PREPRINT

In-service Training for Bilingual Implementation at the University

Mary Griffith, Universidad de Málaga, Spain

Abstract: Research suggests there is an urgent need for a research-driven approach that measures the complex processes involved in bilingual instruction and its effects both on the learning of academic subjects and on the acquisition of English proficiency. These complex processes deal with language, instructional styles, and most importantly, key stakeholders and their training. Universities face a confounding paradox: an educational institution in which teacher training plays a small role. This study approaches the professional development of bilingual instructors in a Computer Science department directly from the context of their classrooms. Focusing on collaborative professional development, this article examines an in-service training project in which participatory action research provides meaningful experiences that add to good teaching practice.

Keywords: Bilingualism, Professional Development, English-Medium Instruction (EMI), Higher Education, Computer Science, Action Research

Introduction

Convergence in higher education began when European universities outlined the desire to promote greater plurilingualism across Europe (Sorbonne Joint Declaration 1998). Since that time, research has called for a more research-driven approach that would consult key actors to underscore the complex processes involved in content and language integrated learning (CLIL). The idea is to examine the effects of bilingual instruction both on the learning of content subjects as well as on the acquisition of English proficiency (Dearden 2014; Griffith and Lechuga 2018). The research frame of this study is in-service training for university professors teaching Computer Science through a second language.

While studies show the importance of follow-up in bilingual instruction, particularly in regard to language support, they rarely assess content professors' needs and expectations. By focusing solely on language, many studies tend to remain distant from real teaching contexts for content instruction at the higher educational level. There needs to be more relevant support for content professors who are non-language specialists. In fact, experts identify this approach as a "blueprint for action designed to be adaptable to the environment in which the university operates" (Marsh, Pavón-Vázquez, and Frigols-Martín 2013, 9). Nevertheless, in higher education, some authors have found that neither institutional language integrated support nor a specific language plan is provided by many bilingual programmes (Barrios, López-Gutiérrez, and Lechuga 2016).

This study is collaborative and began with a dialogue between a language specialist and eight Computer Science specialists who were teaching their subjects through a second language. Seeking to consult the key actors—the content professors—we wanted to explore the professors' needs when faced with the challenges of bilingual instruction. At the same time, the in-service training provided a qualitative assessment of their instruction through a second language. This study explores student responses to bilingual instruction.

Our working hypothesis differs from a traditional hypothesis in that it seeks not only plausible and logical outcomes, but more importantly, practical outcomes. Is effective bilingual instruction measured by only by language performance or can it be measured by instructional aspects? Is effective content teaching equated with effective language teaching? The aim is to

create relevant feedback to support these non-language experts in their endeavour to teach through the medium of English. What we suggest is that perhaps language experts need to be willing to adapt to the specific needs of content professors in this demanding higher educational context to look beyond language learning assessments.

This study continues in the line initiated by Griffith (2017) where the participants form an active part of the research process. The Computer Science professors' needs are at the center of the discussion for this in-service training and these instructors actively participate in the creating of inquiry. Throughout the research process, the content professors shared insights about their professional practice and allowed for a greater validation of the findings.

We have selected a degree outside language areas because we feel less language-centered areas of instruction are missing in many of the discussions about practical bilingual implementation at the university level. We hope to discover how Computer Science contents can activate foreign language learning. The purpose of this research project has been to inform practice and to link student and professor performance in order to assess professors through student data.

It is this student data that will be the focus of this article, and three research questions will provide the framework to assess instruction. First, the interaction data will provide the evidence for process learning by quantifying the student interactions across nine subjects. Next, learner outcomes will provide performance evidence by examining how student results compare with control groups taught in the first language. And finally, students will rate themselves and their instructors after the bilingual experience to provide affective evidence. The qualitative assessment will then be inserted into a template in the final analysis to provide evidence for the program's assessment (Coyle, Hood, and Marsh 2010).

Clarifying Terminology

To effectively combine theory and practice, we begin by clarifying terminology. Most researchers distinguish English-medium instruction (EMI) from CLIL. Theoretically, there is a clear conceptual separation between EMI and CLIL:

> Whereas CLIL is contextually situated (with its origins in the European ideal of plurilingual competence for EU citizens), EMI has no specific contextual origin. Whereas CLIL does not mention which second, additional or foreign language (L2) academic subjects are to be studied in, EMI makes it quite clear that the language of education is English, with all the geopolitical and sociocultural implications that this may entail. Whereas CLIL has a clear objective of furthering both content and language as declared in its title, EMI does not (necessarily) have that objective. (Dearden 2014, 4)

EMI focuses on content assimilation and is instructional, while CLIL focuses on teacher-student discourse and is more interactional than EMI. We have found that bilingual instruction in practice does not exactly fit into this classification, for although the focus in higher education is content assimilation, we encourage interaction as a means to this end. Further, although all teaching is instructional, neither CLIL nor EMI can truly escape the linguistic issues inherent to teaching through a second language (L2) more commonly known as form focus (Long 1991; Dalton Puffer 2007). Bilingual teaching is the combination of instruction and interaction.

Theoretical classifications fail to show the overlap between EMI and CLIL because they fail to represent the complexity of the bilingual classroom. Brinton, Snow, and Wesche (2011) go so far as to classify CLIL typologies at the university level. These authors establish three types: theme-based, sheltered, and adjunct CLIL, and debate whether or not language learning is explicit, implicit, or integrated. Any instructor using a second language in their classroom could argue that all three co-occur.

A review of the literature includes many researchers who have provided key definitions from a wide range of research themes. There is a clear correlation between CLIL and EMI in that our context is higher education and the professors are teaching their contents in English. However, we prefer the wider term CLIL because CLIL has a "purposeful objective of integrating both content and language" (Dearden 2014, 4). Brinton, Snow, and Wesche (2011) and Dearden (2014) have underscored key differences between CLIL and EMI. Llinares, Dafouz, and Whittaker (2007) have pointed out how teaching through a second language differs using instructional methodology. Moving away from the traditional linguistic focus on form, Gajo (2007) has highlighted content-subject competence, while Griffith (2017) has suggested that any project in bilingual education deals with didactics and people as much as it deals with language.

