

# Language Learning through Tasks in a Content and Language Integrated Learning (CLIL) Science Classroom<sup>1</sup>

CRISTINA ESCOBAR URMENETA AND ANTONIO SÁNCHEZ SOLA

*Departament de Didàctica de la Llengua i la Literatura  
Universitat Autònoma de Barcelona*

Received: 28 July 2007 / Accepted: 17 January 2008

ISSN: 1697-7467

**ABSTRACT:** This paper presents the results of a pre-experimental study into language learning in inclusive Content and Language Integrated Learning (CLIL) classrooms at secondary school, in which Science content was taught in English. The pedagogical experience lasted four weeks. The main pedagogical approach followed was the use of learning tasks carried out by students working in dyads.

Through the use different indicators extensively employed in the field, the study attempts to measure the gains that students show in fluency and lexical repertoire in a pretest / treatment / post-test research design. The post-test demonstrated significant progress in the assessed indicators. The study also showed that improvements transcend purely formal linguistic aspects and that it is necessary to find other measurement tools which may help us to capture the extent of those improvements.

The limited scope of the sample does not allow us to make strong claims about the suitability of pedagogical decisions adopted; however the outcomes encourage us to continue to explore how cooperation among students influences learning in CLIL classrooms.

**Key words:** Content and Language Integrated Learning (CLIL), Foreign Language Learning, Inclusive Classroom, Science, Cooperation, Fluency, Lexical Repertoire, Semi-Immersion (SI), Tasks.

## **Aprendizaje por tareas de lenguas extranjeras en la clase de ciencias a través del AICLE**

**RESUMEN:** Este trabajo presenta los resultados de un estudio pre-experimental sobre aprendizajes lingüísticos en aulas inclusivas de Aprendizaje Integrado de Contenidos y Lengua (AICLE) de secundaria, donde se impartieron contenidos de Ciencias Naturales (CN) en inglés, en una experiencia pedagógica de cuatro semanas de duración, cuya principal opción didáctica fue la realización de tareas en diadas de aprendices.

El estudio tuvo como objetivo determinar los avances de los estudiantes en repertorio léxico y fluidez, mediante la utilización de diferentes indicadores ampliamente utilizados en estos ámbitos de investigación. Para ello se recurrió a un diseño del tipo pretest / tratamiento / posttest. El posttest demostró progresos significativos en los indicadores

---

<sup>1</sup> Our gratitude to the anonymous reviewers of our text, who helped us clarify obscure points and contributed to the improvement of the final text. Also our recognition to Ms. E. Moore for her linguistic support and to Dr. M.T. Escobar for her useful advice on statistical issues.

evaluados. También puso en evidencia que el progreso trasciende los aspectos puramente formales, y que se precisan medidas más mejor ajustadas, que puedan contribuir a capturar el alcance de dichos progresos.

La extensión limitada de la muestra no nos permite formular conclusiones definitivas sobre la adecuación de las decisiones pedagógicas adoptadas; sin embargo los resultados nos animan a continuar explorando cómo la cooperación entre estudiantes influye sobre el aprendizaje en las aulas AICLE.

**Palabras clave:** Aprendizaje integrado de contenidos y lengua (AICLE), Aprendizaje de lengua extranjera, Aula inclusiva, Ciencias Naturales, Cooperación, Fluidez, Repertorio léxico, Semi-inmersión (SI), Tareas.

## 1. INTRODUCTION

The construction of European Union and the globalisation of information require that the education systems of member countries guarantee the training of citizens capable of functioning comfortably in more than one foreign language (Council of Europe, 1995).

Complementing the teaching *of* foreign languages with the teaching of non-linguistic subjects using the foreign language as the vehicle in the classroom, that is learning *in* the foreign language, is one option gaining ground in our country in both social<sup>2</sup> and political spheres of education, as results of experimental programmes are made public.

The process is advancing rapidly, and in Royal Decree 1513/2006, of 7<sup>th</sup> December, in which the benchmarks for primary education were set, such practices, which were until now purely experimental, became official.

In the study we present, which comes under the umbrella of the project entitled «*Evaluation of collaborative tasks and the achievement of learning objectives in Science in Foreign Language CLIL classrooms*»<sup>3</sup> [ArtICLE], a series of investigations centred on a) determining the results of teaching-learning activities in terms of students' progress in linguistic and curricular learning, and b) observing classroom practices in order to understand and improve teaching approaches in Catalan public schools. That is, in schools which do not select students on religious, economic or academic grounds, and in a context in which, at least, two languages coexist for day to day communication.

The hypothesis from which the ArtICLE research projects depart is the following:

*Conversation between learners stemming from structured pedagogical tasks shapes a favourable environment for both academic and linguistic learning, in inclusive CLIL classrooms.*

With the aim of verifying (or refuting) this hypothesis, a task-based teaching sequence based on science curricular content was created. Consequently, the main requirement that the

<sup>2</sup> For example, the promotion of curricular content through the means of English was incorporated as a «priority proposal» in the Strategic Accord signed by social forces – business and workers' unions – and Catalonia's government, the Generalitat, in February, 2005.

<sup>3</sup> ArtICLE obtained public funding in the 2004 and 2005 calls for Educational Research and Innovation projects from AGAUR (ARIE 2004-00058 and ARIE 2005-10056).

principle tasks in the sequence had to satisfy was that they were to be carried out by dyads of students working together.

