Challenges in the Implementation of CLIL in Higher Education: From ESP to CLIL in the Tourism Classroom

Retos en la implementación de CLIL en la educación superior: de ESP a CLIL en el aula de turismo

Desafios na implementação do método AICL no ensino superior: do ESP à AICL na sala de aula de turismo

Melita VEGA
Universidad del Azuay, Ecuador.
mvvega@uazuay.edu.ec
https://orcid.org/0000-0002-6501-4011

María de Lourdes MOSCOSO
Universidad del Azuay, Ecuador.
lmoscoso@uazuay.edu.ec
https://orcid.org/0000-0002-5992-9530

Received: 24/04/2019
Accepted by peers: 02/10/2019
Sent to peer review: 25/04/2019
Approved: 03/10/2019

DOI: 10.5294/laclil.2019.12.1.7

To reference this article (APA) / Para citar este artículo (APA) / Para citar este artigo (APA)
ABSTRACT. Content and Language Integrated Learning (CLIL) has often been touted as an effective means of enhancing the language proficiency gains among its learners due to its focus on content over form and higher cognitive demand. However, cautions have been raised regarding the varying conditions and contexts that need to be taken into consideration in order to ensure its effectiveness. This study aimed to analyze the outcome of switching from an English for Specific Purposes (ESP) program to a CLIL program in the fourth and fifth semesters of the School of Tourism at the University of Azuay. Study participants were randomly divided into two groups: a CLIL (experimental) and a non-CLIL group (control), where the former received CLIL instruction and the latter received ESP instruction for an average of five hours per week over a period of two consecutive semesters. The findings revealed no significant increases in language proficiency or differences in achievement between the two groups, thus suggesting that the starting language level of learners influenced the results of the CLIL program.

Keywords (Source: Unesco Thesaurus): Content and Language Integrated Learning; CLIL; English for Specific Purposes; ESP; Content Based Instruction; CBI; intensity; language proficiency.

RESUMEN. El Aprendizaje Integrado de Contenidos y Lengua Extranjera (AICLE) ha sido utilizado con frecuencia como un método adecuado para mejorar la suficiencia en idiomas debido a su enfoque del contenido sobre la forma y una mayor demanda cognitiva. Sin embargo, han surgido precauciones con respecto a las condiciones y contextos que deben ser tomados en cuenta para su eficacia, como los niveles del idioma. El presente estudio tiene como objetivo analizar el resultado de realizar un cambio desde un programa de inglés para propósitos específicos (ESP, por sus siglas en inglés), hacia AICLE en el cuarto y quinto semestre de la Escuela de Turismo de la Universidad del Azuay. Los participantes fueron agrupados al azar en dos grupos: AICLE (experimental) y No AICLE (control); el primero recibió instrucción AICLE y el segundo recibió instrucción ESP con un promedio de cinco horas semanales durante dos semestres consecutivos. Los resultados revelaron que no hubo un incremento significativo en la suficiencia del idioma, ni diferencias en aprendizaje entre los dos grupos. Esto lleva a inferir que el punto de partida en cuanto al nivel de idioma influyó en la eficacia del programa AICLE.

Palabras clave (Fuente: tesauro de la Unesco): AICLE; Aprendizaje Integrado de Contenidos y Lengua Extranjera; ESP; Inglés para propósitos Específicos; Instrucción Basada en Contenidos; CBI; intensidad; suficiencia en idioma.

RESUMO. A aprendizagem integrada de conteúdos e de língua (AICL) tem sido frequentemente usada como um método apropriado para melhorar a proficiência na língua devido ao seu foco no conteúdo em vez da forma e à maior exigência cognitiva. No entanto, foram tomadas precauções em relação às condições e contextos que devem ser levados em consideração para sua eficácia, como os níveis de linguagem. Este estudo tem como objetivo analisar o resultado de uma mudança de um programa de inglês para fins específicos (ESP) para AICL no quarto e quinto semestres da Escola de Turismo da Universidade de Azuay. Os participantes foram aleatoriamente agrupados em dois grupos: AICL (experimental) e Não AICL (controle); o primeiro recebeu instrução AICL e o segundo recebeu instrução ESP com uma média de cinco horas por semana por dois semestres consecutivos. Os resultados revelaram que não houve aumento significativo na proficiência na língua, nem diferenças na aprendizagem entre os dois grupos. Isso nos leva a inferir que o ponto de partida em termos de nível de idioma influenciou a eficácia do programa AICL.