It is worthy to note that "to date as most of the published research confirms, the driving force to integrate language and content in higher education is clearly one sided and comes mainly from linguists, language teachers and teacher educators" (Smit and Dafouz 2012, 8). Most of the research focuses on language learning or top down language policies. However, this study approaches instructional challenges for bilingual instruction from the content instructor's point of view. In fact, this shift is of great interest to content professors when they see that their specific concerns and their complex contents are pushed to the forefront. Their willingness to integrate content and language in their classrooms is crucial to many of the internationalization policies at universities across Europe. So indeed, we have shifted the linguistic focus of most research in bilingualism away from language, towards didactic and affective data. It has not been our intention to "convert" these Computer Science professors into linguists; rather we have sought to collaborate effectively in their teaching practice with relevant input about teaching content through a second language directly from the context of their classrooms (Griffith 2017).

Thus, this study is inserted into the context of language learning, but points to instruction and the actors that enable this instruction to take place. The project began with the proposal to capacitate "professors teach though a second language in a purposeful search for quality instruction" (Griffith 2017, 136). Through the reflection in action, the in-service training encouraged instructors to discover solutions to their instructional challenges. "It's about working towards practical outcomes, and also about creating new forms of understanding" (Reason and Bradbury 2001, 10). Consequently, collaboration is key to the methodological frame.

Method

Several aspects separate action research from other types of research. First and foremost is its focus on using the people involved as researchers. It would seem that the participants "learn best, and more willingly apply" what they have experienced themselves (O'Brien 2001, 1). In fact, it is undeniably relevant to professional practice. For the reflective practitioner, action research is an ongoing process to take informed action (Sagor 2000). The limitations of action research suggest that it is challenging to replicate and non-generalizable. However, while this may be true, the method can be validated through triangulation and is widely accepted in much of the supporting literature (O'Brien 2001; Denzin 2006; Castro Garcés and Martínez Granada 2016; Griffith and Lechuga 2018). By seeking greater relevance for the Computer Science professors and their professional practice, it has been our goal to assess the bilingual program using more than mere language parameters.

Action research usually begins with a problem to solve. In that bilingual instruction is inherently "imperfect," we begin by looking for new ways to assess quality and follow up on this instruction. Quality bilingual implementation cannot be measured by ideal models; nevertheless, this quality must be sought after. In fact, we are searching for ways to measure bilingual instruction by inserting language performance into instruction itself. Research has suggested that bilingual instruction at the university requires a closer follow-up (Fortanet-Gomez 2012). Without limiting the discussion to ideal speech models, this study has assessed bilingual instruction using three different measures.

The dilemma surrounding ideal models of perfect speech is that these patterns tend to measure what "goes wrong" instead of what "goes right." In fact, our results hint that perhaps the "worse" speaker was in fact the "best" instructor. Consequently, "true science" in language assessment aims to measure the deviation from a utopian model that does not, or cannot, exist. While this may work for empirical science, it is not as applicable to research in education because education is an intricate process of ever-changing variables that cannot be held constant. The complex, unrepeatable classroom experience is an extremely valid and useful tool for professional development. Bilingual instruction integrates both content and foreign language and, like all instruction, is never perfect. "Educational success resides in the movement towards the ideal, not in the deviation from this ideal model" (Griffith 2017, 136).

Indeed, this movement towards improved bilingual instruction is the ultimate CLIL objective. This functionality in the foreign language, together with content assimilation, is how the program will be assessed. These measures include: student interaction, student outcomes, as well as student response to the CLIL experience. In turn, the research purpose encompasses the creation of a more reflective practice whereby participants think about what they are doing so as to engage tacitly in this process of continuous learning (Schön 1987).

Data Sample Overview

This section will examine the student sample as well as provide an overview of the three different research questions as they relate to the data collection process to assess this bilingual program. The main focus at this time is how student performance is linked to assessment in bilingual instruction. As previously mentioned, this study continues in the line initiated by Griffith (2017) where the participants form an active part of the research process.

The stakeholders in the research project have been the foreign language specialist/researcher, the professor participants, as well as the students in nine different subjects. Instructors include eight experienced professors who had never taught their subjects in English, while students include both freshman and sophomores who opted to take nine different courses in English. The student groups remained largely constant throughout the assessment, which isolated professor performance more effectively.

Previous research has yielded process evidence using interaction as the main variable and this data will now be analysed qualitatively in correlation across the nine subjects to emphasize how new questions are surging. Next, performance evidence will be assessed with students' final results as compared to the same subjects taught in the first language (L1). Finally, affective evidence will be illustrated through a student survey. By exploring CLIL in practice in a highly contextualized cross sectional case study, this approach has contributed to professional development. Table 1 highlights the evidence collected together with the type of data and analysis.

Table 1: CLIL Template

Evidence Type	**Data**	**Analysis**
Process Evidence	Student Interaction	Quantitatively/Qualitatively Correlated with Context
Performance Evidence	Student Results	Comparative with First and Second Language Groups
Affective Evidence	Student Survey	Qualitative Statistical/Qualitative

Note: Evidence tied to data collection in post hoc analysis, University of Málaga.
Source: Adapted from Coyle, Hood, and Marsh 2010

In their template to evaluate CLIL programs, Coyle, Hood, and Marsh (2010) cite four different evidence features, three of which have been included in our own assessment:

- Process evidence, as measured by student interaction.
- Performance evidence, as measured by learning outcomes.
- Affective evidence, as measured by learners' global assessment of the CLIL experience.

There are many ways to approach CLIL program assessments, but by "taking performance and affective evidence together, we can aim for a fuller cross referenced portfolio using a range of students across the ability range" (Coyle, Hood, and Marsh 2010, 137). We have used different evidence features to validate our assessment of this bilingual program through triangulation (Denzin 2006). Using the three research questions, the evidence features are student interaction, student results, and student opinion.

Process Evidence: Student Interaction

Research has demonstrated that on many occasions, content teaching and foreign language teaching pursue different aims (Barron 2002; Neumann, Parry, and Becher 2002; Squires 2005). This is evident when interaction is examined. With regard to interaction as a measure for process learning, the content professors' point of view proved enlightening. Logically, content specialists have their own expectations when teaching through a second language; we have observed that their concerns are highly pragmatic. Their attitudes and their motivations differ from a linguist's; in fact, this is to be expected when specialists from different areas work together. Findings reveal that interaction is an appropriate measure for language learning, but is less useful to assess content classes.

