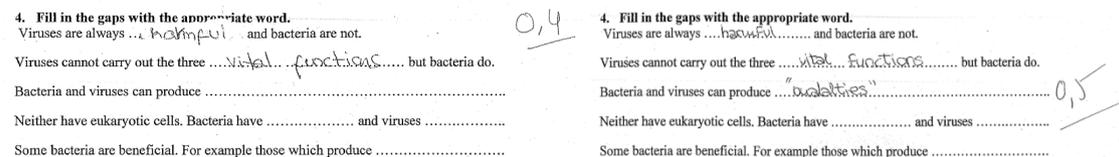


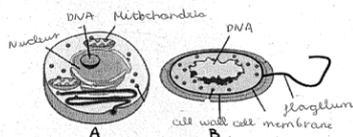
Figure 8. Two samples of students' outcomes which received half of the total mark

In figure 8, points are deducted for empty responses in the sample on the left. The right sample features a similar case, but is granted an extra point for providing the Catalan

alternative (“malalties”) to the target English word (“illnesses”), given that it is one of the target lexical items that students were expected to have learnt throughout the implementation of the unit.

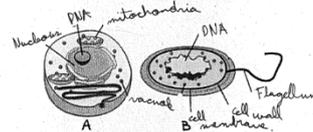
### Item 5

5. Name the parts of the cells.  
Say if they are prokaryotic or eukaryotic. Why? Justify your answers.



The A is a Eukaryotic cell, because it has a nucleus \*  
The B is a Prokaryotic cell, because doesn't have nucleus and  
has a flagellum  
\* and mitochondria.

5. Name the parts of the cells.  
Say if they are prokaryotic or eukaryotic. Why? Justify your answers.

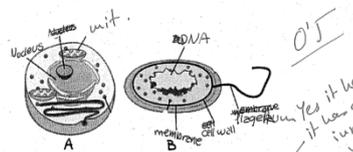


The cell B is prokaryotic because it's simpler than the  
other. The prokaryotic cell doesn't have structures. So  
the cell A is eukaryotic.

Figure 9. Two samples of quality students' outcomes which were granted full mark

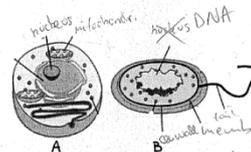
In the samples, the students successfully provided the right labeling for each item demanded. Moreover, both students presented a justification which covered the target knowledge for this item adequately according to the teacher's standards. In the left sample, the teacher adds the subject “it” and changes “have” to “has” in agreement with the subject, but does not deduct points for it. In the right sample, the teacher adds the words “one” and “many” but does not deduct for those issues either.

5. Name the parts of the cells.  
Say if they are prokaryotic or eukaryotic. Why? Justify your answers.



The cell A is a Eukaryotic cell because is bigger and has DNA. Any  
The cell B is a prokaryotic cell because it has DNA and is smaller

5. Name the parts of the cells.  
Say if they are prokaryotic or eukaryotic. Why? Justify your answers.



A is eukaryotic and B is prokaryotic

Why?

Figure 10. Two samples of students' outcomes which received half of the full mark

In figure 10, the student on the left manages to label 4 out of the 8 parts correctly, thus achieving half of the full mark. The student also manages to identify the cells correctly, which gains him 5 more points. However, the student fails to provide a correct justification, which prompts the teacher's comment (“Yes it has. It has got DNA inside the nucleus!”) and deducts the 5 points from the justification part.

In the case of the sample on the right, the student also manages to get 4 out of the 8 parts right and to identify the cells correctly. However, the absence of justification results in a deduction of points and prompts the teacher's comment (“Why?”).

## Item 6

6. Name some physical adaptations that help animals survive to:

- blend in the snow: camouflage
- breathe underwater: gills
- reach leaves high up in the trees: long neck
- grab things and feed: hands
- attract a mate: fur

Now, give examples of species that have got these adaptations.  
Rabbits: camouflage. Gills: fish. Long neck: giraffe. Hands: monkeys. Fur: lynx.

6. Name some physical adaptations that help animals survive to:

- 1 - blend in the snow: withe hair
- 2 - breathe underwater: branquies = Gills
- 3 - reach leaves high up in the trees: the long neck
- 4 - grab things and feed: urpes / ungles
- 5 - attract a mate: beautiful color or plumas / feathers

Now, give examples of species that have got these adaptations.  
wite rabbit 2. fish 3. giraffe 4. talp 5. pabo real

Figure 11. Two samples of quality students' outcomes which were granted full mark

In figure 11, the sample on the left shows that the student got all the lexical items right. The teacher highlights the misspelled word “camouflage” in two occasions but does not deduct points for the misspelling. The sample on the right is also given full score. The teacher circles the misspelled “withe” and “wite” for “white”. The teacher also circles the Catalan “urpes” (“claws”), “ungles” (“nails”), “talp” (“mole”) and the Spanish “pabo real” (*sic.*) (“peacock”). The teacher also adds “feathers” next to its Catalan equivalent “plumas” (*sic.*) and “gills” next to the Catalan “branquies” (*sic.*).

