Vocabulary learning at primary school: a comparison of EFL and CLIL

Elsa Tragant\textsuperscript{a}\textsuperscript{*}, Anna Marsol\textsuperscript{a}, Raquel Serrano\textsuperscript{a} and Àngels Llanes\textsuperscript{b}

\textsuperscript{a}Dept. Filologia Anglesa i Alemanya, Universitat de Barcelona, Barcelona, Spain; \textsuperscript{b}Dept. d’Anglès i Lingüística, Universitat de Lleida, Lleida, Spain

(Received 21 November 2014; accepted 11 March 2015)

Comparative studies in content and language integrated learning (CLIL) often show CLIL students to be at something of an advantage over their non-CLIL peers. However, such studies are often difficult to interpret given problems of cross-group comparability (different schooling systems, different number of instructional hours, bias attributable to selection/self-selection, etc.). This study focuses on a single group of schoolchildren (\(n = 22\)), aged eight years old, that were exposed to English as a foreign language (EFL) instruction in the fall term and to CLIL instruction (Science) in the winter term. The main objectives are to analyze the vocabulary of the class materials and to examine gains in productive lexical knowledge. Our results show that students were exposed to a greater number of words and to more abstract and technical vocabulary in the CLIL materials, but that they made significant progress in vocabulary learning in both contexts. The study also reveals that learning English through Science proved to be a more challenging experience than learning English in the EFL class.

Keywords: CLIL; EFL; second language acquisition; vocabulary; primary school; textbooks

Learning vocabulary is acknowledged to be an important building block in the acquisition of a language, and it is also reported as being closely related to enhanced language proficiency. In studies of children learning a new language in a classroom setting, the analysis of vocabulary learning is crucial since acquiring vocabulary is traditionally one of the main objectives for teachers and materials developers alike in the early phases of second language acquisition. This paper reports a small-scale study of vocabulary learning by a group of primary schoolchildren learning English in a traditional English as a foreign language (EFL) classroom and in a Science class taught in English based on a content and language integrated learning (CLIL) methodology. The study first offers an analysis of the EFL and CLIL materials used in the two classes before presenting a comparison of the vocabulary gains associated with these materials.

Background

Before 1980, European schools that offered the possibility of learning content through a second/foreign language were limited to a minority of elite institutions, but since the

\textsuperscript{*}Corresponding author. Email: tragant@ub.edu

© 2015 Taylor & Francis
Vocabulary and CLIL materials

In spite of the widespread adoption of CLIL in primary and secondary schools throughout Europe, in the early years there was a dearth of specifically designed materials and textbooks (Varkuti 2010; Wegner 2012). Teachers had to either create their own or adapt materials designed for native speaking students. This was far from an ideal situation given that it is recommended that CLIL materials include linguistic adaptations combined with a wealth of layout features, including continuous text, tables and diagrams, bullet points, and visuals (Coyle, Hood, and Marsh, 2010).

Over the past few years, however, the picture has changed somewhat. Today a number of leading ELT publishers print their own textbooks on a variety of subjects, both for primary and secondary schools. Some also produce supplementary materials (i.e., dictionaries, graded readers, teachers’ resources). Given their very recent date of publication, these CLIL materials have, in the main, yet to be evaluated. However, in a study of the CLIL activities included in ELT course books, Banegas (2014) found a tendency toward the oversimplification of contents and a predominance of reading activities and lower level thinking tasks, often related to vocabulary activities. In a comparison of an EFL and a CLIL course book, Schenk (2013) found more vocabulary repetition in the CLIL textbook, although the relative proportion of words in the first frequency band was similar in both.

With the objective of contributing to this line of research, this study compares the vocabulary presented in a CLIL textbook with that presented in an EFL class book. Previous studies of vocabulary in ELT materials can be divided in three groups: those that examine the role vocabulary plays in the syllabus (e.g., O’Dell 1997), those that analyze the aspects of vocabulary knowledge that textbooks give attention to (e.g., Brown 2010), and those that seek to quantify the amount of vocabulary in textbooks (e.g., Jiménez Catalán and Mancebo Francisco 2008). The present study adheres to this last perspective and sets out to describe and quantify the amount of vocabulary schoolchildren are exposed to in a CLIL and EFL course book that target the same student population.

Vocabulary learning and CLIL

Second language learners participating in CLIL programs are reported to benefit in many ways. A European Commission e-newsletter (2013) lists as many as 10 advantages ranging from attitudinal benefits, greater exposure to different classroom practices and perspectives, gains in intercultural knowledge and communication, in addition to enhanced language competence and oral communication skills. According to Dalton-Puffer (2009), CLIL is especially beneficial for the development of linguistic competence, above all lexical development, since students’ L2 production abilities are likely to be stretched further than in a regular EFL class.

