Towards an Understanding of
CLIL Assessment Practices
in a European context:
Main Assessment Tools and the Role of Language in Content Subjects

ANA OTTO

MADRID OPEN UNIVERSITY

anaotto03@yahoo.es

JOSÉ LUIS ESTRADA

UNIVERSITY OF CÁDIZ

joseluisestrada@uca.es

Bilingual Education implies curricular integration along with new teaching procedures. However, a closer look at CLIL contexts shows that, very frequently, these new methodologies have not been integrated in assessment. This article provides a comprehensive overview of CLIL assessment practices in the context of the CAM Bilingual Project in Spain. More specifically, by using responses from two focus groups and comparing them with prior teachers' questionnaires, the study examines the main assessment tools content teachers use in such settings, and the role that language plays in the learning of content subjects. The research findings provide relevant insights in relation with teacher training in bilingual schools and the absence of formative assessment in the context of the study. Therefore, written exams stand out as the most common assessment tool and, furthermore, the students' language level is taken into account in grading the subject. On the basis on these results, a set of recommendations for the teachers in Bilingual Sections of Madrid are proposed.

La educación bilingüe supone una integración curricular además de una verdadera innovación metodológica. Sin embargo, cuando observamos la realidad de los diversos contextos CLIL, comprobamos que, a menudo, estas nuevas metodologías no se han incorporado al ámbito de la evaluación. En este artículo se ofrece un panorama global de las prácticas de evaluación en el contexto del Plan Bilingüe de la Comunidad de Madrid, en España. En concreto, a partir de las respuestas de dos focus groups comparados con las respuestas de los profesores a un cuestionario anterior, el estudio analiza los principales instrumentos de evaluación que los profesores de contenido encuentran en dichos contextos y el papel que la lengua desempeña en el aprendizaje de asignaturas de contenido. Los hallazgos de la investigación aportan datos relevantes relacionados con la formación del profesorado de centros bilingües y la ausencia de evaluación formativa en el contexto del estudio. Así, se observa que el examen escrito prevalece como el instrumento de evaluación más frecuente y que, además, el nivel lingüístico de los alumnos se tiene en cuenta a la hora de calificar la asignatura. A partir de estos resultados se formulan una serie de recomendaciones para el profesorado de las Secciones Bilingües de la Comunidad de Madrid.

KEYWORDS:

CLIL; assessment; focus group; CAM Bilingual Project.

PALABRAS CLAVE:

CLIL; evaluación; focus groups; Programa Bilingüe de la Comunidad de Madrid.

Otto, A.; Estrada, J.L.

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

1. Introduction

The Comunidad de Madrid Bilingual Project (henceforth CAM Bilingual Project) is a state funded program which started in 2004 in Primary schools and was made extensive to the secondary level in 2010. Bilingual high schools in Madrid offer two different tracks: the Bilingual Program and the Bilingual Section. Considered as the real bilingual project for secondary schools regarding the time devoted to the use of English as a vehicular language, the Bilingual Sections in the CAM Bilingual Project are the focus of our research. As in the Bilingual Program, English as a Foreign Language or the so-called Advanced English Curriculum is taught five days a week with a one-hour session each day. This subject substitutes English as a Foreign Language in the first, second, third and fourth grades of Compulsory Secondary Education, and it is aimed at providing students with advanced language skills by covering both English language and literature. As for other subjects taught through the medium of English, the teaching of the Advanced English Curriculum with the rest of the subjects taught in English (Biology and Geology, Geography and History, the tutoring hours and another optional subject) takes at least one-third of the weekly schedule. For a student to be eligible to join the Bilingual Section, s/he is required to certify a minimum level of A2 according to the CEFR although a B1 is highly recommended.

As in other bilingual programs across Europe, the CLIL approach was adopted to teach non-linguistic subjects, except for Mathematics and Spanish Language, using English as a vehicular language. That implied that a conceptual framework for content and language integration needs to go hand in hand with the adoption of new educational approaches and methodologies. However, despite its rapid growth, and the significant involvement of educational authorities, teachers and families, bilingual programs in Europe are still object to improvement concerning aspects such as teacher training, methodologies, the use of appropriate materials, and the way assessment is conducted.

When it comes to assessing students' learning, which is one of the most controversial issues in CLIL, the most common debate arises in the attempt to identify the nature of CLIL assessment (Coyle et al., 2010; Kiely 2009; Järvinen 2009), and how teachers deal with the integration of content and language. Other aspects are related to the methods and tools which are best suited to assessment in CLIL, the best way to measure previous knowledge and/or progression, skills and processes, cognition and culture (Coyle et al., 2010), the need to implement formative assessment (Ball, Kelly & Clegg, 2015) and the role of language in assessment (Llinares, Morton & Whittaker, 2012) among others.

Formative Assessment or Assessment for Learning (AfL) is "the process of seeking and interpreting evidence for use by learners and their teachers, to identify where the learners are in their learning, where they need to go to, and how best to get there" (Assessment Reform Group, 2002, p. 2). As it informs instruction, it can help teachers to motivate students to develop a positive attitude towards content along with a simultaneous improvement in the vehicular language performance. This type of assessment also stands out as having a task-based nature, and for the wider variety of classroom interaction that it promotes (Ball, Kelly & Clegg, 2015, p. 213). Although Formative Assessment is recommended in CLIL, it is necessary to point out that it can also be used along with Summative Assessment, as is still present in some educational contexts. In fact, the combination of both Formative and Summative Assessment can benefit the latter especially when Formative Assessment is based on rigorous planning and uses robust instruments and tools suited to CLIL subjects, leading to more soundly based assessment process (Llinares, Morton & Whittaker, 2012, p. 282).