6. Name some physical adaptations that help animals survive to:

- blend in the snow: \_\_\_\_\_
- breathe underwater: the gills ✓
- reach leaves high up in the trees: \_\_\_\_\_
- grab things and feed: \_\_\_\_\_
- attract a mate: the "plumes" ✓ feathers

Now, give examples of species that have got these adaptations.  
For example, the royal wite use its "plumes" to attract a mate.

6. Name some physical adaptations that help animals survive to:

- blend in the snow: changeable fur ✓
- breathe underwater: gills ✓
- reach leaves high up in the trees: moths
- grab things and feed: get a trunk ✓
- attract a mate: "plomes" feathers

Now, give examples of species that have got these adaptations.  
Gills → wite fish ✓ moths → algunes papallones moths són papallones  
plomes → royald turkey

Figure 12. Two samples of students' outcomes which did not receive full mark

In the left sample, three spaces left blank are deducted from their corresponding score. The use of the Catalan “plumes” (*sic.*) prompts the teacher’s addition of its English equivalent “feathers”, but the teacher does not deduct points for it. The student also gets the points for offering one example of adaptation. Similarly, the teacher also adds “feathers” next to the Catalan “plomes” in the sample on the right. In that sample, a point is deducted for providing the wrong adaptation in the first part. In the second part, a point is deducted for not providing enough examples. One of the examples provided (“moths = algunes papallones” which would be translated as “moths = some butterflies”) is also deemed inappropriate by the teacher, who adds the comment “moths són papallones” (“moths are butterflies”). Moreover, the teacher ignores the response “royald turkey” (*sic.*), which roughly corresponds to a literal translation of the Spanish “pavo real” (“peacock”).

### Summary of the teacher's attitude towards content and language mistakes

The following table presents a categorization the teacher's attitude toward the variety of content and language-related errors that appeared during the sample responses.

Table 2. Categorization of the teacher's attitudes towards content and language mistakes

	Highlighted + Deduction of score	Highlighted	Ignored
Lack of response	X		
Wrong content	X		
Not enough content	X		
Target vocabulary in L1	X		
Errors in target vocabulary	X		
Code-mixing		X	X
Grammar mistakes		X	X
Errors in vocabulary		X	X

From the content perspective, the reasons which led to the deduction of points were: (a) leaving the item in blank, (b) providing an inadequate response and (c) providing an incomplete response. From a language perspective, only errors related to the target vocabulary resulted in a deducted score (misspellings or using their L1 alternatives). The use of the L1 in other contexts, the grammar mistakes and the errors in vocabulary which were not deemed as target vocabulary by the teacher were either just highlighted or ignored.

### Triangulation of the results with the teacher (i)

The following set of quotations belongs to the audio-recorded interview with the teacher and the responses have been translated from Catalan into English. These quotations cover the teacher's personal reflections after correcting the test and her general attitude towards foreign language assessment:

- “Besides focusing on content, I am also taking into account the target language that has been used and worked on throughout the unit.”
- “I am not focusing on problems of form and items which have not been part of the unit and do not pose a serious problem for the overall comprehension of the message and target content.”
- “The help of a language teacher in order to tackle the language assessment is highly necessary. Assessing both the content and the foreign language is just too much for just one teacher.”

The teacher emphasizes the amount of extra work that CLIL assessment demands and the insecurities that arise from not being an expert in foreign language teaching (Lyon, 2013). The attitude towards language assessment corresponds with Höning's (2009) results and the instrumental conception of language in the assessment brings this practice closer to the "European" CLIL strand defined by Coyle, Hood & Marsh (2010).

### Students' performance in relation to the demands set by the teacher

The first part of the second focus of this paper approaches the students' results quantitatively, arranging them by student and test item, which provides a preliminary display of the satisfaction of the teacher in relation to students' responses.

Table 3. Distribution of the teacher's score among students' responses and averages of the group