## 2. THEORETICAL FRAMEWORK

### 2.1. CLIL immersion and semi-immersion programmes

CLIL Immersion and Semi-Immersion (SI) programmes which promote second language learning through *non-linguistic*<sup>4</sup> curricular areas are<sup>5</sup> endorsed by a long research tradition (Brinton et al., 1989; Cenoz et al., 2001; Crandall, 1992; Escobar & Unamuno, 2008; Gajo, 2001; Genesee, 1987; Genesee et al., 1986; Lantolf, 2000; Lüdi & Py, 2002; Mohan et al., 2001; Pérez-Vidal, 1998; Swain & Lapkin, 1982; Siguan, coord, 1985; Unamuno, 2005; Vallcorba, 2005; Vila & Vila, coord, 1998). This abundant corpus suggests that the success of this approach depends on various factors, being quantitative —greater and more intense contact with the target language— cognitive, emotional, psycholinguistic and psychosocial. Some of the goals attributed to CLIL are (Escobar, 2001):

1. To introduce the real world into classrooms.
2. To promote the use of authentic texts and materials.
3. To open the classroom to different types of interaction by promoting the communication of authentic, meaningful information in which content is prioritised over form.
4. To favour the transfer of learning processes.
5. To encourage teachers not to take the comprehension of messages for granted, and therefore, to use of facilitating strategies that aid the comprehension of curricular content.
6. To facilitate contextualised metalinguistic reflection when obstacles to understanding and production arise.
7. To promote the practice of linguistic (comprehension, production and interaction in oral and written situations of use) and discursive abilities (description, argumentation, exposition, etc.).
8. To favour positive attitudes towards genuine plurilingualism.

On the other hand, specific research into the learning of language and content, in *inclusive* classrooms, when the vehicular language is a *foreign language* with low availability in the learners' environment, is in its beginnings and many questions remain unanswered.

---

<sup>4</sup> The term «non-linguistic subjects» is clearly inaccurate as language is the most important medium for academic learning in any subject.

However, the simplicity of the term has made it succeed in CLIL literature and it is currently used by researchers and practitioners alike.

<sup>5</sup> CLIL programmes are also referred to as Content Based or Bilingual. This last term is clearly inappropriate in the context we are describing, as English is the third language of instruction.

## 5.2. Interaction among equals

Classic studies have highlighted the importance of discussion in pairs or small groups in learning in general and in foreign language learning in particular (Barnes & Todd, 1976; Fletcher, 1985; Griggs, 1997; Hoyles et al., 1990; Long, 1983; Mercer, 1995; Pica et al., 1996).

Other more recent studies framed in the sociocultural tradition deal with the impact of activities carried out in pairs on the process of multilingual learning. In this research, the way in which verbal activities exemplify the relationships between linguistic learning and the process of socialisation, as well as the crucial role of verbal interaction in the acquisition of multilingual and academic competences, have been determined (Escobar, 2000, 2002; Gajo & Mondada, 2000; Masats & Unamuno, 2001; Mondada & Pekarek, 2001, 2004; Nussbaum, 1999; Nussbaum, Tusón & Unamuno, 2002; Nussbaum & Unamuno, 2000, 2001 & 2006).

## 5.3. Tasks for language learning and CLIL

ARTICLE adopted a task based approach, following the definition of «task» formulated by Skehan (1998), as it best fit the specific characteristics of the teaching of curricular content in a foreign language. Thus, it was concluded that the type of activities we required for a Science classroom had to comply with the four criteria described by Skehan in his definition of a «task»:

1. *Meaning is primary*: This criterion is, per se, constitutive of a CLIL classroom, given that the point of departure in activity planning is the aims and academic content, and never a discreet linguistic form.
2. *There is a goal which needs to be worked towards*: each task designed should have a communicative and learning aim, easily identifiable for learners.
3. *The activity is outcome-evaluated*: This is the level of efficacy in the resolution of a particular problem, in the communication of a message to a peer or to the rest of the class, which determines the level of achievement in the task.
4. *There is a real-world relationship*: The relationship is double faceted. On the one hand, Science, as a subject, is an unavoidable reality in the school world. On the other hand, news and documentaries about climate change and rainforests as thermo-regulators are frequent in social means of communication and attract the attention of adolescents.

Moreover, it was decided that the planning of the learning activities should be based on the following axes: (a) The design of the tasks would take into account the results of studies into task types (Barnes & Todd, 1976; Long & Porter, 1985; van Lier, 2004), but the final selection prioritised the criteria of relevance and adequacy for the needs and interests of students; (b) The linguistic aims of each activity would derive from the communicative, discursive and formal needs of the specific task; (c) Awareness-raising activities and/or controlled practice of lexical items would precede the carrying out of the task, and were to be called enabling tasks; (d) Self or co-evaluation activities would follow the completion of the task, with the aim of strengthening the learners' awareness of the language used. And (e) The sequence was to include a variety of tasks, in spiral organisation, that would offer learners rich comprehensible

input and which would provide opportunities to use and reuse the language in meaningful production activities.

### 3. THE TEACHING SEQUENCE

The result of the application of these pedagogical principles was the design of a teaching sequence (Escobar, 2008) organised in three task cycles made up of main tasks, enabling tasks and follow-up tasks (figure 1)<sup>6</sup>.

**Fig. 1.**

Task Cycle	Main task	Attributed pedagogical value
Task 1	Information swap about the Amazons between “experts”, using the jigsaw technique.	Those attributed to the “Cooperative learning” pedagogical school (Kagan, 1995): active participation and cooperation between learners.
Task 2	Enquiry-based problem solving activities: formulate a proposal for sustainable development of the Amazons.	Learning through the resolution of problems promotes the use of higher order cognitive abilities (Sellwood, 1989).
Task 3	Oral academic presentation in front of the class.	Task aimed at the activation of students’ cognitive-linguistic abilities (Jorba, Gómez & Prat, 2000). It also acts as the official assessment task, which gives “school meaning” to all the work carried out throughout the sequence.

#### 5.1. The materials

The materials were designed collaboratively by university researchers and teacher researchers based on the tasks types specified in the research project. These materials may be retrieved from: <http://www.clil-si.org/>

Apart from the creation of materials for learners, a «road map» for experimenting teachers was produced, which aimed to guide and standardise teaching actions, and locate the specific moments for data collection on the timeline.