Every classroom observation quantified the student interactions and the results were presented globally to the professors' group. Qualitatively, professors were invited to reflect on the variables of interaction that are outside their control, such as the contents or student attitudes, as well as the variables of interaction that are within their control, such as how they correct students and present materials. Data reveals a striking difference among the sessions in a total number of interactions ranging from six to eighty-one interactions with the same student group (Griffith 2017), pointing to the fact that interaction needs to be re-examined.

As linguists, we consider interaction to be process evidence for comprehension and a valuable measure for bilingual instruction. The cross-comparative aspect revealed a range of student behavior as related to professor behavior and allowed professors to reflect on their own teaching as well as on their colleagues'. Measuring content classes through interaction is a linguistic concern, while at the same time it provides proof of process learning for content assimilation. Interaction proves to be an excellent measure for both language learning as well as content learning. Nevertheless, content professors called for additional measures as a guarantee of quality instruction.

Performance Evidence: Comparative Learner Outcomes

Clearly, the outcomes of learning are the most objective assessment of professor performance. In fact, the pilot study, as well as the diagnostic survey for this in-service training, revealed specific concerns about evaluation through a second language. The department head was particularly concerned with how these groups in English (L2) would compare with groups taught in Spanish (L1). In consequence, students' final results also formed part of professor assessment. Student data from the first language control groups was provided by the department so that a more objective result could be included in the instructional assessment.

Comparative learner outcomes have been examined across eighteen subjects, nine in Spanish (first language) and nine in English (second language), and over 200 results have been examined. The results are clearly indicative of performance evidence for teaching, but the external variables undoubtedly influenced results. Factors such as the class size and the students' capacity play a role; what's more, these particular students have been described as "exceptionally talented and motivated" by their professors (Griffith 2017). We suggest that bilingual classes should be qualitatively compared with students' results in the first language, but again, this cannot be the only assessment. This measure is effective for content assimilation but less so for the second language.

Affective Evidence: Students Assess Instructors and Themselves in CLIL Experience

In order to have a more weighted analysis, we included the students' reactions to compensate for the results from interaction and learner outcomes. There is an interesting twist to CLIL in that most often non-native teachers are teaching non-native students (Dalton Puffer 2007), and both content as well as language learning must somehow be measured. Affective data has been collected by an anonymous student survey whereby each student could assess each professor making them ideal assessors of their professors' performance. Forty-two surveys have been collected for nine different subjects and have yielded unique insight on student CLIL awareness, providing the final validation for this bilingual program. For example, students were well aware which instructor had the "best" English and they were able to separate this skill from the instructor they considered the "best" instructor. Perhaps the most significant point is that effective CLIL instruction has much more to do with proper teaching than with perfect language performance. These results will be expanded upon in the next section. The survey was designed to evaluate both language performance as well as instruction and provides qualitative insight from both a linguistic as well as an instructional point of view.

Data Analysis

There is a clear link between student and professor performance and it is this student data that will be highlighted in this section. Professor performance can be assessed with student interaction, with learner outcomes, or with student surveys. We have chosen to present the multiple variables that influence performance using a comparative analysis in order to triangulate the findings.

Pointing back to the focus of this article, we will now examine the results of the data collected regarding our three research questions. Preliminary analysis for the first two questions can be found in Griffith (2017) and will be expanded upon here briefly. The first aspect will discuss student interaction in nine subjects; the second will compare final evaluation results between the first and second language groups in eighteen subjects; and the last will provide a detailed student assessment of the CLIL experience in the Computer Science department from forty-two surveys. In this qualitative analysis, we will outline the external and internal variables that impact student performance in a comparative assessment.

External Variables

The two external variables that we could not isolate were the class size and the students' capacity. These elements will undoubtedly influence interaction as well as final results. When comparing the English and Spanish groups, one of the main differences is the class size. The bilingual groups were much smaller, allowing for more active participation during class sessions (Griffith 2017). In this context, process and performance variables are related to results, but never completely generalizable.

What we would like to emphasize is that correlation between student teacher ratio and learner performance is significant. If we focus solely on class size, we would highlight the students' answers reflect a more enriched academic environment. Students felt the small class size was an advantage; they thought teachers were somehow more involved with the students and they mentioned the classmates had a better relationship.

In addition to smaller class size, we also noted the individual capacity of students clearly would affect student results. In turn, the second external variable is student capacity. The student sample who opted for these CLIL classes had consistently higher average entrance scores as assessed by the Spanish university access exam. "Consequently, if they began the course already with an advantage, perhaps it is unwise to claim their overall results as part of the CLIL qualitative assessment" (Griffith 2017, 139). We reiterate that our intention was to illustrate that students could learn directly in English and in turn, be assessed in English.

Internal Variables: Process, Performance, and Affective Evidence

In their discussion of assessment for CLIL programs, Coyle, Hood, and Marsh (2010) refer to specific measures for bilingual programmes. In the data collected, we will address process, performance, and affective data to discuss the internal variables. Each of these evidence types has been correlated to a specific research question. In consequence, these three research questions will verify the final assessment for this program. First, will Spanish students interact in a CLIL class? Next, how will CLIL student results compare with control groups taught in the first language? And finally, we examine how students rate themselves and their instructors after the CLIL experience for the overall assessment of the program.

Process Evidence: Do Students Interact?

Gajo (2007) has discussed language knowledge based on classroom interaction. He pointed out that interactional discourse in the classroom is an interface between the language and the content. Associating the exchanges with what teachers do to encourage more and better communication in the classroom, we have both quantified and qualified interaction. The focus has not been on the language itself, but rather on the instructors' ability to incite effective exchanges. In turn, this evidence feature is particularly related to process evidence both for content assimilation as well as for language usage (Griffith 2017).

Pointing to some of the drawbacks in collaboration, we note that traditionally science or "hard" subjects are teacher-centered and humanities or "soft" subjects are learner-centered. There is theory and practice in all instruction, but individual teaching styles play a role as do the contents themselves. Squires (2005, 130) has noted the difference in applied (or practical) courses where "the professions differ from other disciplines in being concerned primarily with *acting* rather than knowing." Consequently, "hard" subjects are usually given in a lecture style with little interaction between teachers and students and tend to focus the contents around knowing. In fact, one of the main differences is the focus on student interaction in more practical courses. While our findings recognize the nature of university style lectures will continue to dominate in many subjects, we do associate interaction in the classroom as proof of process learning.