A number of empirical studies, conducted from within a range of different research traditions, confirm the lexical gains of CLIL students in Europe’s schools. Some even
provide evidence of significant development in the receptive vocabulary skills of preschool children (Buyl and Housen 2014). Adopting the perspective afforded by discourse analysis, Dalton-Puffer (2007) has shown how CLIL students often notice lexical gaps in their own oral output and are able to self-initiate repair, something that is less frequent in typical EFL students. Ruiz de Zarobe’s (2010) work on written production indicates that after the second year, a group of secondary school students enrolled in an intensive CLIL program obtained better vocabulary scores than those recorded by students in the non-CLIL and in the less-intensive CLIL programs. Quantitative analyses of written production in upper secondary education (Jexenflicker and Dalton-Puffer 2010) also show that CLIL students tend to make greater use of less-frequent words than their non-CLIL counterparts and that the former show more lexical variation. Lexical variation in essays was also found to be higher among primary school CLIL students in Agustín Llach and Jiménez Catalán’s study (2007). A different approach to vocabulary research in CLIL is the use of discrete-point vocabulary tests. Within this framework, Jiménez Catalán and Ruiz de Zarobe (2009) found that the difference in favor of the CLIL group was more evident in the case of the most frequent 2000-word test than in that of the 1000-word test among students aged 11–12. Sylvén (2004) and Seregély (2008) also found that CLIL learners outperformed traditional learners in a battery of vocabulary tests. However, and in spite of these positive findings, evidence about the advantages of CLIL have generally been more conclusive in the case of receptive rather than of productive knowledge and skills (Dalton-Puffer 2007; Fusté and Miralpeix 2013; Heras and Lagasabaster 2015). The present study opts to examine specifically the productive knowledge schoolchildren gain when learning vocabulary through CLIL instruction.

This study has, moreover, been specifically designed to avoid one of the most salient problems in existing research, namely the comparability of CLIL and non-CLIL groups, especially when the number of schools involved is small (see Aguilar and Muñoz 2014, and Bruton 2011, for a discussion of other problematic aspects of CLIL research). This is the case of studies in which the two groups are from different cities and/or belong to different school systems as occurs, for example, in Moghadam and Fatemipour (2014). The limitation is often attributable to the fact that the CLIL groups receive considerably more hours of instruction, a frequently unavoidable confounding variable, as in Jiménez Catalán and Ruiz de Zarobe (2009). Another factor that also makes comparisons less valid is when CLIL students are selected or self-selected. In fact, this was the case in Heras and Lagasabaster’s study (2015) of a school in which entering the CLIL program was voluntary. The lack of group comparability is, in general, a major challenge in L2 classroom research, but it is especially difficult – if not impossible – to avoid in the context of CLIL instruction. Aware of these challenges, the opportunity was presented to us of conducting a study of vocabulary gains in which the question of comparability is not a problem since the same group of students underwent a first period of EFL instruction followed by a period of CLIL instruction.

Aims of the study and research questions
The present study focuses on a group of primary school students who followed a program that concentrated their regular English instruction in the fall term and their CLIL instruction in the winter term. The main objectives of the study are to analyze the vocabulary profiles of the instructional materials used in the two teaching contexts and to
compare the vocabulary learning resulting from the two instructional periods. More specifically, we aim to answer the following questions:

(1) How much and what type of target vocabulary are students exposed to in their textbooks during a first term of EFL instruction and a second term of CLIL instruction?
(2) Do students experience any gains in their productive knowledge of the target vocabulary between the start and end of each term?
(3) Do students make more vocabulary gains in one or other of the two programs: EFL vs. CLIL?

In addressing the three research questions above, we focus on the key vocabulary appearing in the teacher’s manuals of the textbooks under analysis. From this point on this vocabulary is referred to as the ‘target vocabulary’ and comprises the words that students are expected to learn through direct instruction.