### 4. AIMS AND RESEARCH QUESTIONS

Once the design of the sequence was finalised, the evaluative research project was planned and developed. The study aimed at assessing the global grade of efficacy of the sequence in promoting the simultaneous learning of language and science. As mentioned previously, in this paper we will concentrate exclusively on linguistic learning, and more specifically on the improvement in fluency and lexical repertoire.

Fluency is usually identified as the area in which semi-immersion programmes offer the best results. For this reason, it was included as one of the indicators to evaluate. As regards

<sup>6</sup> For more information about the sequence, please refer to [www.clil-si.org](http://www.clil-si.org).

lexical repertoire, the acquisition of academic knowledge in a particular discipline is demonstrated by a certain discourse competence: one which nears the ways of speaking and writing specific to the discipline in question. Traditionally, this competence has been defined on lexical grounds.

Communicative approaches, on the one hand, and socio-cultural theories of learning, on the other, situate the individual's communicative activity as the motor of their cognitive and socio-cultural activity. This implies that research should concentrate on the observation of the construction of knowledge in and through communicative action, while the individual is participating in concrete tasks with specific interlocutors and in concrete circumstances. From this perspective, whereas lexical complexity may be considered an indicator of competence, it might not be sufficient for capturing the complexity of learning (Lewis, 1993).

In line with this perspective, we agree that the lexicon may not be a sufficient indicator of progress. Nevertheless, it may not be refuted that one of the main linguistic features characterising each area of academic knowledge is precisely its specific lexical repertoire. Therefore, the identification of improvements in lexical competences is a relevant research aim. Also, lexical repertoire has frequently been an undervalued linguistic competence in research into language learning, in favour of the acquisition of grammar (Laufer, 1986, 1991; Laufer & Nation, 1995; Meara, 1980).

Concretely, the aim of this study is to determine the possible improvement in fluency and lexical repertoire of the participating students when dealing with academic content in Natural Sciences, without them receiving specific linguistic instruction.

This aim may be spelled out in the following research questions:

1. Can improvements in students' capacity to use specific lexical items in academic reading and production activities be determined?
2. Can advances in participants' fluency when they express themselves be determined?
3. In the case that improvements in students' competences are identified, do these have any sort of repercussion for their school grades?

## **5. METHODOLOGICAL FRAMEWORK**

### **5.1. Classroom research**

ARTICLE fits into the field of classroom research (Allwright & Bailey, 1991). Therefore, it is a research project developed in the framework of the intact classroom, without random selection of subjects or a control group. This design offers the advantage of facilitating an approach with low levels of intrusiveness, and therefore, one that is ecological and ethical. The fact that it involves low levels of intrusiveness facilitates the acceptance of the experiment by school masters, as it does not interfere with the normal activities at the school. Data collection, furthermore, does not provoke rejection from the school community: teachers, families or students.

### **5.2. Measurement vs. Interpretation**

In language pedagogy, the contra-positioning of quantitative and qualitative paradigms, in relation to the predominant data treatment methods in a given investigation, is frequent (Camps,

2001). Both paradigms are often presented, simply, as antagonistic and irreconcilable. We agree with Camps that the instruments to explore «the reality of teaching and learning languages may be diverse and complementary» (Camps, 2001:16, our translation).

ARTICLE has produced a series of studies (Corredera, 2007; Escobar & Nussbaum, 2008) formulated from an ethnographic-interpretative approach which centre on processes, based on the qualitative treatment —i.e. identification and categorisation of conversational sequences— of qualitative data —i.e. transcriptions. However, it also aspires to complement this type of research with quantitative research providing data about the quantity and type of learning attributable to the teaching sequence implemented.

### 5.3. A pre-experimental quantitative study

The present research has been developed based on a pre-experimental<sup>7</sup> design of the type **0 X 0** (Brown, 1998; Hatch & Lazarton, 1991; Leedy, 1997, for example), or of pre-test – treatment – post-test, with statistical treatment of quantitative data.

The possible threats to the generalisability of the results attributed to this type of design, which lacks random selection and a control group, are offset by the context, the nature of the research and the limited timeframe in which the experiment was carried out. Figure 2 summarises the threats pointed out by Hatch and Lazarton (1991) and explains the way in which their impact is minimised or annulled in the study we present.

Fig. 2.

<b>THREATS TO THE GENERALISABILITY OF PREEXPERIMENTAL DESIGNS</b> (Adapted from Hatch & Lazarton, 1991: 86)	<b>CONDITIONS IN THE STUDY WHICH ANNUL O OR CONSIDERABLY MINIMISE THE THREAT</b>
<b>HISTORY</b> The interference of educational actions, apart from that which is under study, may cause differences between the pre- and post-test.	During the period of time that the experiment lasted, none of the participating groups studied the topic of rainforests in another subject. The four largest of the six participating groups did not receive any other instruction in English. The two remaining groups (the smallest in number of members) did not cover any of the linguistic content assessed in the experiment.
<b>MATURATION</b> Subjects may have matured between the pre- and post-test.	The short time frame (one month) between the administration of the pre- and post-test limits the possibility of maturation.
<b>MORTALITY</b> Some of the subjects disappear throughout the study, which may affect results.	Carrying out the experiment in natural classrooms caused student absences in some components of the pre-test, in some activities and in some of the components of the post-test. These are habitual classroom conditions, and therefore, were accepted as a constitutive part of the research. Students who had a same test carried out as part of the pre- and post-test were considered valid cases, regardless of whether data from other tests was incomplete, or they had been absent in classes in which the experiment was carried out.

<sup>7</sup> Research design which does not employ random selection or a control group.