Palavras-chave (Fonte: tesauro da Unesco): AICL; Aprendizagem Integrada de Conteúdo e Língua Estrangeira; ESP; inglês para fins específicos; instrução baseada em conteúdo; CBI; intensidade; proficiência na língua.
Introduction

It has been suggested that a second or foreign language is best learned in an environment that places meaning at the forefront rather than accuracy. Such an environment makes the task of achieving the target language level challenging, but achievable through the creation of situations that foster language practice and usage in a meaningful way (Burger, Weinberg, & Wesche, 2013; Butler, 2005; Dupuy, 2000; Krashen, 1982; Snow, 2005). These perspectives, gathered from research in the fields of cognitive psychology, language learning theories, second language acquisition, and psycholinguistics, have served as supportive pillars for the emergence of content and language integrated learning programs that create opportunities for a natural acquisition of a target language through meaningful communication.

Content and Language Integrated Learning (CLIL) was first coined in the mid-1990s in Europe as a result of many changes occurring in the European Union at the time in terms of integration, expansion, and modernization. By applying different strategies to teach language and subject matter simultaneously, CLIL emerged as a game-changer that challenged the prevailing “traditional model of the English language classroom” (Marsh & Frigols, 2012), which had been primarily language-focused. According to Chamot (2014), a CLIL environment is one where non-English speaking students learn a target language as a lingua franca (often referred to English as a Foreign Language or EFL) and use it in an academic setting within a “globalized international context” (p. 79). In other words, the language used as an instrument is one the students only hear and use in the classroom. The main distinguishing feature of CLIL, many authors seem to agree, is its dual focus on both language and content teaching (Banegas, 2012; Eurydice, 2006; Stoller, 2002), which is said to help foster language acquisition more efficiently than teaching the two elements separately (Wolff, 2009 p. 560). Cummins (2013) has also asserted that CLIL programs can be a highly effective approach for learning content and language, since students can achieve high levels of fluency in the target language without sacrificing their knowledge of the curriculum content.
The notion of using meaningful content to learn a language has long appeared in second language acquisition research, particularly in the work of Krashen (1982), who has noted that the goal of learning a second language is better achieved in contexts that generate “comprehensible input” with meaningful information than through memorization of grammatical rules (Krashen, 1982, p. 7). This perspective served as the premise for his input hypothesis theory \((i + 1)\), which contends that language is acquired in different stages that challenge learners’ proficiency levels. Within this context, \(i\) reflects the student’s current proficiency level and \(1\) reflects the next level of proficiency that is slightly beyond the student’s current grasp of the language. For students to be able to advance to the next skill level, they need “comprehensible input” that equals \(i + 1\). If this input is provided, then learning is said to occur as students focus on understanding the content (“natural, communicative and roughly tuned”) not the form (i.e., grammatical rules) by drawing upon other skills, such as sociolinguistic competence and context. According to this hypothesis, “finely-tuned input” in the form of specific grammatical structures will not be the most effective to yield authentic communication (Krashen, 1982, p. 26).

Further foundational pillars for content and language integrated programs such as CLIL can be found in literature regarding contextually appropriate language (Lyster, 2007) and constructivist and cognitive theories. Based on the distinctions presented by Cummins (2008) between Basic Interpersonal Communicative Skills (BICS) and Cognitive Academic Language Proficiency (CALP), it has been suggested that students need to learn content while developing CALP, which involves language manipulation in cognitively demanding situations reduced to a determined context. Since CLIL environments are known for helping foster negotiation of meaning (Lightbown & Spada, 2013), learners will use their higher order thinking skills to analyze, synthesize, discuss, evaluate, and interpret content with and through a second or foreign language. These learning processes bring a multifaceted quality to the roles and responsibilities of learners, who, as Coyle (2009) explains, go beyond the goal of merely acquiring knowledge and skills. Rather, they learn to construct their own knowledge by developing relevant skills that involve comprehension, learning, and critical thinking.
In determining the required elements for learning a foreign language, Coyle suggested the conceptual framework of the 4Cs, which features Content (theme), Cognition (learning and reasoning), Communication (language), and Culture (social consciousness of oneself and others). Within this framework, learners are capable of managing content-related information and use that information by employing higher cognitive processes (Morgado & Coelho, 2013). In Coyle’s (2009) view, a high caliber CLIL requires advancement in knowledge, skills, and comprehension of content, as well as a deep intercultural understanding that encompasses all of the Cs. The objectives sought through CLIL programs focus on intercultural, cognitive, content and language aspects, each of which is given equal weight (see Figure 1).