Methodology and context
The school and participants
The school is a prestigious boys’ school located in a residential area of a medium-sized city in Catalonia (Spain). It is a small school (one class per grade) with spacious facilities and one of its distinctive features is the importance attached to English by its administrators. For example, in the third grade (the target grade in this study), four hours a week are usually devoted to English instruction (the official requirement in Catalonia being two hours per week). In addition, a further three hours a week are normally devoted to the teaching of Science in English in the third grade, which is when CLIL instruction is introduced in this school. For the purposes of this study, however, the school was requested to modify this program and to concentrate English language instruction into the fall term (seven one-hour periods per week) and the instruction of Science in English into the winter term (seven one-hour periods per week). This new program allows us to undertake an independent examination of vocabulary learning in EFL and CLIL classes. In both contexts students worked with a class book/textbook and an activity book, and lessons took place in their regular classroom, except for one lesson each week when the class was held in the school’s computer room.

The participants in the study were 22 boys aged between 8 and 9 (third year in primary education in Spain) and their English/Science teacher. One student was excluded from the study because of serious learning difficulties and, on analyzing the data, he was found to be an outlier. With the exception of one of the fathers and two of the mothers, the rest of the parents have university degrees and 10 of the children study English as an after-school activity. Even though almost all of the children (20 out of the 22) report English to be an important language to learn, when asked to name their three favorite subjects, English was mentioned by just three of the boys. These children are taught English and Science in English by the same teacher, an experienced male teacher with a good command of English and effective management skills. He is a certified English primary schoolteacher who also has a BA degree in English Studies.

Instructional materials and practices
The EFL lessons were based on a class and activity book called Incredible English Kit 3 (Phillips and Morgan 2007). The class book includes a variety of materials: speaking and reading/listening practice, songs and vocabulary, and pronunciation and covert grammar
activities. Vocabulary in this class book is formally introduced on the first page of each unit accompanied by a large illustration, but it is not typologically enhanced in any way on the remaining pages of the unit. In the vocabulary exercises in the activity book, words are printed on a gray background.

Most EFL classes started as a whole class session using the class book, while the last part of the class was often devoted to individual work in the activity book. One session a week took place in the computer room, where students were asked to prepare PowerPoint presentations to illustrate the vocabulary of the unit they were then studying. The teacher covered most of the materials from the class book and the boys were usually asked to complete all the activities in the activity book. At times the teacher was also observed to provide unplanned explanations outside the textbook (i.e., similarities between L1 and L2 words, cross-cultural differences, etc.). During these lessons the teacher almost always addressed the children in English and, when asking them questions, he expected them to answer in English.

The Science lessons were based on a text and activity book called Top Science 3 (Santillana 2011). In the textbook, information is presented in texts and via numerous photographs and drawings, which are complemented with questions, hands-on activities, and written exercises. In this textbook, target words are highlighted in bold in the body of the texts in which content is introduced, and these texts are often accompanied by pictures and drawings. The visuals are usually accompanied by a short descriptive caption, as well as arrows and labels to highlight concepts and vocabulary. In the activity book, words are highlighted individually on a yellow background within a black frame.

In class, most Science sessions started with a teacher presentation of the contents followed by practice activities that were usually carried out with the whole class group. In the same way as the EFL lessons, one session a week took place in the computer room, and it was also devoted to preparing PowerPoint presentations on the vocabulary being studied. The teacher made a selective use of the textbook as he thought some of the texts were too difficult for the students, but all the activities from the workbook were completed and all the target vocabulary from each unit was covered. At times students were observed to engage in spontaneous discussions or to self-initiate questions to the teacher in relation to the content of the lesson (e.g., a discussion in L1 about whether plants can move; a student’s question in L1 as to why salamanders do not die when they lose their tail). However, these opportunities were not always picked up on by the teacher. During these lessons, the teacher almost always used English but, in contrast to the EFL lessons, he sometimes allowed students to use Spanish or Catalan to respond to his questions.

Students’ self-reported evaluations of the English and Science lessons were quite similar, with three quarters of the class saying that they liked them or liked them a lot and only a minority (one or two students) saying they did not like them much or did not like them at all. About the same number of students (roughly half the class) thought English and Science lessons were neither difficult nor easy subjects, although there were a few more students who found the EFL class easier or much easier than the Science class (26% vs. 17.4% respectively).