**Fig. 2. (Cont.)**

<p><b>INTERNAL TEST VALIDITY</b></p> <p>Students may remember the questions in the pre-test when they answer the post-test and therefore improve their results.</p>	<p>The composition and oral production exercises are both open-ended tests. There are not, therefore, any questions to remember.</p> <p>The problem could affect the exercise dealing with the recognition of specific lexical items. The fact that the exercise was made up of 25 items, and one month passed between the first and second administration, decreases considerably the probability that recall of the questions of the pre-test was the determinant variable affecting the answers on the post-test. Further on, it is demonstrated how the statistical results confirm this interpretation.</p>
---	--

## 6. DATA COLLECTION

### 6.1. The sample

The experimentation of the sequence took place between February and June 2005, with a sample made up of six natural, intact compulsory secondary education classrooms in three public high schools in Catalonia, involving four teachers and 127 secondary-school students.

The schools were located in zones with very different social backgrounds and which receive different proportions of students of immigrant origin. The selected classrooms were those taught by teachers participating in the experiment, and in no case was there a selection or exclusion of students.

In the 3 schools, the experiment took place in the period for usual English classes (four groups) or in elective subjects of Science in English that lasted a term (two groups). In all cases, the content covered in the compulsory Science subject was different from that taught in the experiment.

Figure 3 summarises the level of participation of teachers and students in the experiment<sup>8</sup>.

**Fig. 3.**

School	Location	Compulsory secondary education groups	Students	Teachers
TARRAGONA	Tarragona	Two standard 3rd year English classrooms	50	1
LLOBREGAT	Industrial periphery of Barcelona	Two standard 4th year English classrooms	51	1
BARCELONA	Barcelona residential district	Two 1st year elective Science in English classrooms	26	2

### 6.2. Pre-test / Post-test device

A device for the collection of data was created, consisting of the collection of pre-treatment (T1) and post-treatment (T2) data with three different tests, administered in the following order:

<sup>8</sup> In all cases the real names of schools and participants have been replaced by pseudonyms.

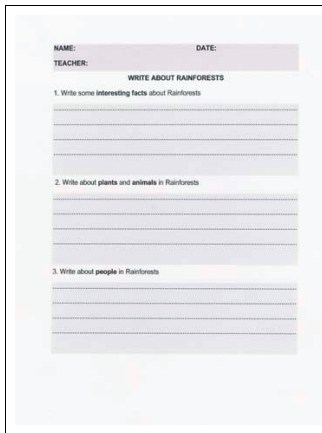


0. Slide show:

Before soliciting any productions, participants were shown a set of 20 slides with images related to tropical rainforests and the Amazons. The stimuli were intended to activate students' prior knowledge about the content. The showing of the slides was carried out in silence, without giving any verbal information in the L2 or L1.

1. TEST 1. Composition: students were requested to produce a written reply to three open-ended questions. The students answered the questions directly in a word processor (figure 4).

Fig. 4.

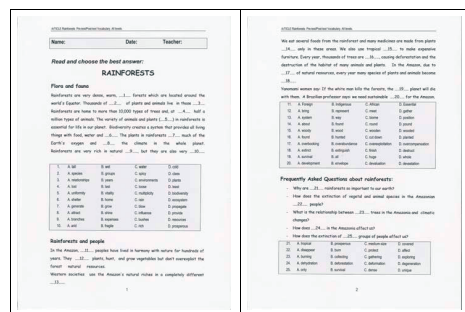


2. TEST 2: Oral monological explanation<sup>9</sup> (Figure 5).
3. TEST 3: Specific lexical recognition test of 25 multiple choice items (Figure 6).

Fig. 5.



Fig. 6.



<sup>9</sup> Oral data was collected from a sub-sample of 52 students. The data obtained in exercise 2 will not be explored in the present paper. This information has been included here in order to adequately contextualise the data collection process. The oral data is in the process of being transcribed and analysed.

The post-test was carried out immediately after finishing the teaching sequence, following the process described for the pre-test point by point, with the exception of the omission of the slideshow prior to doing the exercises.

4. Official school grades in English and Science corresponding to the evaluations immediately prior to and following the treatment.

### 6.3. Valid collected data

Of the 127 students who participated in this study, the following valid data was obtained for the analysis (both pre- and post-treatment data was available and obtained in equal conditions):

#### Basic data:

- Pre-post test written data (compositions) of 73 secondary-school students.
- Pre-post test oral data (monological speech). Sub-sample of 24 secondary-school students.
- Pre-post test of specific lexical recognition (multiple choice test) of 104 secondary-school students.

#### Complementary data:

- Students' school grades in English and Natural Sciences corresponding to the evaluations immediately prior to and following the treatment.

## 7. ANALYSIS

In this section, we will discuss in detail the analysis of the quantitative data obtained with the pre-test – treatment – post-test device, corresponding to the lexical recognition test, composition and school grades in English and Natural Sciences.

As indicators of student progress, a set of measures referring to lexical repertoire and fluency were used. In the case of lexical repertoire, two indicators (2 and 3, below) were selected that provide «intrinsic» information, given that the referential element in the evaluation is always the learner's own repertoire. Two measures (1 and 4, below) were also selected to compare the results obtained by the learner with a corpus elaborated according to external criteria. That is, they provide «extrinsic» information.

### 7.1. Indicators referring to the lexical repertoire

The **multiple-choice** specific lexical recognition test produced **Indicator 1**. The tests were scored and the marks transformed onto a decimal scale. The result is an extrinsic measure of the students' lexical competence.

Indicators 2, 3 and 4 were extracted from the students' **written production**:

**Indicator 2** is an index of lexical richness which indicates the ratio between the number of different words used by the learner and the total number of words produced. This index has been labelled by different authors as Density, Lexical Variation or Type-Token Ratio

(Laufer, 1991; Wolfe-Quintero et al. , 1998). The tool used to determine this *intrinsic* marker was «Advanced Text Analyser»<sup>10</sup> or ATA (see section 7.3. below).

**Indicator 3** corresponds to the total number of unique words that appear in the text and provides a measure of the student's productive lexical capital for the topic written about. This *intrinsic* type of information was calculated with ATA.

**Indicator 4** corresponds to lexical sophistication or the percentage of «sophisticated» terms that the learner uses out of the total number of words produced (Meara & Bell, 2001). Lexical sophistication was chosen as it adds information of an *extrinsic* nature. It was calculated by P-lex.