	1abc	2a	2b	3a	3b	4	5a	5bc	6a	6b	Total
Student 1	10	10	10	0	5	8	8	8	10	2	71
Student 2	10	0	5	10	0	8	5	5	10	10	63
Student 3	3	0	10	10	5	8	5	5	10	10	66
Student 4	8	10	10	10	5	5	8	10	10	10	86
Student 5	5	10	10	10	0	4	5	5	8	10	67
Student 6	2	0	10	10	5	8	2	0	4	4	45
Student 7	10	10	10	10	10	9	10	10	9	10	98
Student 8	3	10	10	8	0	2	5	10	0	2	50
Student 9	10	10	10	5	10	5	10	0	8	10	78
Student 10	7	10	10	0	0	5	3	0	2	0	37
Student 11	10	10	10	0	0	5	6	0	10	10	61
Student 12	5	10	10	10	7	5	5	10	10	8	80
Student 13	10	10	10	10	5	6	5	6	8	4	74
Student 14	10	10	10	10	10	7	10	10	8	10	95
Student 15	5	10	8	10	5	2	6	0	4	4	54
Student 16	10	10	8	10	5	7	10	10	10	10	90
<b>Average</b>	<b>7,38</b>	<b>8,13</b>	<b>9,44</b>	<b>7,69</b>	<b>4,5</b>	<b>5,88</b>	<b>6,44</b>	<b>5,56</b>	<b>7,56</b>	<b>7,13</b>	<b>69,69</b>

To begin with, the average of the total grades is 69.69 points over a total of 100. This leads to the assumption that, taking into account the fact that students needed to achieve a minimum score of 50 points in order to pass, students managed to meet the teacher's expectations to a certain extent, even though they did not manage to achieve a big amount of high quality responses. The fact that the lowest score in the group is 50 points shows that all students managed to pass the test. The highest scores of the group are those of three students, who got totals of 98, 95 and 90 points, showing that some students achieved mostly high quality responses throughout the whole test, meeting the teacher's demands almost exactly.

The test item which shows the highest average is 2b (which required students to extrapolate their knowledge from their experience with animal cells to plant cells). The average of this test item is 9.44 over 10, leading to the assumption that this area was thoroughly covered in the unit so students were mostly able to succeed in this task. The test item with the lowest average is 3b (which asked how species were named by scientists and to provide an example), with an average of 4.5 points. This leads to the assumption that students' mostly did not grasp that point during the unit and perhaps further work on the specific area should have been provided for students to master that content.

### **Triangulation of the results with the teacher (ii)**

The following section includes quotations relating to the teacher's satisfaction towards the students' responses in the test and general impressions after the implementation of the test as a key part of the assessment of the unit.

- “The assessment of the unit puts the motivation of the group at stake.”
- “If many students fail the subject, they will complain that they would have got a much higher mark if the test had been in their L1.”
- “The fact that this is the first test of Science through English that students have taken adds extra pressure when grading students.”
- “I get the impression that students have relaxed in terms of mastering the target content due to the exam being in English.”
- “Students have performed generally lower than I had expected.”
- “I feel that it is necessary to implement an extra session in which we will go over the exam together in order to review and consolidate the target knowledge of the unit.”

These reflections are in this case a key element in order to ascertain the degree of satisfaction of the teacher in relation to the students' outcomes. These reflections underscore the importance of assessment in CLIL practices (in agreement with Serragiotto, 2006) and the diversity of circumstances that the teacher needs to consider before assessing students, besides having solid information on the students' level of competence in the target L2 (Maggi, 2012). The reflections also pinpoint the need for further development of CLIL assessment tools (Poisel, 2007; Barbero, 2012) which can support the CLIL teacher and ease the planning and implementation of CLIL practices.

## Conclusions

The results of the analysis and their triangulation with the teacher's interview support the claims of previous literature on the fact that CLIL assessment practices prove to be a highly challenging area for teachers, who feel insecure about what to assess and how to address language corrections (Hofmannová, Novotná & Pípalová, 2004; Poisel, 2007; Llinares, Morton & Whittaker, 2012). Despite the focus on content in the CLIL practices assessment, research in the area of Science underscores the importance of language in the Science classroom and the opportunities that it can afford students of different learning profiles (Lemke, 1990; Laplante, 1997).

Drawing on that research, it is relevant to extrapolate these findings to CLIL contexts in order to achieve a better grasp on how to assess language in the CLIL classroom and what benefits assessing language may entail for students. The teacher's interview also provides insights on the area of satisfaction, revealing that the mark was, in this case, not totally representative of the quality of the students' responses and how many other factors, such as the students' motivation and the status of the CLIL project on the school added extra pressure to the grading. Therefore, assessment in CLIL proves to be a highly challenging area for teachers, highlighting the need for proper assessment tools and guidelines in order to implement CLIL projects with more security and from a more informed perspective. As a case study, this paper intends to be a contribution to the field which, in agreement with Hönig (2009), stresses the need for further study into the assessment of CLIL practices and teacher development.