Instruments and analysis

Textbook analysis and vocabulary tests constitute the two primary instruments employed in this study. Secondary sources include six classroom observations, two teacher interviews, and a questionnaire with bio and attitudinal data that students completed at
the beginning and end of the school year. The vocabulary analysis of the two course books focused on the target words in the units that were covered during the instructional period under analysis; it was carried out with Lex Tutor. This online program is based on the Lexical Frequency Profiler (Cobb 2014; Heatley, Nation, and Coxhead 2002) and divides the words of texts into first and second thousand levels, academic words, and the remainder or ‘off-list’ words.¹

The vocabulary covered in the EFL class during the fall term was tested at the beginning and end of that term (September-Time 1 and December-Time 2) while the vocabulary covered in the Science class during the winter term was tested at the beginning and end of that term (December-Time 1 and March-Time 2). The December test included items from both the English and Science classes presented at random. In each vocabulary test, the students’ knowledge of 30 items was assessed. Tests included small drawings, next to which students were asked to write a word indicating what was being illustrated; thus, productive vocabulary knowledge was assessed. The initial letter of each word was provided as a prompt. A receptive vocabulary test was also administered but the results proved to be unreliable and, so, the test is not used in this study (see footnote² for an explanation). Because of the format of the productive vocabulary test, only concrete nouns were included (in contrast, the receptive vocabulary test did include adjectives and verbs as well as abstract and concrete nouns). Approximately the same number of items from each teaching unit was assessed in the productive and receptive tests. These items were chosen from the vocabulary component of the syllabus for each unit in the two textbooks. In the EFL test, we included the words that were highlighted on the vocabulary list. The CLIL vocabulary list included more items than the EFL list and they were not highlighted. We included vocabulary items that were easily depicted (and therefore interpreted). See Appendix 1 for the list of words included in the EFL and Science tests – note, these words were the same in both the pre- and post-tests. The tests were tailor-made for this study since no standardized tests could be used, but they were previously piloted with children of the same age to ensure that the pictures were clear and that the length of the test was adequate.

In scoring the tests, correctly produced items were awarded one point, even if words were misspelled. In analyzing the results, related and independent samples t-tests were conducted. The level of significance was set at .05.

Results

Textbook analysis: vocabulary profile

During the fall term, the first six units of the EFL class and activity books (Incredible English 3) were taught in class. They included the following vocabulary areas: personal possessions, clothes, food, places in a town, animals, and adjectives to describe people and music. During the winter term, five units were covered in class from the CLIL text and activity books (Top Science 3) including the following topics: the body, the senses, living things, vertebrate animals, and invertebrate animals. The textbook analysis reported below focuses on the vocabulary that was encountered in these units, since this serves as the pool of vocabulary that students were subsequently tested on as we sought to answer research questions two and three.

A preliminary examination of the wordlists from the two textbooks reveals certain differences. The average number of lexical items per unit is higher in the Science textbook than in the EFL class book (77 vs. 49 words, respectively), but the proportion of multiword lexical items is considerably lower on the Science wordlists (11% vs. 27%
The most frequent multiword items in the EFL class book are fixed expressions, that is, utterances to be learnt as a whole (e.g., Here you are; What’s the matter?; What a mess!), and productive expressions (e.g., I can’t find …; Here’s …). The most common types of multiword units in the Science textbook are phrasal verbs (e.g., to breathe in; to cut off) and nouns preceded by modifiers (e.g., life cycle; magnifying glass).

A further analysis of the two wordlists (including chunks but not productive expressions) by word frequency (tokens) was performed with Lextutor (Cobb 2014). This analysis is based on the total number of target words in the first six units in the EFL class book and the first five in the Science textbook, the number being considerably higher in the latter (a total of 256 vs. 337 words, respectively). Table 1 shows the proportions of the first and second thousand level words as well as the academic and off-list words in both course books. Here, although some differences are apparent, they are not great. Word types forming part of the first and second thousand levels are slightly more numerous in the case of the EFL class book, while the proportion of academic and off-list words are a little higher in the Science textbook. The same trend applies for Latin cognates, with the Science textbook showing slightly higher indexes.

A closer inspection of the items included among the off-list words reveals more significant differences. For example, in the case of the EFL class book there is a good proportion of lexis related to food (e.g., asparagus, biscuits, broccoli), clothing (e.g., tights, trousers, gloves), and everyday objects (e.g., magazine, jar, torch), whereas the Science textbook list includes a good proportion of classifying nouns and adjectives (e.g., carnivore, canine, auditory), words designating processes or abstract concepts (e.g., germinate, height, vibration), and nouns related to human and animal body parts (e.g., iris, sternum, womb), some of which are quite technical.

Finally, the words included on the tests used to measure vocabulary gains (see next section) reflected the vocabulary profiles of the EFL and CLIL course books. The EFL test contained more words belonging to the first and second frequency bands (27% and 36%) than the CLIL test (23% and 27%). In contrast, the proportion of ‘off-list’ words on the CLIL test was higher (50%) than on the EFL test (36%).