Figure 1. 4Cs framework for CLIL

Source: Adapted from Coyle (2009).

Literature Review

The roots of content and language integrated programs have been traced to immersion programs in Canada first implemented in the 1960s (Banegas, 2012; Burger, Weinberg, & Wesche, 2013; Dueñas, 2003; Grabe & Stoller, 1997) with the goal of promoting bilingualism and bi-literacy using French as a medium of instruction for most of the day, thus enabling speakers of English to achieve fluency (Cummins, 2013). It is worth noting the differences between immersion and CLIL, particularly in that the former uses a second language (in the case of the immersion programs in Canada, French, which is the country’s second official language), while the latter uses a foreign language. Other char-
acteristics of immersion programs include the use of a language that is not spoken by the local society, similar language proficiency rates among students (usually low), and the goal of reaching bilingualism (Johnson & Swain, 1997).

Content Based Instruction (CBI) is a term largely used in North America to refer to programs that aim to help improve English proficiency among students whose mother tongue differs from the language that is widely in use in their country of residence (Burger, Weinberg, & Wesche, 2013). Its emergence in the United States was sparked by the arrival of non-English speaking students in the 1970s and came in the form of programs that merged English language instruction and content designed to serve as a pathway to success in other mainstream subjects later on (Crandall & Tucker, 1990). At its core, it is an approach where meaningful subject matter is taught using the target language primarily as a tool, rather than as an object of study. Thus, in such a classroom environment, the amount of specific language instruction can vary considerably, even becoming non-existent.

English for Specific Purposes (ESP) also appears frequently in research related to language teaching through meaningful and useful content (Vega, 2017). Its roots can be traced to the rapid developments made in technology, communications, and economics after the end of the Second World War that later gave way to the need for technical and scientific writing, in which precision was essential. Thus, ESP served to support students in achieving specific communications goals required by their professions or vocations such as nursing, engineering, and medicine, among others (Gonzalez, 2015). The components of an ESP program include the study of a target language within the context of a particular profession, analysis of learner needs, and classroom use of content and materials that might be found in a real-life work environment (Dudley-Evans, 1997). In this sense, ESP differs from CBI and CLIL in that it does not involve teaching a specific subject matter; rather, its focus lies in completing tasks appropriate for specific work situations. For example, in an ESP class focused on hotel management, tasks could include practicing how to provide instructions (through specific vocabulary) regarding cleaning duties.

While immersion, CBI, CLIL and ESP share the common trait of using various types of authentic classroom materials, each approach
differs in how language and content influence syllabus design and class. While specific tasks tend to drive the syllabus design of ESP programs, selected content and aligned language objectives serve as the starting point for CLIL course design activities, as seen in Table 1 (Vega, 2017). While ESP can be deemed as a category of English Language Teaching (ELT), CLIL is said to differ greatly in materials, teacher preparation, and course objectives, most notably in that CLIL places greater importance on content than ESP (Yang, 2016).

### Table 1. Comparison of ESP, CLIL and CBI/Immersion

<table>
<thead>
<tr>
<th>Feature</th>
<th>English for Specific Purposes (ESP)</th>
<th>Content and Language Integrated Learning (CLIL)</th>
<th>Content-Based Instruction (CBI)/Immersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting point for syllabus design</td>
<td>Learner vocational or professional communications needs</td>
<td>Content and Language</td>
<td>Content</td>
</tr>
<tr>
<td>Role of language</td>
<td>Used to achieve a specific objective</td>
<td>Used as a medium of instruction to learn content and language</td>
<td>Used as a medium of instruction to learn content</td>
</tr>
<tr>
<td>Materials</td>
<td>Authentic</td>
<td>Authentic</td>
<td>Authentic</td>
</tr>
<tr>
<td>Focus of in-class activities</td>
<td>Language and vocabulary learning</td>
<td>Content and language learning</td>
<td>Content learning</td>
</tr>
</tbody>
</table>

Source: Compiled by authors, adapted from Vega (2017, p. 30).