However, despite recommendations, and probably due to the variety of CLIL models, the relative novelty of this integrated educational approach, and the lack of established assessment criteria, the small number of studies completed on CLIL assessment (Serra, 2007; Serragiotto, 2007; Hönig, 2010; Wewer, 2014 and Reierstam, 2015) show evidence of significant disparity among the assessment practices conducted in CLIL programs mainly regarding the type of exams and the extent to which they are adapted to students' levels.

With the analysis of teachers' opinions about their assessment practices in the Bilingual Sections of the CAM Bilingual Project, this study aims to address this gap in the CLIL literature, and thus, to analyze the impact that assessment has on teaching and learning.

2. Teachers' focus groups

Teacher focus groups were conducted as part of a mixed-method research combining quantitative and qualitative data on the impact that assessment has on CLIL teaching and learning in bilingual secondary schools in Madrid (Otto, forthcoming). After having gathered initial information through teachers' questionnaires, the focus groups were aimed to clarify aspects about the main assessment tools teachers use and the weight of language in bilingual subjects. Focus group interviews are excellent to complement other quantitative and qualitative research methods as they bring depth into the research, allow the researcher to verify findings from surveys and questionnaires (Vaughn, Schumm & Sinagub, 1996), and because they can help to shed light on aspects which were left unclarified in previous studies or stages of the research. In this work, the focus groups used the phenomenological approach, i.e. to understand the topic of assessment through the perspective of the everyday knowledge and practice of the participants, with the main purpose of making the most of the synergy created in the groups, which is thought to contribute to the free expression of thoughts. In this sense, it is important to stress that bilingual coordinators played a relevant role as they raised the question of assessment among participants,

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

Otto, A.; Estrada, J. L.

and created a favorable climate for the meetings. The bilingual coordinator is, along with the principal and the rest of the school management team, one of the most important agents for the success of a bilingual program. S/he advises the principal and the rest of the management team, and supervises the successful implementation of the academic program of CLIL subjects.

2.1. The participants

The participants in this research are content teachers working in high schools in the CAM Bilingual Project. Teachers are specialists in the following subject(s): Music, Technology, Robotics, Biology, History and Geography, Physical Education and Arts and Crafts, and mostly Spanish native speakers who have certified a minimum of a C1 level of English proficiency which allows them to teach their subjects through English. As for their training and experience, they come from different backgrounds, and have different levels of experience, being some of them novice interim teachers recently arrived in a bilingual school, and some others veteran teachers coming from the first bilingual high schools in the MEC-British Council Project or from other schools which became bilingual in the recent years.

2.2. The work with the focus group

Two focus group interviews were conducted in two different schools consisting of 12 and 15 teachers each. The focus groups were carried out in order to refine and explore in depth some of the information gathered in a previous step of the research: the teachers' questionnaire, in which teachers stated to use a majority of written tests, and highlighted the lack of common guidelines in relation with language issues. The bilingual coordinators, being conscious of the importance of CLIL assessment for school life, invited all the members of the bilingual team i.e. content teachers, language teachers and language assistants in the first focus group, and content teachers and language assistants in the second focus group, with the main goal to facilitate teacher cooperation, draw further conclusions, and comment on future suggestions of improvement. However, as the teachers' questionnaires had previously made it clear that content teachers were the only ones assessing content subjects, the questions were strictly designed for them, so the rest of the group had an observer status. The focus group interviews were conducted in Spanish so that teachers would benefit from a relaxing atmosphere and could feel free to express their own views. The discussions were focused but some scope for individual perspectives was also considered beneficial, according to what Krueger (1994) calls "**the interview guide**" which provides subject areas and the possibility of freely exploring, and asking questions, depending on the participants' answers. Responses were analyzed focusing on the key questions driving the focus group, but attention was

also paid to additional comments by teachers as they help to understand their views and keep their conversation going smoothly. After the two groups were conducted, abridged transcripts were created with the most relevant and useful portions of the conversations. These transcripts were then analyzed using the constant comparative analysis (Krueger & Casey, 2009) to identify the most important trends or ideas by participants about the topic of assessment. Likewise, the questions were organized to move easily from one topic to the next, and special emphasis was laid on using non-technical vocabulary to promote teacher interaction at all times.

Group interaction was based on a list of topics pertaining to the main obstacles teachers find in CLIL assessment, the instruments they commonly use, and whether language competence has a direct influence on the grade they assign to a student, piece of homework/test or any tool they may use for assessment. Attention was also given to the way teachers deal with the absence of CLIL assessment guidelines, which was a common complaint according to data obtained from teachers' questionnaires and informal conversations; i.e. whether they communicate with colleagues in their department and/or at school to know how to deal with assessment issues, and whether they have coordinated in that matter or have reached any agreements so far on topics such as the role of the foreign language in CLIL assessment or the aspects that could be penalized (if any) in assessing the language.

3. Results

3.1. First focus group interview

The **first focus group (FG1)** interview took place in March 2015 in the library of the High School in eastern Madrid. It involved 12 teachers -permanent and temporary staff- along with the Bilingual Coordinator. In this first group, the discussion focused mainly on the weight of the English language in CLIL subjects along with the criteria teachers have to correct language aspects and the teachers' roles. Teachers' views revealed that they found it extremely difficult to assess content knowledge without taking language proficiency into account. In fact, as they pointed out during the focus group session, the difficulties which Bilingual Education can entail in terms of students' production in the foreign language has always stood out as a controversial topic in the school, which attracted most teachers' interest. Consequently, this issue had been previously discussed on many occasions during school meetings since the implementation of the bilingual program three years earlier. Finally, in the academic year 2013-14, the teaching staff agreed on an "**Improvement Plan for Writing Skills**" to be used by all content teachers in both non-bilingual and bilingual groups. The plan was aimed at improving writing skills in English and Spanish and for that purpose, it was initially devoted to agree on joint rules for the presentation and organization of students' class notebook and academic work, as well as for the outline of exams and

Otto, A.; Estrada, J.L.