**Vocabulary gains**

Gains in students’ productive knowledge of EFL words from Time 1 (beginning of the fall term) to Time 2 (end of the fall term) are examined first. The mean score at Time 1 was 8.9 for the whole sample (n = 22), increasing to 20.1 at Time 2. The related samples t-test indicates that the difference between these two scores is significant with a large effect size, according to Cohen’s $d$ (see Table 2). When we repeated this same test with

| Frequency          | EFL textbook ($n = 256$ tokens) % | Science textbook ($n = 337$ tokens) % |
|--------------------|----------------------------------|---------------------------------------|
| 1k words           | 39.22                            | 33.51                                 |
| 2k words           | 27.83                            | 21.73                                 |
| Off-list words     | 31.76                            | 40.31                                 |
| Academic words     | 1.18                             | 4.45                                  |
| Latin cognate index| 33.91                            | 39.47                                 |
the subsample of students not taking English lessons as an extra-curricular activity ($n = 11$), the results were equally significant. It should be noted that in both cases the standard deviations are quite high, showing considerable variability among students.

Gains in students’ productive knowledge of Science words from Time 1 (beginning of the winter term) to Time 2 (end of the winter term) are examined next. The mean score at Time 1 was 11.8 for the whole sample ($n = 23$), increasing to 20 at Time 2. The related samples $t$-test indicates that the difference between these two scores is significant with a large effect size (see Table 3). When we repeated this same test with the subsample of students not taking English lessons as an extra-curricular activity ($n = 11$), the results were equally significant.

Finally, a comparison of the gains between the EFL and CLIL vocabulary tests is drawn. In order to do this, the gains for each student were obtained by subtracting their scores at Time 1 from those obtained at Time 2. The results from an independent samples $t$-test show that students’ mean gains are significantly higher in the EFL context and that the magnitude of this difference is large (see Table 4). In the EFL tests students were able to produce an average of 11 more words at Time 2 than at Time 1, whereas in the CLIL tests students were able to produce an average of eight more words at Time 2 than at Time 1.

| Table 2. Gains on EFL vocabulary test (fall term). |
|-----------------------------------------------|
|                                | Time 1 mean (SD) | Time 2 mean (SD) | $t$ | $p$ | $d$ |
| Whole sample ($n = 22$)          | 8.9 (5.6)        | 20.1 (6.7)       | 10.4 | .000 | .8 |
| Subsample (extra-curricular ss excluded) ($n = 11$) | 8.1 (6.3)        | 20.9 (5.8)       | 8.7  | .000 | .9 |

| Table 3. Gains on CLIL vocabulary test (winter term). |
|-----------------------------------------------|
|                                | Time 1 means (SD) | Time 2 means (SD) | $t$ | $p$ | $d$ |
| Whole sample ($n = 23$)          | 11.8 (4.5)        | 20 (4.5)          | 13.1 | .000 | .9 |
| Subsample (extra-curricular ss excluded) ($n = 11$) | 10.5 (4.6)        | 19.4 (4.3)        | 8.9  | .000 | .9 |

| Table 4. Gains on EFL and CLIL tests. |
|-----------------------------------------------|
|                                | Mean gains (SD) | $n$ | $t$ | $p$ | $d$ |
| EFL context                     | 11.6 (4.6)      | 21  | 4.4 | .000 | .5 |
| CLIL context                    | 8.1 (2.9)       |     |     |     |     |

**Discussion and conclusions**

A comparison of the target vocabularies that students were exposed to in the EFL and CLIL contexts revealed a number of differences. A frequency analysis of the target vocabularies showed slight differences between the EFL and Science materials, with the
former presenting somewhat higher proportions in the first and second frequency word bands and the latter presenting slightly higher proportions of Latin-based and off-list words. We would expect these differences to become more marked in materials targeting higher level students as contents become more specialized and topics are dealt with in greater depth.

A more detailed analysis revealed other more obvious differences. Thus, the CLIL materials included roughly 25% more items than the EFL materials, which probably reflects the greater role played by content and information in the former. In terms of word types, the differences identified probably reflect the text types included in the EFL and CLIL materials. The frequent presence of stories, songs, and dialogues in the EFL materials would explain why multiword units often included chunks and productive language (more typical of oral language), while the predominance of expository texts in the CLIL materials would explain the presence of nominal premodification and abstract language (more typical of descriptive language and the scientific genre). In terms of the visual presentation of vocabulary, the target words in the CLIL materials were more often and more prominently highlighted than those in the EFL materials. An analysis of the words used in the vocabulary tests that the students took confirmed this difference between CLIL and EFL vocabularies.