In this study, we ask whether or not communication in the classroom can be an effective measure for both language and content learning when combined with other measures. In one two-hour session, the maximum number of exchanges by students was eighty-one, while the minimum was six (Griffith 2017). What we suggest is that communication lies at the core of all instruction and this interaction can be an effective measure both for language learning as well as for content learning. The professor in-service training used interaction to assess professor performance and framed these exchanges as a didactic aid. The overall student interactions are shown in Figure 1.

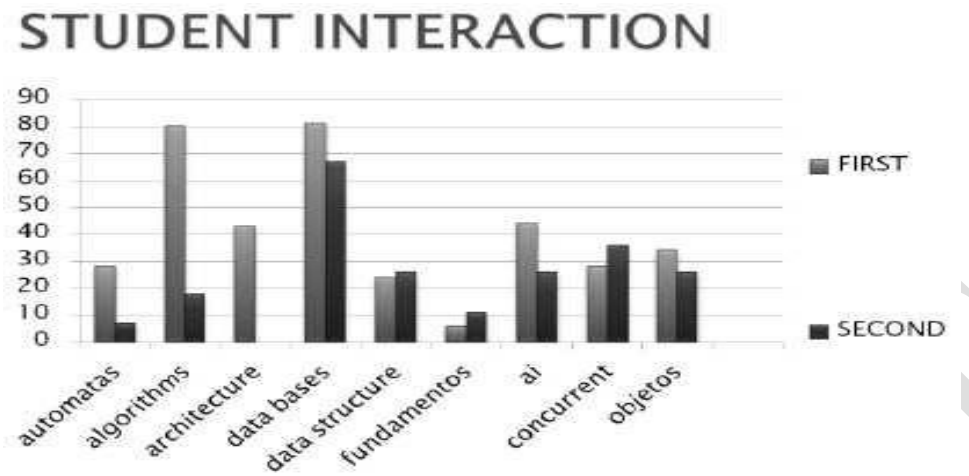


Figure 1: Student Interaction in Nine Subjects, University of Málaga
Source: Griffith 2017, 140

This type of analysis is particularly challenging when taken cross-curricular. Research has demonstrated that on many occasions content teaching and language teaching have different objectives (Barron 2002; Neumann, Parry, and Becher 2002). We examined student groups, contents, and professors' teaching style using interaction as one of the measures for content assimilation. When put into context, an enormous range of communicative exchanges occurred. However, the variables are difficult to isolate to reach any one conclusion. Without the qualitative focus, data is inconclusive when taken out of context. Should we measure professor performance, student performance, content assimilation, language performance, or all of the above? A more inclusive, qualitative approach allows for a more weighted analysis and a better description of the complex context.

Regarding student performance, there are significant differences when comparing the freshman and sophomore student groups. Findings show an increased participation in the first year when comparing two programming courses, "Fundamentals of Programming" with "Programming with Objects," offered in different semesters. These findings suggest a student's progression over time, but it is unclear if this may well be related to the students, to one of the instructors, or even to the contents themselves.

In addition, this study revealed many features about teaching style, but these features remain challenging to isolate. When examining one instructor in two different subjects with a similar student group, our findings do not support a strong influence of a consistent teaching style. For instance, when comparing "Data Bases" with "Concurrent Systems," this same instructor had different results with different contents, suggesting the contents themselves have an impact on student interaction. So once again, our findings, rather than fully answering our questions, lead us to new ones. Does interaction depend on language performance, per se, or does it depend on contents or even teaching style? More research is warranted to measure this more effectively.

Finally, regarding course contents, results showed that the more practical sessions yielded higher interaction, while the more theoretical math classes such as "Automatas" or "Fundamentals of Programming" had fewer interactions. But perhaps more than the contents, it is the university context that determines the instruction. Even when maintaining the student group constant, a lecture in the subject "Algorithms" showed only seventeen exchanges as opposed to a practical session with eighty interactions (Griffith 2017). Even though language teachers use student participation as proof of learning, content professors must necessarily lecture. Research in higher education has suggested that the nature of university lecture is part of the context; not all classes can be interactive.

The goal has been to prove that students will interact if given a chance and to present this data as process evidence comparatively across the nine subjects. Interaction is not just a way to measure language performance. It is a way to explore the role of student motivation, to reveal the difficulties in certain subjects, as well as to examine more fully what professors do to encourage more effective communication in their classrooms. Through this reflection in action, the Computer Science professors were surprised that classroom interaction could validate quality instruction, just as the language specialist was surprised they had not considered it an essential part of instruction (Griffith 2017). Each side had something to learn.

Perhaps the most significant finding was the realization that foreign language teachers seem to measure quality instruction with interaction, but outside language teaching, instructional assessment must use different values. The content professors began to see interaction as a valuable instructional tool both for content assimilation as well as for language learning. The language specialist was forced to step back from her original assumption. Good language teaching uses interaction as a way to assess effective strategies, while good content teaching looks to assess instruction using learning outcomes.

Performance Evidence: Student Results Compared

The second research question deals with performance evidence and seeks to assure content. In this section, we will discuss the students' final results across the nine bilingual subjects and compare the learner outcomes to the other nine groups taught in the first language. The faculty dean requested this comparison between the Spanish group (L1) and the English group (L2) to guarantee the results from this first year of teaching in English. For qualitative assessment of the bilingual instruction, the final results of the nine subjects should be comparable to the same subjects taught through the first language. Findings show that all students in CLIL classes performed equally or better in the final evaluation results as seen in Griffith (2017).

This analysis takes an integrated approach; undoubtedly, foreign language performance is related to the final results, but it is not limited to this feature. The fact that contents can be actively assimilated and measured as separate from the language competence is significant for content classes in Computer Science. Cognitive research underscores that the idea is separate from the code with which it is represented. What is significant here is that according to the subject material, language issues play varying roles. For instance, in Computer Science, it has been relatively easy to separate language form from content issues, whereas in Law and other Social Sciences, this separation is more difficult. Findings reveal that it is crucial to measure content assimilation as superior to language assimilation in the university context. The focus on the language form will depend on each unique instructional situation.