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

project work. Those actions led teachers to agree on the assessment criteria regarding writing skills and grammar mistakes in exams and students' work. After having analyzed typical mistakes and having created a framework for written proficiency, both assessment and grading criteria were modified accordingly in all the subjects, and families were informed about these guidelines through the students' school diary.

> **"The difficulties which Bilingual Education can entail in terms of students' production in the foreign language has always stood out as a controversial topic in the school, which attracted most teachers' interest."**

Regarding **the role of English in CLIL assessment**, teachers overtly showed their concerns about the topic and immediately started asking about the existence of general guidelines as they complained about the lack of information and teacher training in CLIL issues.

"We don't have much idea about it, to be honest. What are we supposed to do about assessment?". (FG1-M)

They also emphasized that their main goal as content teachers in relation with language is that students are successful in acquiring academic vocabulary or what they term as "CALP, the specific vocabulary from their subjects". In this sense, it is interesting to notice that although CALP (Cognitive Academic Language Proficiency) is more than just academic vocabulary, teachers tend to simplify the concept to refer to the specific language of the subject:

"We always emphasize the vocabulary of the subject. Students have to learn it and know how to use it to express content. In Music, for instance, it is essential to know ordinal and cardinal numbers, they learnt that in Primary Education. As for the new concepts, or definitions, etc. above all, they are names in Italian. Well, I suppose I can overlook some spelling mistakes". (FG1-A)

When asked about **error treatment in CLIL subjects**, all the members of the focus group seemed to be clearly concerned about how to deal with language errors as they commented on the most typical grammar mistakes - the -s in the third person singular, starting a sentence using "that" which is obviously Spanish-like word order:

"Some mistakes need to be fixed immediately. Otherwise, they go viral…" (FG1-J).

Nevertheless, although teachers recognize the need to correct students while speaking, most of the teachers tend to favor intelligibility over accuracy. In this regard, it is interesting to see the tendency they show to contrast accuracy and fluency as if the first did not help the latter in the process of content expression, as can be seen in the following comment:

"I usually focus on whether the writing is easy to understand. I go for comprehensibility because CLIL is a communicative approach". (FG1-C)

In fact, accuracy in writing had also been a controversial issue they had been discussing for years. As the different departments were not in agreement on the best ways to deal with language mistakes in CLIL subjects; i.e. whether they should just be highlighted or also marked down, they asked the English language department for advice. Apparently, although the English teachers had not agreed on a taxonomy of errors themselves, this request proved useful for them so as to identify common mistakes which were later used to design the improvement plan for written skills. However, despite these agreements, it might be the case that in current practice, each teacher corrects what s/he finds appropriate depending on the level, the subject and the group with a focus on fluency over accuracy:

"I sometimes come across sentences with no -s in the third person singular but they express so much content knowledge that for me it's fine, it is enough" (FG1-R). Another teacher states: *"I know there were some agreements about the way we correct but we also need to look at other aspects which have not been considered, and which are also necessary".* (FG1-E)

In this sense, and regarding the joint rules they agreed on the improvement plan for written skills, it is interesting to notice that although the plan was globally perceived as positive, some teachers complain that there is more flexibility in CLIL subjects than in Spanish:

"This is like when a student goes and starts a definition using "when". We don't accept that in non-bilingual groups. Students can't start a definition using "when" in Spanish. But then we allow them to do that in English. You can even find a definition like that in a textbook! So of course, I believe we take comprehensibility rather than accuracy or grammar mistakes into account". (FG1-M)

The discussion also raised issues about **the role of the content teacher** as opposed to that of the language teacher, and it revealed the fact that content teachers seem to be uncomfortable when correcting and grading language mistakes:

"I am afraid if I devote too much time to check and fix English mistakes, I will end up being a teacher of English. However, my students sometimes don't know how to express content in my subject…" (FG1-P).

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

Otto, A.; Estrada, J. L.

Finally, an additional difficulty that teachers have dealing with **the weight of English in CLIL** is that they are also afraid that in some situations their language level might not be good enough, and they might make mistakes that students could repeat, as one the teachers states:

"Sometimes, I also need to have a grammar or a dictionary around when I am grading exams. Yes, that happens sometimes, to make sure this guy is writing this and that the correct way. How am I supposed to do that if I am not sure to have that proficiency level in English? I am a Science teacher, not an English teacher". (FG1-C)

As in some of the comments from the teachers' questionnaire, the participants also expressed their concerns about the difficulties they find when **selecting appropriate assessment tools for CLIL contexts**. Despite the presence of the Improvement Plan for Written Skills in this school, the general procedure for assessment criteria in Spanish Secondary Education is set by the didactic department which usually comprises non-bilingual and bilingual groups. Thus, exam formats and assessment tools are usually designed for non-bilingual groups, namely, tests including essay questions. These essay parts might be problematic for bilingual groups even in the case of Bilingual Sections where students need to express content knowledge through productive skills - being writing the preferred mode - which is challenging since the language level in English is lower than in Spanish:

"The main problem is that regardless of whether you have bad, good or excellent materials, when it comes to assessment tools, I mean the way exams and tests are designed, it's completely different. I don't know about you, but I can't expect my students will be able to write in English the way they would write in Spanish". (FG1-A)

> **"Exam formats and assessment tools are usually designed for non-bilingual groups, namely, tests including essay questions. These essay parts might be problematic for bilingual groups even in the case of Bilingual Sections where students need to express content knowledge through productive skills."**

As for **the absence of clear guidelines for CLIL assessment**, comments showed that teachers agree that the Ministry of Education or Regional Government of Madrid should offer specific guidelines regarding assessment regulations for bilingual schools in the CAM Bilingual Program. As respondents put it, the assessment tools designed for non-bilingual groups are not in line with bilingual education, and a great deal of effort needs to be made to create specific CLIL materials which are not mere translations from Spanish. Apart from that, in the absence of guidelines, more freedom should be given to bilingual schools so that assessment tools, methods and criteria can be set apart from those recommended by the didactic department which are common for both bilingual and non-bilingual schools. In fact, a common complaint by parents, they assert, is that bilingual students can have easier exams than their non-bilingual partners, which some people think can devalue Bilingual Education:

"Besides, we have that pressure from the parents. When families come, they tell us non-bilingual students have much more difficult exams, essay-type exams while bilingual groups sometimes do that, but not always, they have these matching activities, more visual support...But we are aware we can't expect the same linguistic level in the other groups, the Spanish groups, that's a fact". (FG1-M)

Regarding **alternative assessment tools**, such as the portfolio and peer and/or self-assessment, which are usually recommended for CLIL (Wewer, 2014), their absence is quite noticeable according to teachers:

"We correct the activities at the end of the term, we assess the didactic units. This is the best way to check they were working on a regular basis. No, we don't really use the portfolio". (FG1-A)

Another teacher points out:

"I don't know about the rest of the teachers in the department, but I don't use self or peer-assessment. The students do know about their progress because the activities are corrected in class. Activities are always corrected here". (FG1-O)

3.2. Second focus group interview

The **second focus group (FG2)** took place in the meeting room of the High School in a town in the South of Madrid. It included 15 content teachers, five language assistants and the bilingual coordinator who expressed her wish to include all the members in the bilingual team in the meeting. It is important to point out that this high school has extensive experience in Bilingual Education since it was one of the first MEC-British Council Project centers back in 2006 until they became part of the CAM Bilingual Program in 2010. This has given the teaching staff a deeper understanding of CLIL methodology, materials and the functioning of a bilingual school and above all, a strong commitment by all members in the bilingual group to work

Otto, A.; Estrada, J.L.

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

in collaboration with each other as will be shown later on.

Although the questions were the same as in the first focus group, before discussing the weight of English in CLIL assessment, the conversation started with the main **assessment tools** they use for CLIL subjects, and the assessment and grading criteria. In this regard, all the teachers indicate they use both open-ended and closed questions: fill in the gaps, multiple choice questions, short questions and answers and essay type questions:

"I usually combine the two: short and essay-type questions. The multiple-choice type and longer questions. And I add images so that they can complete the task with the help of visual support. I do it that way because I know there are also visual students, and they learn this way, I don't want the final grade to be so influenced by the CLIL methodology". (FG2-N)

As can be observed from the quote above, teachers are conscious that the lack of proficiency in the foreign language might hinder the expression of content, and thus apart from traditional essay-type questions, they try to offer some matching or multiple-choice questions in which students can demonstrate content knowledge and skills without being burdened by linguistic issues. Also, in more practical subjects such as Technology or Arts and Crafts, students are asked to solve problems or demonstrate skills. Again, the main goal for teachers seems to be vocabulary knowledge since students are required to master the specific academic vocabulary from a subject:

"There are some questions in which they have to write a definition so that I can see they master the concept, they have understood the subject". (FG2-MO)

Other assessment tools which respondents use in order to give prominence to language in content subjects are **oral presentations**. This is a regular requirement in most subjects since students need to prepare them on a monthly basis whilst some others ask for group expositions once a week. When asked about the criteria to assess oral expositions, teachers agree that the focus lies on content knowledge, presentation skills such as the ability to create a good Power Point presentation, and to address the audience appropriately. Besides, they recognize they assess fluency over accuracy; i.e. they expect students to be able to express themselves with acceptable fluency according to their level although they might make some mistakes or inaccuracies:

"I guess the most important thing is whether they know how to express content knowledge in English. Rather than reading from their cue notes, they have to be able to speak fluently and confidently, and of course, to know the vocabulary". (FG2-S)

Oral presentations are important because they allow students to show understanding of the subject and express it. In relation with content expression, and in order to abandon

memorization in favor of fluency in oral presentations, some teachers also expressed their concerns about the students' need to develop critical thinking and skills as is noticed in Bloom's taxonomy where students can move from LOTS (Low Order Thinking Skills) - remembering and understanding knowledge- to upper-level HOTS (High Order Thinking Skills), in which they are able to apply, analyze, evaluate and create from the knowledge they have acquired:

"Then I can see if they understood a historical fact. I check they were able to understand not just memorize concepts and facts, to understand that a historical fact comes as the result of other direct previous factors. This is the type of knowledge that people in our department acknowledge is difficult to measure by means of a multiple-choice test". (FG2-R)

Another teacher points out:

"The most important thing is the message. The message should be transmitted in a clear way. In this sense, I'd say it is important to demonstrate they understood the main contents, that important information was assimilated. They also have to be able to reflect critically, in terms of cognition". (FG2- E)

In Arts and Crafts, for instance, teachers state that portfolios are used to measure students' progress, but no additional information was offered on the topic. On the other hand, teachers reveal that the use of self and peer-assessment techniques are not current tools yet.

In relation with the selection of assessment tools, no difficulties were highlighted. Nevertheless, teachers noted that they sometimes miss good materials for exams and tests in their textbooks. Although the quality of materials has improved over the past years, some teachers complain that most CLIL materials are translations from Spanish textbooks and consequently, the assessment tools do not serve Spanish CLIL contexts very well.

As regards **informal assessment**, class notebooks are of high importance for teachers in order to check students' daily work. This process of gathering students' pieces of work is rather systematic among teachers in the school. The weight of these assessment tools is set by the department and it is also made public and sent to first and second graders' families at the beginning of the academic year so that both students and parents know about the school' assessment and grading criteria in advance. These notebooks are measured using quantitative marks along with some qualitative comments which students can read and learn from.

Informal assessment, teachers assert, is complemented with other tools such as class observation, checklists, students' behavior and active class participation and interest - known as "attitudinal contents" in Spanish secondary education. Criteria for informal assessment is also set by the department - not the bilingual team - as is common for both non-bilingual and bilingual groups, and it can amount to approximately

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

Otto, A.; Estrada, J. L.