## 7.2. Fluency indicators

The total number of words (**indicator 5**) and the total number of sentences (**indicator 6**) per text produced by the informants were computed by ATA. Both of them are *intrinsic* measures.

## 7.3. Computer-based tools

As mentioned above, the computer-based tools used to extract and process the indicators were the following:

1. Advanced Text Analyser (ATA) is an analysis tool for text content that counts and classifies words from a text according to different criteria, without taking any external texts into consideration. The ATA tool also provides the total number of words produced by learners and the total number of sentences.
2. P\_lex (Meara & Bell, 2001) is a computer program that calculates the index of lexical sophistication of a text based on the proportion of «easy» and «difficult» words it contains. In order to do so, P\_lex classifies the words into four categories, according to the frequency of use of such words.
3. The data obtained by means of ATA and P\_lex were processed using the statistical programme SPSS version 11.

## 7.4. Results obtained from T-test

A difference of means test was applied to the figures obtained through the processes described above in order to disprove (or confirm) the null hypothesis (Hatch & Lazarson, 1991):

**H<sub>0</sub>: The progress in evaluated linguistic abilities is null.**

<sup>10</sup> Advanced Text Analyser is available at: <http://usingenglish.com/resources/text-statistics.php>; and P-lex is available at: <http://www.swan.ac.uk/cals/calsres/lognostics.htm>

Figure 7 brings together the main results obtained for these indicators.

Fig. 7.

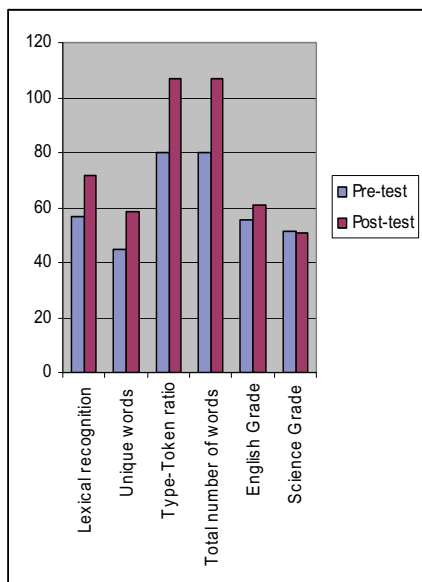
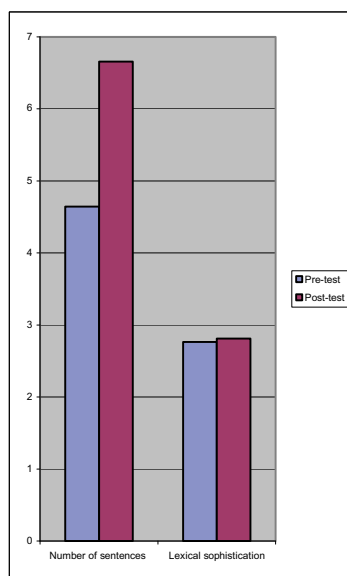
	1	2	3	4
	N	T-test	Improvement	Significance level
Lexical recognition	104	+1.5269	26.951 %	.000
Total number of unique words	73	+13.1918	29.2705%	.000
Type-token ratio	73	-.16915	-0.2746%	.223
Number of sentences per text produced	73	+2.0137	43.14%	.001
Total number of words produced by informant	73	+27.0411	33.9174%	.000
Lexical sophistication (P_lex)	70	+0.0446	1.6131%	.761
English grade	126	+.2698	9.7408%	.003
Science grade	127	-.0157	-0.6135%	.851

**Column 1** refers to the number of valid subjects used to calculate the statistic for each of the indicators.

**Column 2** expresses numerically the difference in results obtained between the two tests, according to the formula: Improvement (I) = Post-test result – Pre-test result.

**Column 3** shows the percentage of improvement in each indicator. The data in this column show that the students improved on 4 indicators. The most remarkable advance was produced for the two fluency indicators, with a 43% improvement in the number of sentences produced per student. Improvements in lexical richness, both in the recognition of terms (27%) and in the total number of unique words (29%), were also notable. Progress demonstrated by the two fluency indicators (27% and 29 %) was similar, which gives consistency to the results of each category. A slight but interesting advance of 10% was also observed for the grades obtained by students in the English subject.

In the above cases the results are (**column 4**) statistically significant (.000 to .003), indicating a statistical reliability in the results of over 99.7%. The results are hardly conclusive for the remainder of indicators. The slight improvement in lexical sophistication (I=1.6131; s=.761) and the worsening in Science grades (I=0.6135; s=.851) and lexical variation (I=0.2746; s=.223) are only very slightly, if at all, statistically significant.

**Fig. 8:** Indicators with a range higher than 10.**Fig. 9:** Indicators with a range below 10.

## 8. DISCUSSION

### 8.1. On school grades

The lack of progress in students' grades in the subject of Science was consistent with our expectations, given the characteristics of the students evaluated: as mentioned earlier (fig.3), the experiment took place in regular English class time (4 groups) or in the term-long elective Science in English (2 groups) subject. In all cases, the content dealt with in the compulsory Science subjects was different from that covered in this experiment. In such circumstances and given the short time span, it is improbable that a significant change be produced in student grades in their regular Science classes<sup>11</sup>.

On the other hand, the above result gives special relevance to the positive evolution of about 10% in the grades obtained by students in their regular English classes. Various hypotheses could explain this fact:

- a) Students, in fact, improved their linguistic competences to the degree indicated by their teachers. This improvement may derive exclusively from the properties of the treatment, or in conjunction with other factors such as:
  - The fact that the experimentation process involved systematic self- and other-observation may have provoked greater awareness among participating teachers of their students' capacities, that may have previously gone unnoticed. And / or ...

<sup>11</sup> It may be suggested that the Natural Science grades have somewhat acted as a control variable.