Similar research at this same university has revealed that both professors, as well as students, face foreign language challenges (Barrios, López-Gutiérrez, and Lechuga 2016). Our findings do not corroborate this in the Computer Science department but did reveal this same challenge in other departments such as Tourism, Economics, and Law. At this time, we have followed up on this discrepancy and point out that each department has a unique set of variables and can conclude that active language support will be crucial to successful bilingual implementation. What each department needs may vary considerably.

In Figure 2, the comparison of the nine subjects offered in both English and Spanish shows the significant "degree of internal consistency across all subjects" (Griffith 2017, 140). Our findings corroborate Fortanet-Gomez (2012, 58), who has noted that the university context requires a "closer follow-up." Following this author, we do urge that bilingual instruction be held to the same standard of quality as its monolingual counterpart. The next section will complete the validation of the program assessment with more qualitative data.

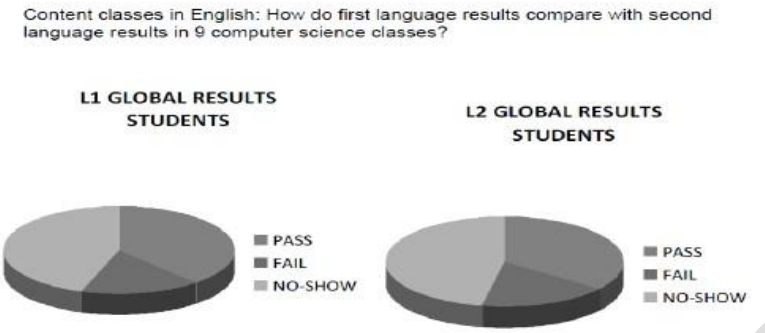


Figure 2: Comparative Exam Results across Nine Subjects Taught in First (L1) and Second Language (L2)
Source: Griffith 2017, 140

Affective Evidence: A Student Survey

In this section, we will continue to explore how to assess bilingual instruction through student data. The third and final research question refers to the data collected in a student survey. After receiving their final marks, students were asked to fill out an anonymous online questionnaire to assess learning and evaluation in this CLIL experience. Variables included student capacity, information regarding language performance for both students and professors, information regarding evaluation, as well as the overall assimilation of course contents. Forty-two students participated in the survey designed to explore the students' perspective of this CLIL experience. Several filter questions allowed us some additional information about student capacity, not only in L2 but also overall academic prowess.

Question 4 (Figure 3) allowed data to be extracted by subject so that teachers could receive any specific comments related to their particular course. We see a much higher response rate of surveys from the first-year students; 41 percent of the total sample for only two subjects. The remaining 59 percent of respondents came from the seven second-year groups. We will elaborate on the findings globally in the subsequent sections.

STUDENT SURVEY CLIL EXPERIENCE

4 WHAT IS THE COURSE YOU ARE EVALUATING?

- FIRST SEMESTER_Fundamentals of Programming 24%
- FIRST SEMESTER_Data Bases 14%
- FIRST SEMESTER_Data Structures 12%
- FIRST SEMESTER_Algorithms 7%
- FIRST SEMESTER_Automatas 7%
- FIRST SEMESTER_Computer Architecture 10%
- SECOND SEMESTER_Artificial Intelligence 2%
- SECOND SEMESTER_Sistemas de Concurrencia 7%
- SECOND SEMESTER_Programacion orientada a objetos 17%

- Total 100% 42/42

Figure 3: List of Participating Subjects and Percentage of Respondents
Source: Griffith 2015

Question 5 (Figure 4) was designed for students to self-assess their own CLIL experience. Five different aspects were measured. The first was to rate the importance of English in their overall education. The average result showed 7.9 out of ten. The second aspect assessed student comprehension of class lectures, which yielded a score of 8.8 out of ten. The third element considered was student participation in class activities, which produced a score of 6.9 out of ten.

The fourth aspect dealt with student assimilation of course contents. Students' average opinion of their assimilation was 8.3 out of ten. And finally, the last element considered was the overall improvements in English after taking the course. Here it is important to note that we are considering students' perception of learning and not measuring this learning. According to student opinion, improvements in L2 yielded a score of 6.8 out of ten. We see from these five elements a high degree of student satisfaction with the CLIL instruction received.

STUDENT SURVEY CLIL EXPERIENCE

5. RATE THE FOLLOWING ASPECTS OF THIS CLASS IN ENGLISH.
10 being the highest possible value

Rate the importance of this subject in L2 for your overall education.	7.9
Rate your comprehension of the class lectures.	8.8
Rate your participation in class activities.	6.9
Rate your assimilation of course contents.	8.3
Rate improvements in your English level after taking this course.	6.8

Figure 4: Students Rate the Importance of Classes in the Second Language
Source: Griffith 2015

Question 6 (Figure 5) presupposes that CLIL can be challenging. We asked students to rate how challenging they found four variables. The first aspect considered students' foreign language proficiency. Students' average response showed 3.1 out of ten, meaning that they did not consider this aspect very challenging. The second variable measures the professor's foreign language capacity. Again, students did not perceive this as challenging, showing an average score of three out of ten. Next, students were asked to consider the complexity of the course contents as challenging. In this aspect, respondents produced an average score of 5.6 out of ten, making the content complexity the most challenging aspect considered in this question. Finally, students were asked to consider being evaluated in L2 as challenging. Answers showed that this was the least challenging aspect with a score of 2.8 out of ten. It was interesting to note some of the initial concerns revealed by the professors' diagnostic survey included evaluation and foreign language proficiency; however, both proved to be less challenging than anticipated.

Figure 5: Students Rate Challenges of CLIL Instruction
Source: Griffith 2015

Question 7 (Figure 6) assessed one of the specific objectives of the network. We established in our initial action plan to teach as we were going to evaluate. This question is closely tied with this objective and explored the students' point of view about learning and evaluation. Three different aspects relate to final evaluation. When asked if they felt capable of using their second language in the final exam, students answered with a score of 9.1 out of ten. Students produced a score of 8.5 when they considered if the class and practice sessions had promoted learning. Finally, in their assessment of whether the virtual material presented by the instructor helped them for the final exam, students responded with an average score of 8.8 out of ten. Without a doubt, students felt the teachers had prepared them for final evaluation; surprisingly, they were extremely confident about being evaluated through L2.