20% of the final mark. According to the data from the teachers' questionnaires, the rest can be obtained by one or more written tests, which shows a big prevalence of written tasks over oral tasks and other forms of assessment.

Moving on to the weight content teachers assign to English in CLIL assessment, as in oral presentations, teachers (overtly) focus on fluency over accuracy but they insist that in production activities, the students' level is taken into consideration:

"In assessment, language is part of the final grade, but the most important aspect is always content, and as such it is considered over the English language". (FG2-L)

Apparently, students with a good command of English do not have difficulties in expressing content knowledge. The problem arises with those students who are less proficient in English and whose final grade can be affected by their English level. It might be the case - they point out - that these students find that the foreign language represents an additional challenge and they could (possibly) obtain better results in non-bilingual programs.

In both oral and written productive skills, some actions and agreements have been made. Contrarily to the criteria in some other schools, where the weight of English in content subjects is clearly specified by each department, some general joint rules have been agreed from the introduction of the so called "Action Plan". This Plan was implemented in the academic year 2014-15 as a strategy to prevent the fossilized errors which teachers observed had started to be rather common among 3rd and 4th graders. The teachers worried that students' language proficiency might be compromised by an overt focus on fluency, and consequently, a group of English teachers supported by the bilingual coordinator met to agree on criteria to grade language mistakes in both English as a foreign language and CLIL subjects so that they could subtract from two to four points in the exam or final mark. Although typical mistakes are the same for all subjects, they are penalized differently depending on whether they occur in content subjects or in English as a foreign language, English teachers being stricter regarding language accuracy. Nevertheless, apart from the criteria in the "Action Plan", teachers point out that some additional factors regarding students' level, effort and attitude are also taken into account. The language mistakes in this plan are the ones which teachers supposedly consider for assessing and marking down students' written output in essays and exams (See **Appendix**).

Finally, another problematic issue was how to deal with **language mistakes** especially during students' oral participation in class and oral presentations. At this point, they asked about European guidelines on this subject matter, at the same time that they insisted on the importance of accuracy, and they pointed out that some errors cannot be overlooked and need to be corrected immediately:

"I have this group, they are the best group in the 4th grade (4º ESO). And then there are these two boys who are so confident, self-assured, they have very fluent English but they make mistakes all the time, so I also need to stop them at times. Otherwise, they would think they are doing it fine and they aren't…" (FG2-F)

About the duality between fluency and accuracy, some teachers clarify it is still fluency over accuracy the criterion that prevails among them, and that they tend to let students talk without correcting unless it is a very serious mistake. One teacher exemplifies her teaching procedure when she describes the way these mistakes can be later retrieved in class and come under scrutiny as in the "Language Clinic" (Coyle, Hood & Marsh, 2010) which, as she points out, is very common practice in this high school. As for the type of mistakes which have been typified in the Action Plan, evidence shows that the focus is on grammatical accuracy, namely correct verb tenses, the obligation to include the subject at the beginning of declarative sentences - a typical mistake among Spanish students- and correct comparative and superlative forms, to name just a few.

4. Discussion

The focus groups offered an in-depth view and understanding of the topic of CLIL assessment in Madrid (Spain), which clearly has the challenge of following the same guidelines that in non-bilingual schools even if the bilingual program deals with a different reality.

4.1. Main assessment tools

According to the data collected, the most frequent assessment tools are exams combining multiple choice and essay type questions, and offering visual support. This emphasis on written exams is not common in Pre-primary and Primary education contexts in other European countries (Serra, 2007; Hönig, 2010) where oral tasks prevail, and specifically avoided in others such as in the German state of Baden Wurttenberg, where students are assessed through oral tasks and activities. However, they are frequent in Upper Secondary Education in Sweden (Reierstam, 2015) because they are easier to grade, and in the Spanish context, mainly due to the predominance of standardized exams in education as compared to other countries (TALIS, 2013). Unlike assessment in some Primary Education CLIL contexts where the testing methods are adapted to the students' level of language development (Zangl, 2000), the testing methods in the context of this study are the same for all type of learners. This is probably because the students in the Bilingual Sections have an advanced level if compared to the students in the Bilingual Program (usually a B1 level in the two first academic years, and B2 in the two last academic years), and because Spanish mainstream education tends to assess students

Otto, A.; Estrada, J.L.

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

uniformly regardless of students' characteristics. On the other hand, class notebooks, consisting mainly of written homework (essays, reflections on experiments, timelines, projects, etc.) are very highly considered among secondary teachers to check students' skills or practical knowledge over time. Likewise, regarding alternative assessment tools, namely self and peer-assessment and portfolios, which are recommended for CLIL contexts as well as by the law in force (LOMCE), timid movements are being made to implement them in content subjects. Nevertheless, their use is still very limited or even inexistent in some schools as is also common in other countries (Hönig, 2010). According to some informal conversations held with teachers after the focus groups sessions were completed, the reasons for not using self and peer- assessment are often relative to the lack of consistency these tools seem to have for teachers, and the students' lack of training in their use. The same can be said about the portfolio, which in contrast with the mere compilation of activities presented in class notebooks typical of the Spanish context, should involve reflection on the part of the students. For the practical implementation of these tools, apart from specific training, the teachers need to accept them as valid assessment tools, and therefore include them in the final grade so that students develop reflection skills, and see their purpose in the subject. Since educational changes and tools are slowly implemented, it is hoped that to compensate for the supremacy of written exams, and to conduct assessment in a formative way, more efforts will be made to include alternative assessment tools in the near future.