- External observation may have influenced students' commitment to the tasks. In both cases the observer's paradox (Labov, 1972) would be partly responsible for the improvement in grades.
- b) The high grade of implication on behalf of teachers in the experimentation and the desire of all good teachers that their students «put on a good performance» before external observers may have provoked certain doses of self-fulfilling prophecies or the Pygmalion effect (Strom, 1971).

## 8.2. On fluency

The notable increase in the number of words produced by students, and the even more notable improvement in the number of sentences per text produced, suggest that the activities carried out had a significant positive effect on the improvement of students' fluency, in the context of Science academic learning in the foreign language. This conclusion is, in itself, positive and encourages us to continue our research into this approach.

## 8.3. On lexical repertoire

The increase in the two indicators measuring lexical richness – 27% for specific lexical recognition and 29% for the total number of unique words – is very similar. This consistency among both values evidently points to clear progress in the students' lexical repertoire, at the same time as it removes importance from the doubts about a possible lack of internal validity for the lexical recognition test attributable to possible recall during the post-test of the questions already formulated on the pre-test. This (fig. 2) is precisely one of the weaknesses ascribed to pre-experimental designs, based on the lack of a control group.

On the other hand, we interpret that the low grade of statistical significance for the other two indicators of lexical richness (lexical variation and sophistication) demonstrate that such indicators have not functioned well with the target population and hardly seem adequate to measure the progress of this collective. They have therefore been removed from the evaluative study of task efficacy.

The poor functioning of the index of lexical density or variation (type-token ratio) may have been simultaneously influenced by two factors. Firstly, the measure is highly sensitive to the length of texts, given that the proportion of new words that appear in a text tends to decrease as the length of the text increases (Laufer, 1991, Malvern & Richards, 1997; Miralpeix, 2002; Richards, 1987; Wachal & Spreen, 1973; Wolfe-Quintero et. al., 1998). This is due to the fact that as the length of the text increases, items in closed grammatical classes (articles, pronouns, prepositions, etc.) tend to be repeated, meaning that variation is lessened. The ostensible increase in the length of the texts in the post-test compared to the pre-test therefore affects this indicator, whose value consequently decreases, showing lower lexical mastery on the post-test. A possible solution to offset this problem would be to compare texts of the same length. However, this solution would only make sense if the only cause of the poor functioning of the measure was text length.

Unfortunately, we have also detected a second factor which undesirably influenced the index of lexical density or variation, and which we had not identified until now in the bibliography consulted: the grade of grammaticalisation of the texts obtained in the post-tests is obviously greater than in the pre-tests. This higher grade of grammaticalisation is the cause of the

percentual increase in the repetition of articles, auxiliary verbs, prepositions and other closed-class items, even if we compare texts of the same length.

Indeed, a strategy often observed in language learners in beginning levels is that of writing their messages in lists of content words, none or very few of which are repeated. Productions with a higher level of grammaticalisation necessitate the repetition of those words which express syntactic relationships between content words. Figure 10 shows an example of two texts with a different grade of grammaticalisation, which mean to express the same message.

**Fig. 10.**

Low grammaticalisation production	High grammaticalisation production
<i>People live rainforest are savage and poor. Children have ugly: piercings, tattoos<sup>12</sup>.</i>	<i>The people who live in the rainforest are savage and poor. The children are ugly. They have piercings and tattoos.</i>
Length: 12 words; Different words: 12/12 (all of them).	Length: 20 words; Different words: 16/20.
<b>Type-token ratio = 1</b>	<b>Type-token ratio = 0.8</b> Different words in the first 12 words: 10/12. <b>Type-token ratio = 0.83</b>

The example clearly illustrates that the length of the text affects the ratio between the total number of different words and total number of words produced. The index of grammaticalisation also affects this ratio, as observed in the total number of different words in the above texts, when the total number of words is kept the same.

The progress in lexical sophistication as measured by P\_lex has little statistical significance (.761) and therefore, no conclusions on learning can be extracted from that figure. Our hypothesis is that this is due to the inadequacy of P\_lex for capturing lexical sophistication in the ARTICLE corpus, formed by academic texts with a high proportion of specialised terminology (Escobar, 2007).

## 9. CONCLUSION

With all of the precautions required by a pre-experimental design in mind, and despite the disappointing functioning of some of the indicators, the results of the data analysis presented in this paper support our optimism for the adopted approach. Students' progress in terms of fluency and lexical repertoire is evident. The results are especially striking if we consider the students' evolution (between 10 and 43% according to the indicators) over only one month (between 8 and 9 class periods); students who had previously studied more than 600 hours of English. These provisional results invite us to continue designing and piloting materials along similar lines, and to design and execute research projects of a greater scope.

<sup>12</sup> Taking one student's initial production. The second text has been elaborated based on the first to illustrate the point we make. Real post-tests also show the evolution in the students' understanding of facts and the subsequent choice of words. We will not go into this point here.

The analysis has also highlighted the need for more precise measures, better adapted to the type of students and pedagogical treatment targeted in this study, and covering a greater field of text characteristics. Continuing in line with the hypothesis presented at the beginning of this paper, and after identifying the strengths and weaknesses of the system of indicators used, it seems appropriate to carry on with the comparative study of pre- and post-tests through the exploration of the morphosyntactic and rhetorical characteristics of the texts to complement the study carried out until now. In this theoretical line of work, we have already begun to identify indicators capable of capturing the changes observed between the pre- and post-test.

The generalisation of CLIL programmes is a major shift in European education, which, like any great innovation in education, may involve significant risks. Enthusiasm for the unavoidable promotion of plurilingualism among citizens can not leave aside rigorous research into methodological options adopted, or their results. It is the responsibility of universities, and not only of public administrations, to carry out conclusive research to check whether the adopted pedagogical options produce the desired results, and to understand the reasons why such results are obtained or not. We plan to continue exploring this area of research.