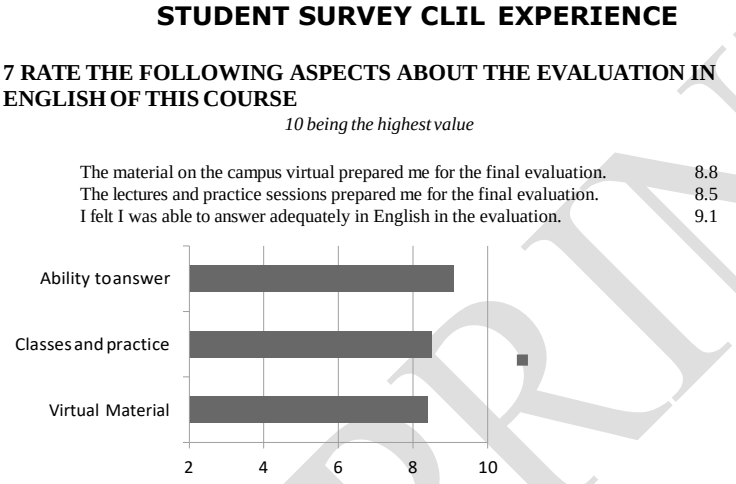


Figure 6: Students Rate Evaluation in CLIL Experience
Source: Griffith 2015

Question 8 (Figure 7) is an open question designed to add student opinion about evaluation. We have created smaller subgroups to facilitate analysis. The majority of respondents, over 50 percent thought assessment in CLIL should be the same as if it were in L1. Other answers about how to be evaluated included six responses for more continual assessment, four responses calling for more focus on form issues, one request for more time on exams, and one requesting more help deciphering exam questions. The remaining responses were either very personal or null. It would seem that the instructors were more concerned about evaluation issues than the students themselves.

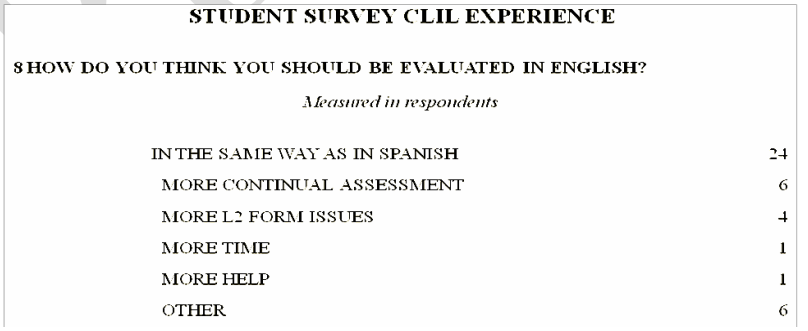
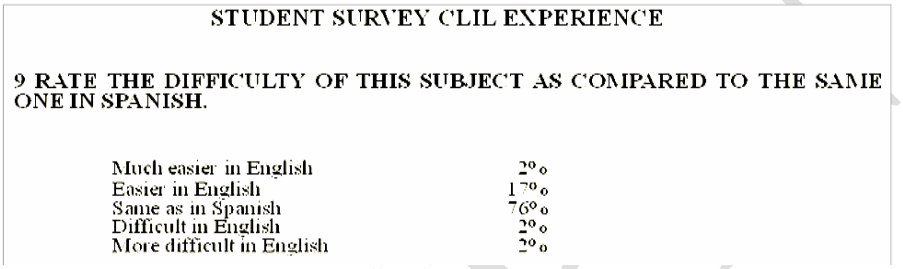


Figure 7: Student Opinion about Being Evaluated in a Second Language
Source: Griffith 2015

Question 9 (Figure 8) assesses the difficulty for evaluation of complex contents in a second language. The opening seminar of the in-service training began with the premise that "CLIL instruction might be harder, but it is better for students." The student responses were quite surprising given our assumptions at the beginning of the project. When comparing to the same subject taught through a first language, 76 percent of respondents found no difference in difficulty between completing this course in either language. Even more surprisingly, 17 percent found the course easier in L2 than in L1. We used the Likert scale to capture the range of intensity for the measurement of difficulty in our five options. Perhaps by not removing the neutral option, we led respondents to this choice, albeit unknowingly. We suggest the students' results are indicators of highly motivated students, who as one student put it, are "the type of person that easily gets distracted if [they] don't feel some pressure, so it was helpful to take it in English…." (anonymous student, professor support network). This same question cross-referenced with other faculties yielded diverging results suggesting that this particular group was unique in their assessment.


STUDENT SURVEY CLIL EXPERIENCE

9 RATE THE DIFFICULTY OF THIS SUBJECT AS COMPARED TO THE SAME ONE IN SPANISH.

Much easier in English	2%
Easier in English	17%
Same as in Spanish	76%
Difficult in English	2%
More difficult in English	2%

Figure 8: Students Compare Difficulty between First and Second Language Classes
Source: Griffith 2015

The survey queried the reasons students have for taking CLIL courses as well as their plans to take more courses offered in English. 98 percent of respondents said they would opt to take classes in L2 if they were available, while 2 percent were undecided. The open format yielded some interesting data about the overall assessment of the course. Thirty-two respondents answered that the experience had been good or excellent. Three found it difficult, particularly at the beginning. Three cited personal reasons, while two offered precise recommendations on how to improve. Finally, two respondents offered negative comments regarding the course contents and the practicality of the subject. Ultimately, this more qualitative date demonstrates the degree of student satisfaction. 95 percent of respondents would recommend taking these courses in L2 to other students (Griffith 2017).

Summarizing the students' reactions to the CLIL experience proved enlightening for teachers. The overall positive responses given by students provided these CLIL professor participants with some very motivating data. Normal student questionnaires take months to process, and teachers rarely get to see what students' more open responses would be. One teacher commented that he highly valued student opinion, noting in his survey: "I would like to know the students' opinion and point of view, though not directly but through the coordinator or a third party in order to obtain non-biased information." The teaching/learning experience is one of reflection and the in-service training provided the participants with the opportunity to reflect.