> **"For the practical implementation of these* tools, apart from specific training, the teachers need to accept them as valid assessment tools, and therefore include them in the final grade so that students develop reflection skills, and see their purpose in the subject."**
>
> *** self and peer-assessment, and portfolio**

4.2. The role of language in the assessment of content matter

Another significant issue was raised in relation with the role that language plays in the assessment of content matter. Although content teachers recognize language is paramount in the expression of content and skills, they do not consider themselves as language experts, and thus feel they might not be in a position to deal with language-related aspects, as will be discussed further. Language awareness is also observed in the creation of school guidelines for correction and weighing of language due to the absence of official recommendations. In this sense, they insist they focus on academic vocabulary along with grammar, and do not penalize language mistakes unless the message is not clear. However, assessing the language does not necessarily entail that language-related aspects are present in daily teaching practice. In fact, apart from commenting on students' language mistakes in exams from time to time, language is not visible in class as happens in other European CLIL contexts where teachers recognize the relevance of language in daily teaching practice as a preparation for content expression in exams (Reierstam, 2015), and a tool for learning in general. Thus, in the context of our study, even if errors are treated by means of the "language clinic", the objectives teachers present refer exclusively to content and not language, and in the need to compensate for students' deficiencies, teachers opt for simplifying or reinforcing content objectives. This invisibility of language (Llinares et al, 2012) in the class contrasts with the prominence it has in exams, and it shows the lack of alignment between teaching practice and assessment.

The **lack of focus on language** may be attributed to several factors. Firstly, language objectives and tasks are still absent in some CLIL models (Hönig, 2010), and scarce in most CLIL textbooks and materials (López Medina, 2016; Martín del Pozo & Rascón Estébanez, 2015; Kelly, 2010). Secondly, listening and speaking skills still receive little attention in Secondary Education assessment in Spain (García Laborda & Fernández Alvarez, 2011). Thirdly, teachers are usually reluctant to be made responsible for the language in CLIL, a role they think suits the language teacher best. This is also common in other countries such as Slovakia (Gondová, 2012), probably due to their background as content specialists, which usually implies a lack of training in language pedagogies, and because of their lack of confidence in their own language skills (Clegg, 2012). This tendency to overlook language issues and take them for granted can be explained because of the teachers' lack of language awareness (Andrews, 2007; Pavón, 2010). In fact, although content teachers master the topic and the academic registers, they see language as a natural part of the text, and are already trained to using academic literacy, which prevents them to notice the difficulties students might encounter in dealing with academic texts. Besides, another factor impeding language visibility is that, as teachers point out, students have a limited vision of subjects and when content teachers highlight language-related issues, students tend to see them

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

Otto, A.; Estrada, J. L.

as adopting the English teachers' roles. It also seems that students are not used to seeing teachers collaborating with each other, and thus they consider content teachers as the only ones responsible for the subject, which contrasts with the recommendations of subject integration by recent Spanish regulations, and the cross-curricular approach necessary in bilingual education. Teacher collaboration and coordination are, in fact, commonplace in other countries (TALIS, 2013) such as Italy and Austria, where content teachers and language teachers can co-assess the subjects (Serragiotto, 2007; Hönig, 2010).

> **"Although content teachers master the topic and the academic registers, they see language as a natural part of the text, and are already trained to using academic literacy, which prevents them to notice the difficulties students might encounter in dealing with academic texts."**

5. Conclusions

The purpose of this study has been to shed some light on one of the most contested issues in CLIL, assessment, and how it is conducted in practice in the context of Bilingual Sections of the CAM Bilingual Project. This section is divided into two different parts: first, some conclusions are drawn from the results of this research. The conclusions have been contrasted with best practice suggestions from other CLIL contexts, and the informal conversations with teachers and students about the difficulties they face in their daily assessment practices. Second, some recommendations are included concerning assessment practice and the treatment of language issues.

5.1. Main conclusions

As was pointed out in the discussion, despite recommendations about the implementation of formative assessment in CLIL, practices according to the answers from the focus groups demonstrate that assessment is conducted in a summative way. Assessment

does not serve to inform instruction, and the main tools being used to assess students in content subjects still conform to traditional assessment patterns mostly in the form of written tests, leaving communicative language competence behind. Thus, although the impact of CLIL can be observed in aspects such as the increase in the number of oral activities in daily teaching practice, and the implementation of accommodation strategies catering for students with limited foreign language proficiency, this impact is not as evident in relation to assessment practices. Assessment in this study does not exclusively depend on issues suited to Bilingual Education but also on assessment legislation for Secondary Education, which undoubtedly exerts a significant influence on current assessment practices. In fact, the PAU/EvAU exam (the entry exam to access Higher Education) has a big impact on Secondary Education, and it shapes assessment practices (Rodríguez-Muñiz, Díaz, Mier & Alonso, 2016; Zakharov, Carnoy & Loyalka, 2014). Due to this washback effect, CLIL assessment tends to follow the same patterns typical of non-bilingual groups as regards the main assessment tools and exam format. To start with, the EvAU exam in Madrid is conducted in Spanish, a fact that commonly worries teachers, students and families because of the effect that bilingual education might have on content learning, and students' expression in their L1. Second, although attempts have been made to introduce listening tasks in English as a Foreign Language, this entry exam consists predominantly of written tests. Even though Bilingual Education is already well established in the Madrid Region after ten years' experience, these Secondary Education standardized exams are common for both bilingual and non-bilingual groups, a fact which might lead teachers to adopt more traditional approaches suited to the entry exam format to train students accordingly in the long term. On the other hand, regarding the role of language in assessment, this study has evidenced that the foreign language is assessed as separate from content issues, and it is not necessarily linked to the achievement of content-based learning objectives (Mohan & Huang, 2002). Finally, it is also important to stress that the student's language level plays a major role in assessment as the vehicle of expression in most assessment tools.

> **"The main tools being used to assess students in content subjects still conform to traditional assessment patterns mostly in the form of written tests, leaving communicative language competence behind."**

Otto, A.: Estrada, J.L.