Despite the dual focus on both language and content that is inherent in CLIL, in practice, the needs of learners, teacher characteristics and institutional expectations can all affect how much weight is given to these two elements in the classroom (Yang, 2016). Therefore, a CLIL program could be designed to range between two ends of a spectrum where one end is more content-centric (“strong” CLIL) and the other is more language-centric (“weak” CLIL) (Paran, 2013; Ravelo, 2014), or somewhere in between (see Table 2). The ability to move along this spectrum is said to afford a great deal of flexibility in the classroom; however, this flexibility can also be seen as both a “strength and a potential weakness” (Coyle, 2007, p. 546).
CLIL has been described as a “limited novelty” (Bruton, 2013, p. 589) that bears a significant resemblance to the Communicative Language Teaching (CLT) approaches that surfaced in the 1970s and 1980s and emphasized learning grammatical rules and practical social norms so students could achieve “meaningful communication” in real-life scenarios, such as learning how to order at a restaurant (Richards, 2006, p. 3); however, this comparison has been tempered by the argument that authenticity has a greater presence in the CLIL classroom than in the CLT classroom (Coyle, Hood, & Marsh, 2010).

Table 2. Met's Continuum of Content and Language Integration

<table>
<thead>
<tr>
<th>Content-based language teaching:</th>
<th>Language-driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>A continuum of content and language integration</td>
<td></td>
</tr>
<tr>
<td>Content-driven</td>
<td>Language-driven</td>
</tr>
<tr>
<td>Content is taught in L2.</td>
<td>Content is used to learn L2.</td>
</tr>
<tr>
<td>Content learning is priority.</td>
<td>Language learning is priority.</td>
</tr>
<tr>
<td>Language learning is secondary.</td>
<td>Content learning is incidental.</td>
</tr>
<tr>
<td>Content objectives determined by course goals or curriculum.</td>
<td>Language objectives determined by L2 course goals or curriculum.</td>
</tr>
<tr>
<td>Teachers must select language objectives.</td>
<td>Students evaluated on content to be integrated.</td>
</tr>
<tr>
<td>Students evaluated on content mastery.</td>
<td>Students evaluated on language skills/proficiency.</td>
</tr>
</tbody>
</table>


While there is much research focusing on the use of CLIL in primary and secondary education, its rise in university education merits further attention, as this is where content can be studied with a much wider lens and a more practical purpose in mind. According to Tudor (2008), CLIL is a productive learning method in terms of validity and brings a significant motivational value to the table. In a university context, students of non-linguistic academic programs may not have an interest in learning a second or foreign language; however, the ability to do so with their academic area of interest serving as a context for learning may prove to be a motivating factor (Vilkanciene, 2011, p. 112). Thus, it has been argued that the use of real material leads to more authentic reactions among learners (Arment & Perez-Vidal, 2015) and that using a target language for a purpose that goes beyond the classroom can have a positive impact on learner motivation (Hunt,
This authenticity delivered through content and language can offer students a flexibility through which their most highly favored elements can help improve their proficiency in their least favorite areas (Banegas, 2012).

In a review of empirical studies on CLIL and other content-based approaches featuring experimental and control groups, Vega (2017) found reports of varying degrees of improvement among CLIL learners in areas such as writing, reading, and speaking. The review also noted that while gains in these areas support the purported benefits of and success of content and language integration programs, findings on greater linguistic competence have been mixed. Other authors' findings have been equally varied; some have contended that improvements in grammatical accuracy in the CLIL classroom are lower than those achieved in speaking and listening (Pica, 2010; Maíz-Arévalo & Domínguez-Romero, 2013), while others have found that a CLIL environment offered a better learning ground for linguistic gains than a mainstream language class (Arment & Perez-Vidal, 2015; Lasagabaster & Doiz, 2016). A large scale study of over 1,000 primary and secondary school students in Andalusia by Lorenzo, Casal, and Moore (2010) found CLIL learners outperformed their Non-CLIL peers in overall language proficiency (reading, writing, speaking, and listening) while another longitudinal study of 403 students at a secondary school over a four-year period by Arribas (2016) found no statistically significant advantage in vocabulary gains in students exposed to CLIL via additional hours of instruction compared to their Non-CLIL counterparts. While the CLIL group reported listening and speaking as the most improved skills, these were the result of self-evaluations. The lack of statistically significant scores was attributed to a poorly implemented execution of the CLIL program, in which the number of hours of exposure to CLIL varied considerably from year to year.