With the student feedback, we did discover an alternative way to assess teacher performance by moving outside the researcher's point of view. Professors commented that this more qualitative and timely feedback proved enlightening. Undoubtedly, attitudes impact CLIL implementation as much as language proficiency. From within the instructional setting, we have had access to a reality where we could better inform practice by linking professor and student performance. One student's remarks sum up the overall student reaction: **"**Hardly could be better, from professor to learning and of course evaluation. I have no suggestions in mind except for other professors to act in a similar way." These students' comments coupled with the

coordinator's assessment shifted the focus away from language towards learning and encouraged the CLIL instructors to improve their professional practice empowering them to teach their subjects in English. Pointing back to our working hypothesis, effective bilingual instruction encompasses much more than language proficiency. Crucially, professor performance conceptually supports course content; it is this shift in the focus from language to the effective communication of ideas that indicates quality content instruction.

Conclusions

No one project could fully develop the complexity inherent in every classroom; however, we can conclude that a plurilingual university must actively engage content professors who are non-language specialists. Research has shown "that there is an urgent need for a research-driven approach which consults key stake-holders (…) which measures the complex processes involved in EMI and the effects of EMI both on the learning of academic subjects and on the acquisition of English proficiency" (Dearden 2014, 2). In the introduction, we suggested that a practical approach to professional development in education readily connects to the real context and to the challenges instructors face daily. By focusing research directly in university classrooms, we have created a relevant in-service training using action research.

Experts call for the need of such a plan that combines classroom initiatives with professional development: "Without appropriate teacher education programs, the full potential of CLIL is unlikely to be realized, and the approach would be unsustainable" (Coyle 2012, viii). The ultimate goal has been content professors' commitment to plurilingual implementation. Although this project has discussed language as well as teaching, it has most directly dealt with the people responsible for CLIL implementation.

There are a series of assumptions we have pointed out in the development of our project. We do assume that content professors need help in CLIL instruction, we do support the fact that contents can be learned directly through a second language, and we have observed that there is reluctance, or perhaps just fear of the unknown, from subject professors approaching CLIL instruction. "In Spain, these [CLIL] programs are a challenge, as the level of communicative competence in foreign languages is poor (…), and teachers may be reluctant to implement them for various reasons" (Rubio and Hermosin 2010, 107). Based on these assumptions, this study used participatory action research to validate the efficiency of the bilingual program by using professors' direct experience teaching Computer Science through a second language together with student data.

By assessing professors with student data, the purpose of this research project has been to inform practice and to link student and professor performance. It is this student data that has been the focus of this article, and three research questions have provided the research framework. First, will Spanish students interact in a CLIL class? We have observed quite a range of interactions and note that content-based subjects inserted into the real university context offer both lecture and interactive sessions, as corroborated by Squires (2005). At the same time, we question whether interaction measures language learning more effectively than it measures content learning. The second aspect explored how CLIL student results compare with control groups taught in the first language. In all cases, findings yielded similar or superior results in final evaluation when comparing eighteen subjects taught in both the first and second language. And finally, a qualitative survey measured how students rate themselves and their instructors after the CLIL experience. For professors teaching in English for the first time, this more qualitative feedback produced an insightful assessment and provided these professors with a unique opportunity for reflection in action.

The in-service training and facilitative collaboration gave these professors the support they needed when approaching bilingual instruction. Through action research, the professors built on their expertise and enriched the shared group experience as a whole. Linking student and professor performance to assess professors through student data has validated the quality

assessment of CLIL instruction. We note that this data forms part of an ongoing project at several Spanish universities where content professors' needs are being taken up by language specialists. We feel these collaborations can enlighten linguists and challenge us all to discover effective measures for bilingual implementation that move beyond language performance.

At the theoretical level, bilingual models seemingly are a panacea to our deepest desire to manage multiple languages. Deductively, we seem to seek utopian multilingualism where everyone communicates freely with no one language dominating the rest. However, at the practical level, CLIL models are challenging, complex, and ever-changing. Linguistic planners run the risk of falling victim to the false promises of plurilingual intentions, so we can only humbly conclude that, without exception, all language learning methods make promises they cannot keep (Decco 2001). The reality of non-native teachers teaching non-native students is that imperfection is part of the model and the model is continuously being constructed and informed by practice.

Action research is a disciplined process of inquiry conducted by and for those taking action. Indeed, action research is informed practice, and the success of plurilingual implementation depends on "the commitment to action by university professors within the specific context of their departments, their contents, and their students" (Griffith 2017, 142). The ongoing challenge for bilingual teaching is how to provide support for instructors who are not specialists in the second language.

Dearden (2014) has reported that in many countries, the educational infrastructure does not specifically support quality bilingual instruction. For example, there are few organizational or pedagogical guidelines that might lead to effective bilingual teaching and learning. But most importantly, there is little or no specific bilingual instructional content in teacher education programmes or in professional development (in-service) courses. Language learning is taken for granted by content specialists, and linguists seem to focus solely on language learning. This project suggests that content professors and linguists should collaborate.

Can interaction and learner outcomes combine with language performance to assess bilingual instruction? Is efficient content teaching equated with effective language teaching? And can students help reveal assessment features for the bilingual classroom? These are the questions that drive our ongoing research in different educational settings. In fact, bilingual instruction "is not a fixed concept but one that is evolving as an increasing number of countries adopt it as a system of education" (Dearden 2014, 7). Our project's aim has been to support content teachers in order to ensure quality bilingual instruction. Quality should be measured by progressive improvements in materials, in student interaction, as well as learner outcomes. Defining quality and functionality must be constructed from the bottom-up. There are indeed, many "right" ways to approach bilingual instruction in practice, whereby practitioners discover their solutions.

We suggest that more orientation and practical support is needed for the ideas of multilingual convergence to become a reality, and to achieve this, greater collaboration is warranted between content and language specialists. As a language specialist placed into a context of content specialists, I must remark on how much I continue to learn about content instruction through this research project. Collaboration is sometimes accepting your role as an "outsider" and combining your strengths with the strengths of others.