An examination of the gains students made between the start and the end of the two-term period demonstrated that there was a considerable growth in the productive vocabulary of the learners in both contexts, thus confirming the overall positive benefits of CLIL instruction on vocabulary learning (Buyl and Housen 2014; Jiménez Catalán and Ruiz de Zarobe 2009; Seregély 2008; Sylvén 2004). However, a comparison of the gains between the two contexts showed that, on average, students learned more words from the EFL lessons than from the CLIL lessons and that this difference was significant. The fact that students used their L1 more often than in the EFL lessons to answer the teacher’s questions may account, in part, for these results. Likewise, the fact that the vocabulary in the EFL class book included more frequent words of probably greater relevance to students’ daily lives may be another explanatory factor. In fact, it seems likely that the vocabulary in the EFL textbook could more readily become part of the existing everyday word network in the child’s mental lexicon. In contrast, some of the vocabulary in the CLIL textbook may not yet have an appropriate web of words in the child’s mental lexicon to which they might easily connect. Even though receptive knowledge could not be tested in this study (see footnote ii), and it is one of its limitations, it is possible that the impact of CLIL would have been more marked in the case of receptive than productive vocabularies, as reported elsewhere (Ruiz de Zarobe 2011). Additional limitations that need to be taken into account when interpreting the results are that the study involved a small sample of language learners and a few units from the respective textbooks. The fact that the vocabulary test was restricted to concrete words made the test less representative of the vocabulary students were exposed to in their class books, this being especially pertinent in the case of CLIL. A further limitation is the fact that important factors affecting vocabulary learning were not taken into account in the analysis of the data such as word length, frequency of use, textual context, among others, etc.

Despite the scale of the study, we feel it has a number of implications for further research. First, the research design employed here, with the number of hours of instruction in CLIL and EFL being constant, strengthens the validity of the study. The design is also methodologically innovative, at least in comparative studies involving CLIL, and it has proved to be feasible in practice. Second, in the process of designing the
vocabulary tests, we faced major difficulties, to the point of having to discard the receptive vocabulary test. Further research needs to pay special attention to how CLIL vocabulary might be assessed with young children. The fact that some of this vocabulary typically includes abstract and technical terms is an added difficulty when it comes to conducting assessments with this age group. Finally, some of the aspects of classroom practice that were briefly described above in contextualizing the results of this study point to some intriguing questions. For example, our observation data suggest certain differences in classroom interaction between the CLIL and EFL sessions, there being more teacher-fronted teaching during the CLIL than the EFL sessions. There also seemed to be more student-initiated moves in the CLIL sessions, in spite of them being in the students’ L1. Another issue of interest is the difference in difficulty of the EFL and CLIL textbooks, as identified to us by the teacher. A subsequent analysis of a random unit from the two textbooks confirmed the teacher’s impressions: there being an average of 9 and 5.1 words per sentence in the CLIL and EFL texts and an average grade level in terms of comprehension difficulty of 4.2 and 0.6, respectively. Another issue worth exploring is the profile of the teacher and how this affects classroom practice. In this study, the teacher was a language teacher not a science teacher and this might explain the fact that he was observed to provide more unplanned explanations during the EFL than during the CLIL lessons. He also had students spend time during the CLIL sessions on vocabulary practice (PowerPoint presentations), an activity that was also carried out in the EFL sessions and which is more characteristic of a language than a science class. The significance of these aspects of classroom practice could be of special interest for textbook development and teacher training and require further exploration.

All in all, there seems to be sufficient indications in the study reported here to suggest that learning Science in English was a more difficult experience for our eight-year-old learners than learning the language in the EFL class. The CLIL textbook was considerably more difficult than the EFL class book, and the teacher was more permissive with the use of L1 when teaching Science. There were also more target words to be learned in each of the CLIL units, and the analysis with Lextutor revealed there to be slightly more academic, Latin cognate, and off-list words in the Science vocabulary. An examination of the off-list words also revealed that the Science textbook included more words designating processes and abstract nouns in contrast to the more everyday topics covered in the EFL class book. The complexity of the Science lessons was also reflected in students’ self-reports, with a few more students finding the EFL class easier or much easier and none finding Science to be easier. Given this evidence, it would be simplistic to interpret our analysis of word gains in isolation without acknowledging that CLIL and EFL lessons offered quite different learning contexts. Besides, it should not be forgotten that vocabulary learning in the context of CLIL is a more complex process, since it is not limited just to learning new words but it also implies understanding new concepts and performing tasks with different cognitive demands.