A review by Graham, Choi, Davoodi, Razmeh, and Dixon (2018) of twenty-five studies examining the outcomes of CLIL versus Non-CLIL environments in secondary and higher education found nearly equal numbers of cases where CLIL courses produced advantageous gains over Non-CLIL courses and vice versa, thus highlighting the potential effect of methodological flaws, such as uneven evaluation criteria and various mechanisms for implementation of CLIL. Indeed, the mixed
findings on CLIL have prompted suggestions that its success is subject to a variety of factors such as learning context, current language proficiency, cognitive skill level, age of learners, as well as prominence and potential use of the target language outside the classroom (Banegas, 2014; Nieto, 2016; Dalton-Puffer, 2011; Paran, 2013; Sylvén, 2013). Regarding proficiency, many authors have noted that the language levels of learners must be up to par to be able to process the content studied in a CLIL environment to mitigate the risk of potential failure or even withdrawals should their deductive capabilities fall short (Apsel, 2012; Bruton, 2013; Lorenzo, 2007). However, there does not appear to be a consensus as to what exact language level is required or what age level is appropriate to introduce CLIL courses (Sylvén, 2013). According to Paran (2013), “selective implementation” of students also plays a significant role in the success of CLIL programs (p. 325), particularly since many students are either pre-selected to enroll in a CLIL program based on their English level and/or content knowledge, or select themselves as participants and thus, may be more motivated to succeed because they are already “high achievers” (p. 326). Other authors who not only acknowledge the impact of student selection in CLIL program outcomes (Broca, 2016; Bruton, 2013), but also recommend caution when examining prior research comparing proficiency levels among students grouped into CLIL and non-CLIL groups (Merino & Lasagabaster, 2018) also share this viewpoint. The effect of time and intensity of instruction on proficiency gains among CLIL students has also been studied (Merino & Lasagabaster, 2018) with some authors going so far as to suggest that positive CLIL outcomes may solely be the result of additional exposure (Graham et al., 2018).

Method

Considering the critical success factors in the implementation of a CLIL program and the mixed findings as to its results, the present study aimed to analyze the outcome of a CLIL program within the School of Tourism at the University of Azuay. To this purpose, the following research questions were posed at the onset of the study:
1. To what extent can CLIL improve English language proficiency among second-year Tourism students at the University of Azuay?

2. How do second-year Tourism students perceive CLIL as an approach to English language learning?

**Site and context**

The study was conducted in the School of Tourism, a four-year undergraduate program at the University of Azuay in Ecuador. The intervention and data collection took place over the course of two consecutive academic semesters: March – July 2018 and September 2018 – February 2019. It is worth noting that, up until the start of the study, English had previously formed part of the core Tourism curriculum through an English for Specific Purposes (ESP) approach. As a side note, ESP was removed from the School of Tourism at the time of the study; according to university regulations instituted thereafter, newly enrolled students are required to demonstrate proficiency at a B2 level through an exam, external certifications from recognized language entities, or by taking English as a Foreign language courses at the Language Unit of the University. In the March – July 2018 semester, students were allotted a total of six hours of class for English and five in the September 2018 – February 2019 semester, as per the curriculum set out by the Academic Council.

**Participants**

The participants in the study were n=17 students enrolled in an ESP course in the fourth semester of the Tourism program in the March 2018 – July 2018 academic period. It is worth noting that the Tourism program does not open every semester; therefore, the pool of students available to participate at the time of the study and readily accessible to the researchers were those enrolled in second and fourth semester. Since the objective was to analyze the outcomes of a CLIL approach in the Tourism program, it was decided that the fourth-semester students would be suitable candidates as they had been attending university and taking English courses for a year and a half. For the second semester, the sample size was reduced to 15 students due to student withdrawals from the university. Their ages ranged from 19 – 21 and they...
were randomly split into two groups: a CLIL group who would receive classes through the CLIL approach with content related to tourism and a Non-CLIL group who would continue receiving the ESP classes with the textbook assigned by the university and serve as a control group. The native language of all students was Spanish and their English proficiency levels varied. One of the researchers was the regular ESP professor in the Tourism program while the other was a professor in the School of International Studies invited to collaborate on the project. The switch to CLIL as part of the intervention constituted a new learning approach for the study participants.