Acknowledgement

We are grateful to the Computer Science Department (ETSI Informática) at the University of Málaga who provided funding together with the invaluable access to real CLIL classrooms in their own desire to give support to their teaching staff through this network. Without a doubt our sincerest appreciation goes to the professor participants in this project.

REFERENCES

Barrios, Elvira, Aurora López-Gutiérrez, and Clotilde Lechuga. 2016. "Facing Challenges in English Medium Instruction through Engaging in an Innovation Project." *Revista Procedia—Social and Behavioral Sciences* 228: 209–14. https://doi.org/10.1016/j.sbspro.2016.07.031.

Barron, Colin. 2002. "Problem-Solving and EAP: Themes and Issues in Collaborative Teaching Venture." *English for Specific Purposes* 22: 297–314. https://eric.ed.gov/?id=EJ668074.

Brinton, Donna, Marguerite Snow, and Marjorie Wesche. 2011. *Content-based Second Language Instruction.* Ann Arbor, MI: University of Michigan Press.

Castro Garcés, Angela Yicely, and Liliana Martínez Granada. 2016. "The Role of Collaborative Action Research in Teachers' Professional Development." *PROFILE Issues in Teachers' Professional Development* 18 (1): 39–54. https://doi.org/10.15446/profile.v18n1.49148.

Coyle, Do. 2012. "Forward." In *CLIL in Spain: Implementation, Results and Teacher Training*, edited by David Lasagabaster and Yolanda Ruiz de Zarobe, i–viii. Newcastle upon Tyne: Cambridge Scholars Publishing.

Coyle, Do, Phillip Hood, and David Marsh. 2010. *CLIL Content and Language Integrated Learning.* Cambridge: Cambridge University Press.

Dalton Puffer, Christine. 2007. *Empirical Perspectives on CLIL Classroom Discourse.* Frankfort: Peter Lang.

Dearden, Julie. 2014. "English as a Medium of Instruction—A Growing Global Phenomenon." *Executive Summary.* British Council. http://englishagenda.britishcouncil.org/continuing-professional-development/cpd-managers/english-medium-instruction-growing-global-phenomenon.

Decco, Wilhelm. 2001. "On the Mortality of Language Learning Methods." Given as the James L. Barker lecture, Brigham Young University, Provo, UT, November 8, 2001.

Denzin, Norman. 2006. *Sociological Methods: A Sourcebook*. Piscataway, NJ: Aldine Transaction.

Fortanet-Gomez, Inmaculada. 2012. "Academics' Beliefs about Language Use and Proficiency in Spanish Multilingual Higher Education." *AILA Review* 25 (1): 48–63. https://doi.org/10.1075/aila.25.04for.

Gajo, Laurent. 2007. "Linguistic Knowledge and Subject Knowledge: How does Bilingualism Contribute to Subject Development?" *The International Journal of Bilingual Education and Bilingualism* 10 (5): 563–81. https://doi.org/10.2167/beb460.0.

Griffith, Mary. 2015. "Integrating Content and Language in Higher Education: Professor Support Network." PhD diss., University of Málaga.

———. 2017. "Tapping into the Intellectual Capital at the University." *Universal Journal of Educational Research* 5 (12A): 134–43. https://doi.org/10.13189/ujer.2017.051320.

Griffith, Mary, and Clotilde Lechuga. 2018. "Contrasting Learner Language in Social Sciences Education: A Case Study at the University of Málaga." *Universal Journal of Educational Research*, 6 (10): 2364–72. https://doi.org/10.13189/ujer.2018.061035.

Llinares, Ana, Emma Dafouz, and Rachel Whittaker. 2007. "A Linguistic Analysis of Compositions Written by Spanish Learners of Social Sciences in CLIL Contexts." In *Diverse Contexts Converging Goals. Content and Language Integrated Learning in Europe*, edited by David Wolff and David Marsh, 227–37. Frankfurt: Peter Lang.

Long, Michael. 1991. "Focus on Form: A Design Feature in Language Teaching Methodology." In *Foreign Language Research in Cross-Cultural Perspective*, edited by Kees de Bot, Ralph Ginsberg, and Claire Kramsch, 39–52. Amsterdam: John Benjamins.

Marsh, David, Victor Pavón-Vázquez, and María Jesus Frigols-Martín. 2013. "The Higher Education Languages Landscape: Ensuring Quality in English Language Degree Programmes." *Higher Learning Research Communications* 5 (1): 4–10. https://doi.org/10.18870/hlrc.v5i1.241.

Neumann, Ruth, Sharon Parry, and Tony Becher. 2002. "Teaching and Learning in their Disciplinary Contexts: A Conceptual Analysis." *Studies in Higher Education* 27 (4): 405–17. http://dx.doi.org/10.1080/0307507022000011525.

O'Brien, Rory. 2001 "An Overview of the Methodological Approach of Action Research." In *Theory and Practice of Action Research*, edited by Richard Richardson, 1–18. Joao Pessoa: Universidade Federal da Paraíba.

Reason, Peter, and Hilary Bradbury. 2001. *The Handbook of Action Research.* London: Sage Publications, Ltd.

Rubio, Fernando, and Manuel Hermosin. 2010. "Implantación de un Programa de Plurilingüismo en el Espacio Europeo de Educación Superior: Análisis de contexto y detección de necesidades" [Implementation of a Plurilingual Plan in the European Higher Education Area]. *Revista de Educación* [Journal of Education] 12: 108–28. http://hdl.handle.net/10272/5325.

Sagor, Richard. 2000. "What is Action Research?" In *Guiding School Improvement with Action Research*, 3–11. Alexandria, VA: ASCD (Association for Supervision and Curriculum Development). http://www.ascd.org/publications/books/100047/chapters/What-Is-Action-Research%C2%A2.aspx.

Schön, Donald. 1987. *Educating the Reflective Practitioner.* San Francisco, CA: Jossey-Bass Publishers.

Smit, Ute, and Emma Dafouz. 2012. "Integrating Content and Language in Higher Education: An Introduction to English-Medium Policies, Conceptual Issues and Research Practices across Europe." *AILA Review* 25 (1): 1–12. https://doi.org/10.1075/aila.25.01smi.

Sorbonne Joint Declaration. 1998. "Joint Declaration on Harmonisation of the Architecture of the European Higher Education System." http://www.ehea.info/media.ehea.info/file/1998