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

5.2. Recommendations

Given the lack of research on CLIL assessment, the different CLIL realities among countries, regions and even schools, and the fact that the type of formative assessment recommended for Bilingual Education has not been translated into real practice in some educational contexts, there is an urgent need to create some guidelines for CLIL assessment. What follows is a series of recommendations for improving assessment in CLIL in general, and to deal with linguistic aspects in content subjects in particular so that the language can be made visible along with content knowledge and skills.

Previous research on CLIL has concluded first that assessment should be conducted in a formative way, by means of carefully selected assessment tools depending on the learning goals. Second, that regardless of the treatment given to the language in CLIL, linguistic elements are paramount in the expression of content and skills and as such, they cannot be separated from content. The present study agrees with previous findings in all these regards. However, as CLIL is an umbrella term covering a broad range of scenarios, for adequate assessment in CLIL, the particular context in question should also be taken into account. The following guidelines are suited to the Bilingual Sections in the CAM Bilingual Project:

1 **Specific guidelines and policies for Bilingual Education are urgently needed** given the fact that the general ones from the Ministry of Education and the Madrid Regional Government refer to mainstream education and as such, they are insufficient for the reality of assessment in Bilingual Secondary Education. These guidelines might come from the educational administration or in their absence, the secondary schools in the CAM Bilingual Project could agree on a model and basic CLIL guidelines to deal with assessment in general, and the role and weight of the vehicular language in particular.

2 **Assessment should mirror daily practice. The type of exams (if any) and the questions in them should be similar to the ones students deal with on a daily basis in that they are rooted in real life.** In this regard, more innovative assessment tasks in line with formative assessment are needed for a variety of reasons: first, to abandon the prevalence of the traditional exam, which does not always allow the integration of competences in real-life, in favor of more task-based learning using for instance portfolios and journals. Second, to allow the students to show content knowledge and skills in a meaningful way, focusing not just on the final product but also on the process. Third, to assess language "for a real purpose in a real context" (Coyle et al., 2010: 131). Likewise, although oral tasks are already implemented in the CLIL lessons, more efforts should be made to include them in assessment practice and thus, to give them more weight in the final grade.

3 **If language production is still so present in CLIL assessment tools, as is the case in Social Sciences, maybe more writing components such as clause-linking strategies, nominalization and cohesion can be included as part of the curriculum planning** (Boscardin et al., 2008: 7). These genre-based activities which are aimed to make the linguistic structures of academic language explicit to students need to be stressed by content and language teachers, and ideally reinforced by language assistants.

4 As content teachers' opinions reveal the lack of language and CLIL pedagogies typical of content teachers' background (Dalton-Puffer, 2013), **more teacher training is needed in the context of the study to give the language aspects the importance they deserve.**

5 In this scenario of traditional standard exams, and the lack of CLIL curricular guidelines for real integration of content, language and skills, **more efforts are clearly needed so that content and language teachers work in collaboration with each other.** Collaboration among teachers is recommended in the current educational law (LOMCE, 2013) as one of the signs of an effectively integrated and integrative curriculum, and by CLIL research. (Pavón & Ellison, 2012; Kelly, 2014; Otto, 2017) ■

6. References

Andrews, S. (2007). *Teacher Language Awareness.* Cambridge, UK: Cambridge University Press.

Assessment Reform Group, The (2012). *Assessment for Learning: Ten Principles.* Retrieved from http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/doc/Afl_principles.pdf

Ball. P., Kelly, K. & Clegg, J. (2015). *Putting CLIL into Practice*. Oxford, UK: Oxford University Press.

Boscardin, C. K., Jones, B., Nishimura, C., Madsen, S., Park, J. (2008). *Assessment of Content Understanding Through Science Explanation Tasks, CRESST Report 745*. Los Angeles, USA: University of California Graduate School of Education and Information Studies, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Clegg, J. (2012). Preface. In Quartapelle, F. (ed.) *Assessment and Evaluation in CLIL* (pp. 5-7). Retrieved from: www.aeclil.net.

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

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

Otto, A.; Estrada, J. L.

Dalton-Puffer, C. (2013). A Construct of Cognitive Discourse Functions for Conceptualising Content-language Integration in CLIL and Multilingual Education. In *European Journal of Applied Linguistics*, 1(2), pp. 216-253. DOI 10.1515/eujal-2013-0011

García Laborda, J. y Fernández Álvarez, M. (2011). Teachers' opinions towards the integration of oral tasks in the Spanish University Examination. In *International Journal of language studies*, 5(3), pp. 1-12. Retrieved from https://files.eric.ed.gov/fulltext/ED523843.pdf

Gondová, D. (2012). Identifying the Needs of CLIL Teachers at Lower and Upper Secondary Schools. In *Journal of Interdisciplinary Philology*, 3(3), pp. 5-22.

Hönig, I. (2010). *Assessment in CLIL: Theoretical and Empirical Research.* Saarbrücken, Germany: VDM Verlag Dr. Müller.

Järvinen, H. M. (ed.), (2009). *Handbook: Language in Content Instruction*. Turku, Finland: University of Turku. Retrieved from https://broadyesl.files.wordpress.com/2012/09/language-incontent_handbook_en.pdf

Kelly, K. (2010). CLIL in natural science subjects: Language and task. In *Ikastaria*, 17, pp. 91-108. Retrieved from https://core.ac.uk/download/pdf/11503455.pdf

Kelly, K. (2014). *COOP E-CLIL*. Retrieved from https://www.teachingenglish.org.uk/sites/teacheng/files/Language%20teachers%20working%20with%20subject%20teachers%20-%20TE.pdf

Kiely, R. (2009). *CLIL: The Question of Assessment.* Retrieved from https://www.unifg.it/sites/default/files/allegatiparagrafo/20-01-2014/kiely_r._clil_assessment.pdf

Krueger, R. A. (1994). *Focus Groups: A Practical Guide for Applied Research* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Krueger, R. A., & Casey, M. A. (2009). Focus Groups: *A Practical Guide for Applied Research* (4th Ed.). Thousand Oaks, CA: Sage Publications.