**Study design and procedure**

Since the aim of the study was to examine the outcome of implementing an CLIL program in the Tourism program at the University and learn more about students’ perceptions of the approach, a mixed-methods design was adopted. According to Creswell (2014), the use of qualitative and quantitative methods can often offer a deeper perspective on an issue than by using either of the methods in isolation. Thus, both qualitative and quantitative instruments including proficiency tests and questionnaires were used for data collection.

As for course materials, the ESP courses used a Tourism ESP textbook where each unit presented a tourism-related topic covered through different sections, such as grammar, vocabulary, reading, listening, speaking and writing to provide students with the opportunity to practice the grammar and the vocabulary presented. Class activities included writing short texts, creation of short dialogues to practice specific skills (e.g., making a complaint, asking for clarifications at an airport), and grammar exercises, such as fill in the blanks, circle the correct option, and unscrambling words. Unit tests contained in the text measured comprehension of grammar, vocabulary, and the four skills.

The CLIL courses focused on tourism-related content inspired by the topics mentioned in the ESP textbook (e.g., basic concepts of customer service, the airline industry, destination marketing, etc.). The teachers further expanded the topics into self-contained content units using additional sources (such as business administration textbooks) and prepared appropriate in class materials for each unit that included
theoretical presentations via PowerPoint and video, a case study for analysis in groups, a self-assessment content quiz, and a speaking/writing assignment (see Appendix A for sample classroom materials). The final module featured an oral presentation assignment carried out in pairs. In total, eight modules of content were divided into 10 units delivered over the course of the two semesters. The content was as follows:

» March – July 18, 2018: client service, human resources, leadership, and hotel administration.

» September 2018 – February 2019: public relations, the communications process, cultural differences, and tourism marketing (2 units).

While the CLIL modules featured select grammatical structures (the same as those featured in the ESP textbook) embedded in the case studies read by the students and listed as requirements in the written/speaking assignments (e.g. modals, phrasal verbs, present perfect tenses), direct grammar instruction was not made a core focus of the class lessons and no grammar/vocabulary quizzes were included in the CLIL curriculum. Thus, based on Met’s (1999) continuum of content and language integrated learning, the approach in the design of the CLIL curriculum was primarily content-driven.

The researchers, who had prior experience with the CLIL format, took turns teaching both the CLIL and non-CLIL groups each week, thus ensuring that each group had an equal amount of class time with each so as to avoid potential bias. Thus, while one researcher worked with the CLIL group using the content material, at the same time, the other worked with the Non-CLIL group using the regular ESP textbook. Classroom activities in the Non-CLIL group included grammar and vocabulary quizzes, writing assignments and oral presentations. The CLIL group completed written assignments and multiple-choice questionnaires based on case studies that emphasized reading skills, group critical thinking exercises, and oral presentations. While direct grammatical instruction was provided to the Non-CLIL group as dictated by the ESP textbook, it was provided to a lesser degree to the CLIL group in the form of one on one and group guidance during the completion of in-class written assignments and advance review of group presentations. Content mastery was not the primary objective of this study; however, content-related questions were included in all official midterm and final examinations during both semesters for the CLIL group.
Data collection instruments

The intervention started with the administration of the Cambridge English Placement Test (CEPT), which is an adaptive online test that assesses grammar, oral and reading comprehension skills and provides precise scores based on the Common European Framework of Reference for Languages (CEFR). Since the CEPT serves as the official placement mechanism for all students wishing to take English courses in the Language Unit at the university, it was deemed a suitable and reliable pre-test to ascertain the level of English of all study participants at the onset of the intervention. At the end of the second semester, the CEPT was administered again as a post-test to obtain final scores that could be compared in the data analysis phase. It is worth noting that the CEPT evaluates general English proficiency only, so it does not feature specific content tailored to any particular subject area. It also does not evaluate speaking or writing skills.