In sum, the present study has examined the CLIL and EFL approaches with young learners of English following a design that guarantees the comparability of the groups under analysis. The analysis of the vocabulary in the course books used by the students revealed marked differences, indicating that CLIL and EFL materials offer complementary sources of input. The results have also shown that both approaches offer valid contexts for the development of productive vocabulary knowledge and that Science in English is a more challenging learning experience for most children.
Acknowledgment
We would like to thank Imma Miralpeix for her insightful comments.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by the Spanish Ministry of Education under grants [FFI2010-18006, FFI2013-40952-P], and by the ‘Generalitat de Catalunya’ [2014 SGR 1089].

Notes
1. Off-list words are those that belong to the third frequency band and beyond and which are not considered as being academic words.
2. At each testing time, students were also evaluated on their receptive vocabulary knowledge. Yes/no vocabulary recognition tests (following Meara 1992) were developed and administered together with the productive vocabulary tests. Each vocabulary recognition test included 60 words, 20 of which were imaginary. Unfortunately when these tests were scored, using a formula that adjusts for guessing, it was apparent that there had been too much guessing, which meant that results were not reliable and had to be discarded.
3. The Flesch–Kincaid Grade Level formula was used to calculate comprehension difficulty. The formula is based on the number of words and the number of syllables per word and results correspond to the number of years in education generally required to understand a text. For more information see https://readability-score.com/.

References
Aguilar, A., and C. Muñoz. 2014. “The Effect of Proficiency on CLIL Benefits in Engineering Students in Spain.” International Journal of Applied Linguistics 24 (1): 1–18. doi:10.1111/ijal.12006.
Agustín Llach, M. P., and R. M. Jiménez Catalán. 2007. “Lexical Reiteration in EFL Young Learners’ Essays. Effect of the Type of Instruction.” International Journal of English Studies 17 (2): 85–103.
Banegas, D. L. 2014. “An Investigation into CLIL-related Sections of EFL Coursebooks: Issues of CLIL Inclusion in the Publishing Market.” International Journal of Bilingual Education and Bilingualism 17 (3): 345–359. doi:10.1080/13670050.2013.793651.
Brown, D. 2010. “What Aspects of Vocabulary Knowledge Do Textbooks Give Attention to?” Language Teaching Research 15 (1): 83–97. doi:10.1177/1362168810383345.
Bruton, A. 2011. “Is CLIL so Beneficial, or Just Selective? Re-evaluating Some of the Research.” System 39: 523–532. doi:10.1016/j.system.2011.08.002.
Buyl, A., and A. Housen. 2014. “Factors, Processes and Outcomes of Early Immersion Education in the Francophone Community in Belgium.” International Journal of Bilingual Education and Bilingualism 17 (2): 178–196. doi:10.1080/13670050.2013.866627.
Cobb, T. 2014. Web Vocamprofile. Accessed May 29, 2014. http://www.lextutor.ca/vp/.
Coyle, D., P. Hood, and D. Marsh. 2010. CLIL Content and Language Integrated Learning. Cambridge: Cambridge University Press.
Dalton-Puffer, C. 2007. Discourse in CLIL Classrooms. Amsterdam: John Benjamins.
Dalton-Puffer, C. 2009. “Communicative Competence and the CLIL Lesson.” In Content and Language Integrated Learning: Evidence from Research in Europe, edited by Yolanda Ruiz de Zarobe and Rosa Maria Jiménez Catalán, 197–214. Clevedon: Multilingual Matters.
European Commission Communication. 2004. Promoting Language Learning and Linguistic Diversity: An Action Plan 2004–2006. Accessed December 31, 2014. http://www.saaic.sk/eu-label/doc/2004-06_en.pdf.
European Commission (Languages). 2013. *Content and Language Integrated Learning*. Accessed August 15, 2013. [http://ec.europa.eu/languages/language-teaching/content-and-language-integrated-learning_en.htm](http://ec.europa.eu/languages/language-teaching/content-and-language-integrated-learning_en.htm).

Eurydice. 2006. *Content and Language Integrated Learning (CLIL) at School in Europe*. Brussels: European Commission.

Fusté, J., and I. Miralpeix. 2013. “Assessing the Vocabulary Knowledge of Young EFL Learners in Instructional Settings.” Paper presented at the 37th AEDEAN Conference, Oviedo, November 13–15.

Heatley, A., I. S. P. Nation, and A. Coxhead. 2002. *Range and Frequency Programs*. [http://www.victoria.ac.nz/lals/staff/paul-nation](http://www.victoria.ac.nz/lals/staff/paul-nation).