## 10. REFERENCES

- Allwright, R.L. & K.M. Bailey (1991). *Focus on the language classroom: an introduction to classroom research for language teachers*. Cambridge: Cambridge University Press.
- Barnes, D. & Todd, F. (1976). *Communication and learning in small groups*. Londres: Routledge & Kegan Paul.
- Brinton, D.M., Snow, M.A., & Wesche, M.B. (1989). *Content-based second language instruction*. New York: Harper & Row.
- Brown, H.D. (1998). *Understanding research in second language learning*. Cambridge: Cambridge University Press.
- Camps, A. (Coord.). (2001). *El aula como espacio de investigación y reflexión. Investigaciones en Didáctica de la Lengua*. Graó: Barcelona.
- Cenoz, J. et al. (Eds.). (2001). *Cross-linguistic influence in third language acquisition*. Clevedon: Multilingual Matters.
- Corredera, A. (2008). «Uso de la lengua extranjera en tareas para aprender ciencias en inglés: ¿hablan en inglés?» En Monroy, R. y Sánchez, A. (eds.) *25 Años de Lingüística Aplicada en España: Hitos y Retos / 25 Years of Applied Linguistics in Spain: Milestones and Challenges*. Murcia: Servicio de Publicaciones de la Universidad de Murcia, 295-307.
- Council of Europe [COE] (1995). *White Paper on Education and Training*. Available at: <http://europa.eu.int/en/record/white/edu9511>.
- Crandall, J. (1992). «Content-centered instruction in the United States». In: *Annual Review of Applied Linguistics 13*: 111-126.
- Escobar Urmeneta, C. (2000). *El portafolio oral como instrumento de evaluación formativa en el aula de lengua extranjera*. Tesis doctoral UAB. Available at: [www.clil-si.org](http://www.clil-si.org).
- Escobar Urmeneta, C. (2001). «Los programas de aprendizaje integrado de contenidos curriculares y lengua». In: Nussbaum, L. & Bernaus, M. (eds.): *Didáctica de las lenguas extranjeras en la educación secundaria obligatoria*. Madrid: Síntesis, 229-232.
- Escobar Urmeneta, C. (2002). «Attention and the processing of comprehensible input in communication tasks among secondary school learners». In: Nizgorodcew, A. (ed.). *Beyond L2 Teaching. Research studies in second language acquisition*. Krakow: Jagiellonian University Press.



- Escobar Urmeneta, C. (2007). «Mesurament de competències lingüístiques en aules AICLE, en contextos multilingües: allò que els números amaguen.». In: Guasch, O. & Nussbaum, L. (eds) *Aproximacions a la noció de competència multilingüe*. Bellaterra: Servei de Publicacions de la UAB, 133-146.
- Escobar Urmeneta, C. (2008). «Talking English to learn Science. A CLIL experience in Barcelona». In: Dooly, M. (ed.) (2008) *How we're going about it. Teachers' voices on innovative approaches to teaching and learning languages*. Cambridge: Cambridge Scholar Press, 154-169.
- Escobar Urmeneta, C. y Nussbaum, L. (2008). «Tareas de intercambio de información y procesos de aprendizaje en el aula AICLE.». In: Camps y Milian: (coord.) *Miradas y voces. Investigaciones sobre la educación lingüística y literaria en entornos plurilingües*. Barcelona: Graó, 159-178. Catalan translation: «Tasques d'intercanvi de informació i processos d'aprenentatge en l'aula AICLE». In: Camps & Milian: (coord.) *Mirades i veus. Investigacions sobre l'educació lingüística i literària en entorns plurilingües*. Barcelona: Editorial Graó, 167-187.
- Escobar Urmeneta, C. & Unamuno, V. (2008). «Languages and language learning in Catalan Schools: From the bilingual to the multilingual challenge». In: Hélot, C. & de Mejia, AM (eds.): *Forging Multilingual Spaces: Integrating Majority and Minority Bilingual Education*. Clevedon: Multilingual Matters. (Forthcoming).
- Fletcher, B. (1985). «Group and individual learning of junior school children on a micro-computer-based task». In: *Educational review* 37: 251-261.
- Gajo, L. (2001). *Immersion, bilinguisme et interaction en classe*. Paris: Hatier.
- Gajo, L. & Mondada, L. (2000). *Interactions en contexte*. Fribourg: Presses Universitaires Fribourg Suisse.
- Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Cambridge, MA: Newbury House.
- Genesee, F. Lambert, W.E. & Holobow, N.E. (1986). «Adquisición de la segunda lengua mediante inmersión: el enfoque canadiense». In: *Infancia y aprendizaje* 33: 27-36.
- Griggs, P. (1997). «Metalinguistic Work and the Development of Language Use in Communicative Pairwork Activities Involving Second Language Learners». In: *Eurosla 7. Proceedings*, 403-415.
- Hatch, L. & Lazarson, A. (1991). *The Research Manual*. Los Angeles: Newbury House.
- Hoyle, C., Sutherland, R. & Healy, I. (1990). «Children talking in computer environments: new insights on the role of discussion in mathematics learning». In: Durkin & Shine (ed.). *Language and mathematics in education*. Milton Keynes: Open University Press.
- Jorba, J., Gómez, I. & Prat, A. (eds). (2000). *Hablar & escribir para aprender. Uso de la lengua en situaciones de enseñanza-aprendizaje desde las áreas curriculares*. Madrid: Síntesis.
- Kagan, S. (1995). «Dimensions of Cooperative Classroom Structures». In: Slavin & al. (eds.): *Learning to Cooperate, Cooperate to Learn*. New York: Plenum Press, 67-96.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lantolf, J. P. (ed.). (2000). *Sociocultural theory and second language learning*. Oxford: Oxford University Press.
- Laufer, B. (1986). «Possible changes in attitude towards vocabulary acquisition research». In: *IRAL XXIV*(1): 69-75.
- Laufer, B. (1991). «The development of L2 lexis in the expression of the advanced learner». *The Modern Language Journal* 75 (4): 440-448.