López-Medina, B. (2016). Developing a CLIL Textbook Evaluation Checklist. In *Latin American Journal of Content & Language Integrated Learning*. Vol. 9(1), pp. 159-173. doi:10.5294/laclil.2016.9.1.7

Llinares, A., Morton, T. & Whittaker, R. (2012). *The Roles of Language in CLIL*. Cambridge, UK: Cambridge University Press.

Martín del Pozo, M. A. & Rascón Estébanez, D. (2015). Textbooks for Content and Language Integrated Learning: policy, market and appropriate didactic. In *Foro de Educación*, 13(18), pp. 123-141. Retrieved from https://dialnet.unirioja.es/descarga/articulo/5153348.pdf

Mohan, B. & Huang, J. (2002). Assessing the Integration of Language and Content in a Mandarin as a Foreign Language Classroom. In *Linguistics and Education* 13(3), pp. 407–35.

OECD (2014), *TALIS 2013 Results: An International Perspective on Teaching and Learning*, TALIS, OECD Publishing, Paris.

Otto, A. (2017a). Assessment Issues in CLIL: What You've Been Wondering but Were Afraid to Ask. In *RDIM*, 1, pp. 1-13. Retrieved from https://app.rdim.es/archivos/publicacion/e8a4d38d3b45a4831eebde1c2dcbe47e.pdf

Pavón, V. (2010). The introduction of multilingual teaching in Andalusia: heading towards a newly proposed methodology. In *Journal of Border Educational Research*, 8(1), pp. 31-42.

Pavón, V. & Ellison, M. (2013). Examining Teacher Roles and Competences in Content and Language Integrated Learning (CLIL). In *Linguarum Arena*, 4, pp. 65-78. Retrieved from http://ler.letras.up.pt/uploads/ficheiros/12007.pdf

Rodríguez-Muñiz, L. J., Díaz, P, Mier, V., Alonso, P. (2016). Washback Effect of University Entrance exams in Applied Mathematics to Social Sciences. *In PLoS ONE 11*(12) e0167544. DOI:10.1371/journal.pone.0167544.

Reierstam, H, (2015). *Assessing Language or Content? A comparative study of the assessment practices in three Swedish upper secondary CLIL schools.* (Licentiate thesis in Subject Matter Education at the Department of Education and Special Education). Gothenburg, Sweden: Faculty of Education, University of Gothenburg. Retrieved from https://gupea.ub.gu.se/bitstream/2077/40701/4/gupea_2077_40701_4.pdf

Serra, C. (2007). Assessing CLIL at Primary School: A Longitudinal Study. In *International Journal of Bilingual Education and Bilingualism*, 10:5, pp. 582-602. DOI: 10.2167/beb461.0

Serragiotto, G. E (2007). Assessment and Evaluation in CLIL. In D. Marsh & D. Wolff (eds.), *Diverse Contexts-Converging Goals: CLIL in Europe*, pp. 271-286. Frankfurt, Germany: Peter Lang.

Vaughn, S., Shay Schumm, J. & Sinagub, J. (1996). *Focus Group Interviews in Education and Psychology.* California, USA: SAGE.

Wewer, T. (2014). *Assessment of young learners' English proficiency in bilingual content instruction CLIL* (Doctoral Dissertation). Turku, Finland: University of Turku. Faculty of Education/Department of Teacher Education. Retrieved from https://www.doria.fi/bitstream/handle/10024/96838/AnnalesB385Wewer.pdf?sequenc2

Zakharov, A., Carnoy, M., Loyalka, P. (2014). Which teaching practices improve student performance on high-stakes exams? Evidence from Russia. In *International Journal of Educational Development*, 36, pp. 13-21. Retrieved from https://www.hse.ru/mirror/pubs/lib/data/

Otto, A.; Estrada, J.L.

Towards an Understanding of CLIL in a European Context:
Main Assessment Tools and the Role of Language in Content Subjects

access/ram/ticket/28/15453029081db2971d3120c607b98f41
1215cb382f/1-s2.0-S0738059314000078-main.pdf

Zangl, R. (2000). Monitoring Language Skills in Austrian Primary (Elementary) Schools: A Case Study. In *Language Testing*, 17(2), pp. 250-260. DOI: 10.1191/026553200675232246.

Appendix

Action plan for language mistakes

Action Plan for correcting grammatical errors.

BILINGUAL PROGRAM
1st and 2nd ESO
■ Omission of the subject in a sentence
■ Subject-verb agreement
■ Correct use of "there is/there are"
■ Correct use of verb tenses, particularly of irregular verbs
■ Correct use of possessive forms
■ Correct use of the auxiliary verbs "do/does/did" in interrogative and negative sentences
■ Correct use of WH-questions
■ Correct use of "some/any"
■ Correct use of demonstratives (this-that-these-those)
3rd and 4th ESO
■ Comparatives and superlatives
■ Verb tenses (present/past/perfect tenses)
■ Modal verbs
■ Relative pronouns or adverbs
■ Linking words
N.B. For each mistake in an exam, 0,10 will be deducted up to 1 point

BILINGUAL SECTION
1st and 2nd ESO
■ Omission of the subject in a sentence
■ Subject-verb agreement
■ Correct use of "there is/there are"
■ Correct use of verb tenses, particularly of irregular verbs
3rd and 4th ESO
■ Omission of the subject in a sentence
■ Subject-verb agreement
■ Correct use of "there is/there are"
■ Correct use of verb tenses, particularly of irregular verbs
■ Correct use of the auxiliary verbs "do/does/did" in interrogative and negative sentences
■ Correct use of WH-questions
■ Correct use of demonstratives (this-that-these-those)
■ Relative pronouns
N.B. For each mistake in an exam, 0-10 will be deducted up to 2 points