## References

- Barbero, T. (2012). Assessment tools and practices in CLIL. *Franca Quartapelle*, 38-56.
- Bennett, J., Hogarth, S., Lubben, F., Campbell, B., & Robinson, A. (2010). Talking science: The research evidence on the use of small group discussions in science teaching. *International Journal of Science Education*, 32(1), 69-95.
- Carr, W., & Kemmis, S. (1986) *Becoming critical: education, knowledge and action research*. Lewes, Falmer.
- Council of Europe. (2001). *Common european framework of reference for languages: learning, teaching, assessment*.
- Coyle D., Hood P., & Marsh D. (2010) *CLIL, content and language integrated learning*. Cambridge: Cambridge University Press.
- Departament d'Educació de la Generalitat de Catalunya. (2008). *Curriculum Educació Secundària Obligatoria*.  
[http://phobos.xtec.cat/cda-monestirs/web/media/curriculum\\_eso.pdf](http://phobos.xtec.cat/cda-monestirs/web/media/curriculum_eso.pdf) (accessed 18 October 2013).

- Erickson, F. (2006). Studying side by side: Collaborative action ethnography in educational research. In G. Spindler & L. Hammond (Eds.), *Innovations in educational ethnography: Theory methods and results* (pp. 235-258). Mahwah, NJ: Lawrence Erlbaum.
- Evnitskaya, N., & Morton, T. (2011). Knowledge construction, meaning-making and interaction in CLIL science classroom communities of practice. *Language and Education*, 25(2), 109-127.
- Evnitskaya, N. (2012). *Talking science in a second language: The interactional co-construction of dialogic explanations in the CLIL science classroom*. Doctoral Dissertation (unpublished). Universitat Autònoma de Barcelona.
- Fuentes, M.A. (2011). *Pathways to CLIL assessment: an approach to the idiosyncrasies of classroom-based evaluation practices in a Science in English CLIL classroom*. Research in Language and Literature Teaching Official Master's Degree Dissertation. (unpublished). Universitat Autònoma de Barcelona.
- Hofmannová, M., Novotná, J., & Pípalová, R. (2008). Assessment approaches to teaching mathematics in English as a foreign language (Czech Experience). *In Depth*, 3(4), 5.
- Hönig, I. (2009). Assessment in CLIL –A case study. *Current Research on CLIL*, 3, 36.
- Laplante, B. (1997). Teaching science to language minority students in elementary classrooms. *NYSABE Journal*, 12, 62-83.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Lee, O., Maerten-Rievea, J., Buxton, C., Penfield, R., & Secada, W. (2009). Urban elementary teachers' perspectives on teaching science to English language learners, *Journal of Science Teacher Education*, 20(3), 263-286.
- Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing Corporation.
- Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, 2(4), 34-46.
- Llinares A., Morton T., & Whittaker R. (2012). *The roles of language in CLIL*. Cambridge: Cambridge University Press.
- Lyon, E. G., Bunch, G. C., & Shaw, J. M. (2012). Navigating the language demands of an inquiry-based science performance assessment: Classroom challenges and opportunities for English learners. *Science Education*, 96(4), 631-651.
- Lyon, E. G. (2013). What about language while equitably assessing science? Case studies of preservice teachers' evolving expertise. *Teaching and Teacher Education*, 32, 1-11.
- Maggi, F. (2012). How things started and developed. *Franca Quartapelle*, 18-28.
- Morton, T. (2012). Classroom talk, conceptual change and teacher reflection in bilingual science teaching. *Teaching and Teacher Education*, 28(1), 101-110.
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328, 463-466.
- Poisel, E. (2007). Assessment modes in CLIL to enhance language proficiency and interpersonal skills. *VIEWZ*, 16(3), 43-46.
- Serragiotto G. (2006). La valutazione del prodotto CLIL. In F. Ricci Garotti (Ed.), *Il futuro si chiama CLIL: una ricerca interregionale sull'insegnamento veicolare* (pp. 159-160). Trento: IPRASE del Trentino.

- Smith, K. (2010). Talk it up! Developing students' oral scientific literacy. Poster presented at the annual Capstone Project, Montana State University, Bozeman, MT.
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553-573.
- Stenhouse, L. (1975). *An introduction to curriculum research and development*. London: Heinemann.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. New York, NY: Routledge.

---

<sup>1</sup> This study has been funded by the MICINN (R+D+i EDU2010-15783 project).

<sup>2</sup> <http://grupsderecerca.uab.cat/clilsi/>

<sup>3</sup> Further information available at <http://grupsderecerca.uab.cat/clilsi/content/dale-apecs>

---

**Author's Reference:**

**Miquel Àngel Fuentes** finished his Master Degree in Research in Literature and Language Teaching (2011) and is currently PhD candidate at the Universitat Autònoma de Barcelona. He works as a secondary school teacher while carries out his research, focused on CLIL as an empowerment tool for students with different learning profiles.

**Email:** [miquelangel.fuentes@gmail.com](mailto:miquelangel.fuentes@gmail.com)

---

To cite this article:

Fuentes, M. (2013). Which score is adequate: Approximation to the assessment rationale used in a Science through English CLIL written test. *Bellaterra Journal of Teaching & Learning Language & Literature*, 6(4), 54-73.