- Laufer, B. & Nation, P. (1995). «Vocabulary size and use: lexical richness in L2 written production». In: *Applied Linguistics* 16-3: 307-323.
- Leedy, P.D. (1997). *Practical research. Planning and design*. Upper Sadler River: Merrill-Prentice Hall (6th edition).
- Lewis, M. (1993). *The Lexical Approach*. LTP. Hove: England.
- Long, M. H. (1983). «Native speaker/non-native speaker conversation and the negotiation of comprehensible input». In: *Applied Linguistics* 4/2: 127-141.
- Long, M., & Porter, P. (1985). «Group work, interlanguage talk, and second language acquisition». *TESOL Quarterly* 19: 207-28.
- Lüdi, G. & Py, B. (2002). *Être bilingüe*. Berna: Perter Lang (second edition, revised).
- Malvern, D. & Richards, B. (1997). «A new measure of lexical diversity». In: L. Eubank, L. Selinker & M. Sharwood Smith (eds.): *Evolving models of language*. Clevedon, Multilingual Matters: 58-71.
- Masats, D. & Unamuno, V. (2001). «Constructing Social Identities and Discourse through Repair Activities». In Forter-Cohen, S. & Nizegorodcew, A. (eds.), *Eurosla Yearbook 2000*. Amsterdam: Benjamins.
- Meara, P. (1980). «Vocabulary acquisition: A neglected aspect of language learning». In: *Language teaching and linguistics: Abstracts*. 13: 221-46.
- Meara, P. & Bell, H. (2001). «P\_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts». In: *Prospect* 16(3): 5-19.
- Mercer, N. (1995). *The guided construction of knowledge*. Clevedon, Avon: Multilingual Matters. Spanish translation: *La construcción guiada del conocimiento*. Barcelona: Paidós, (1997).
- Miralpeix, I. (2002). *The use of P\_lex to assess lexical richness in compositions written by learners of English as an L3*. Treball de recerca DEA. Universitat de Barcelona. Unpublished.
- Mohan, B. et al. (2001). *English as a second language in the mainstream: teaching, learning and identity*. New York: Longman.
- Mondada, L. & Pekarek, S. (2001). «Interactions acquisitionnelles en contexte». In : *Le français dans le monde, n° especial*: 107-137.
- Mondada, L. & Pekarek, S. (2004). «Second Language Acquisition as Situated Practice: Task Accomplishment in the French Second Language Classroom». In: *The Modern Language Journal* 88, iv: 501-518.
- Nussbaum, L. Tusón, A. & Unamuno, V. (2002). «Procédures d'évitement des difficultés de langue dans l'interaction entre apprenants». In: Cicurel, F & Véronique, D. (eds.). *Discours, action et appropriation des langues*. Paris: Presses de la Sorbonne Nouvelle, 147-162.
- Nussbaum, L. & Unamuno, V. (2000). «Fluidité et complexité dans la construction du discours entre apprenants de langues étrangères.». In: *Aile*, 12: 27-49.
- Nussbaum, L. & Unamuno, V. (2001). «Sociolinguistique de la communication entre apprenants.». In: Castelotti, V. (ed.): *De une langue à de autres*. Rouen: Presses Universitaires de Rouen, 59-80.
- Nussbaum, L. & Unamuno, V. (2006). *Usos i competències multilingües d'escolars d'origen immigrant*. Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona.
- Nussbaum, L. (1999). «Emergence de la conscience linguistique en travail de groupe entre apprenants de langue étrangère.». In: *Langages*, 134: 35-50.
- Pérez-Vidal, C. (1998). «Towards Multilingualism and Content Language Integrated Learning in Spain». In: Marsh, Marsland and Ralfers (eds.) *Future Scenarios in Content and Language Integrated Learning*. Jyväskylä: University of Jyväskylä, 54-61.

- Pica, T. et al., (1996). «Language Learners Interaction: How Does it Address the Input, Output, and Feed-back Needs of L2 Learners?». In *TESOL Quarterly*, 30/1: 59 – 84.
- Richards, B. (1987). «Type/token ratios: what do they really tell us?». *Journal of Child Language* 14: 201-209.
- Sellwood, P. (1989). «The role of problem solving in developing thinking skills». *The Technology Teacher*, 49(3): 3-10.
- Siguan, M. (coord) (1985): *Enseñanza en dos lenguas y resultados escolares*, Barcelona: ICE-Publicacions de la Universitat de Barcelona.
- Skehan, P. (1998). Task-based Instruction. In: *Annual Review of Applied Linguistics*, 18: 268-286.
- Strom, R. (1971). *Teachers and the Learning Process*. Englewood Cliffs: Prentice-Hall.
- Swain, M & Lapkin, S. (1982). «Evaluating Bilingual Education.». In: *Canadian Case study*. Clevedon: Multilingual Matters.
- Unamuno, V. (2005): *L'entorn sociolingüístic i la construcció dels repertoris lingüístics de l'alumnat immigrat a Catalunya*. Noves SL, Primavera.
- Vallcorba, J. (2005). *La situació de la llengua als centres educatius: accions per consolidar-la i potenciar-la*. Paper presented at Jornades de llengua i ensenyament. Barcelona, October.
- van Lier, L. (2004) *The Ecology of Language Learning*. Workshop at the UAB. March 8-16, 2004.
- Vila Moreno F.X. & Vila Mendiburu, I. (coord). (1998): *Bilingüisme i educació*, Barcelona: Proa-UOC.
- Wachal, R.S. & Spreen, O. (1973). «Some measures of lexical density in aphasic and normal language performance.». In: *Language and Speech* 16: 169-181.
- Wolfe-Quintero, K. et al. (1998). *Second Language development in writing: Measures of fluency, accuracy and complexity*. Hawaii: University of Hawaii at Manoa.