Heras, A., and D. Lagasabaster. 2015. “The Impact of CLIL on Affective Factors and Vocabulary Learning.” *Language Teaching Research* 19 (1): 70–88. doi:10.1177/1362168814541736.

Jexenflicker, S., and C. Dalton-Puffer. 2010. “The CLIL Differential. Comparing the Writing of CLIL and Non-CLIL Students in Higher Colleges of Technology.” In *Language Use and Language Learning in CLIL Classrooms*, edited by Christiane Dalton-Puffer, Tarja Nikula, and Ute Smit, 169–189. Amsterdam: John Benjamins.

Jiménez Catalán, R. M., and R. Mancebo Francisco. 2008. “Vocabulary Input in EFL Textbooks.” *RESLA* 21: 147–165.

Jiménez Catalán, R. M., and Y. Ruiz de Zarobe. 2009. “The Receptive Vocabulary of EFL Learners in Two Instructional Contexts: CLIL vs. Non-CLIL Instruction.” In *Content and Language Integrated Learning. Evidence from Research in Europe*, edited by Yolanda Ruiz de Zarobe and Rosa Maria Jiménez Catalán, 81–92. Clevedon: Multilingual Matters.

Marsh, D., ed. 2002. *CLIL/EMILE. The European Dimension. Actions, Trends, and Foresight Potential*. Jyväskylä: University of Jyväskylä.

Meara, P. 1992. *EFL Vocabulary Tests*. Swansea: Centre for Applied Language Studies.

Moghadam, N. Z., and H. Fatemipour. 2014. “The Effect of CLIL on Vocabulary Development by Iranian Secondary School EFL Learners.” *Procedia - Social and Behavioral Sciences* 98: 2004–2009.

O’Dell, F. 1997. “Incorporating Vocabulary into the Syllabus.” In *Vocabulary and Language Teaching*, edited by Norbert Schmitt and Michael McCarthy, 258–278. Cambridge: Cambridge University Press.

Phillips, S., and M. Morgan. 2007. *Incredible English Kit 3. Classbook and Activity book*. Oxford: Oxford University Press.

Ruiz de Zarobe, Y. 2010. “Written Production and CLIL: An Empirical Study.” In *Language Use and Language Learning in CLIL Classrooms*, edited by Christiane Dalton-Puffer, Tarja Nikula, and Ute Smit, 191–209. Amsterdam: John Benjamins.

Ruiz de Zarobe, Y. 2011. “Which Language Competences Benefit from CLIL? An Insight into Applied Linguistics Research.” In *Content and Foreign Language Integrated Learning: Contributions to Multilingualism in European Contexts*, edited by Yolanda Ruiz de Zarobe, Juan Manuel Sierra, and Francisco Gallardo del Puerto, 129–154. Bern: Peter Lang.

Santillana. 2011. *Top Science 3: Textbook and Activity Book*. Madrid: Santillana/Richmond.

Schenk, J. 2013. “A Lexical Approach to Coursebook Analysis: How an EFL Coursebook and a CLIL Coursebook Approach Vocabulary Acquisition.” Unpublished Master’s thesis, Universitat de Barcelona.

Seregély, A. M. 2008. “A Comparison of Lexical Learning in CLIL and Traditional EFL Classrooms.” Unpublished MA thesis, Universität Wien.

Sylvén, L. K. 2004. *Teaching in English or English teaching? On the Effects of Content and Language Integrated Learning on Swedish Learners’ Incidental Vocabulary Acquisition*. Göteborg: Department of English, Göteborg University.

Varkuti, A. 2010. “Linguistic Benefits of the CLIL Approach: Measuring Linguistic Competences.” *International CLIL Research Journal* 1 (3): 67–79.

Wegner, A. 2012. “Seeing the Bigger Picture: What Students and Teachers Think About CLIL.” *International CLIL Research Journal* 1 (4): 29–35.
Appendix 1. Items included in vocabulary tests

Test items from the EFL textbook:

biscuits, bridge, butterfly, cap, cheese, coat, comb, dress, eggs, flower, glasses, gloves, grasshopper, guinea pig, ladybird, leaf, library, pillow, rabbit, river, scarf, sleeping bag, snake, strawberries, swimming pool, toothbrush, torch, towel, trainers, wood.

Test items from the CLIL textbook:

ant, bat, bear, beetle, brain, branch(es), cow, crab, dolphin, ear, eye, eyebrow, eyelash(es), fly, horse, jellyfish, lamb, mussel, nose, root, seed, skull, spider, squirrel, stems, sunflower, tear, tongue, whale, wings.