At the end of both academic semesters, a questionnaire was administered to the CLIL to gather information on their personal experiences and perspectives throughout the intervention. The questions ranged covered topics such as students’ prior history with CLIL classes (if any), perceived difficulties associated with this approach, and a self-assessment as to which skills they felt had improved the most as a result (see Appendix B). To complement the information gathered in the questionnaires, a group of 5 students were selected at random from the CLIL group for an informal focus group. To encourage more openness with answers and ease of expression, the questionnaire and focus group were administered in Spanish.

Results

Quantitative analysis

To determine the English proficiency levels of all participants in the control and experimental groups at the start and end of the study, two
online CEPTs were administered in the computer laboratory at the university: a pre-test at the beginning of the March – July 2019 semester and a post-test at the end of the September 2018 – February 2019. As shown in Table 3, four students in the CLIL group maintained their placement score according to the CEFR, while three increased and one decreased. In the Non-CLIL group, four maintained their placement score, while two decreased and one increased.

### Table 3. CEPT exam results (pre and post tests)

<table>
<thead>
<tr>
<th>CEPT Score</th>
<th>CEFR Placement</th>
<th>CEPT Score</th>
<th>CEFR Placement</th>
<th>CEPT Score</th>
<th>CEFR Placement</th>
<th>CEPT Score</th>
<th>CEFR Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>A1</td>
<td>10</td>
<td>A1</td>
<td>31</td>
<td>B1</td>
<td>20</td>
<td>A2</td>
</tr>
<tr>
<td>32</td>
<td>B1</td>
<td>35</td>
<td>B1</td>
<td>43</td>
<td>B2</td>
<td>50</td>
<td>B2</td>
</tr>
<tr>
<td>19</td>
<td>A1</td>
<td>25</td>
<td>A2</td>
<td>25</td>
<td>A2</td>
<td>27</td>
<td>A2</td>
</tr>
<tr>
<td>37</td>
<td>B1</td>
<td>41</td>
<td>B2</td>
<td>30</td>
<td>B1</td>
<td>32</td>
<td>B1</td>
</tr>
<tr>
<td>23</td>
<td>A2</td>
<td>21</td>
<td>A2</td>
<td>14</td>
<td>A1</td>
<td>19</td>
<td>A2</td>
</tr>
<tr>
<td>27</td>
<td>A2</td>
<td>34</td>
<td>B1</td>
<td>15</td>
<td>A1</td>
<td>8</td>
<td>Below A1</td>
</tr>
<tr>
<td>15</td>
<td>A1</td>
<td>11</td>
<td>Below A1</td>
<td>15</td>
<td>A1</td>
<td>17</td>
<td>A1</td>
</tr>
<tr>
<td>28</td>
<td>A2</td>
<td>24</td>
<td>A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Compiled by authors

The average numeric score achieved by the participants in the CLIL group in the pre-test was 24.88/50 and 25.13 on the post-test, thus showing a slight increase of 0.25. The average score on the pre-test and post-test for the non-CLIL group was 24.71/50, thus indicating neither an increase nor a decrease in achievement. In comparing the mean CEPT scores in the post-test of the CLIL and non-CLIL groups, a slight difference of 0.42 was observed (see Figure 1). Based on the numeric scores, the average CEFR level of all the participants at the start and at the end of the study was A2 (20–29 points).
To determine whether the numeric results were statistically significant, the SPSS Version 23 program was used. The Wilcoxon statistical tool was considered appropriate for two related data samples. The level of significance, also known as the critical value, was set at 0.05 (p<0.05 or a 95% confidence interval), thus indicative of a 5% probability that the results are due to chance. According to Doane (2016), the $p$ value approach to testing a mean contends that a small $p$ value (close to 0.00) will likely contradict a null hypothesis ($H_0$). In other words, the $p$ value is a measure of the strength of the evidence against the null hypothesis. Thus, should the $p$ value be less than the selected level of significance ($\alpha$), the null hypothesis can be ruled as false. For the study, the null hypothesis ($H_0$) was based on there being no change in the mean CEPT pre and post test scores for the CLIL and Non-CLIL groups, while the alternative hypothesis ($H_1$) was based on there being differences between these scores.

Table 4 shows the outputs of the Wilcoxon signed ranks test for the CLIL group. The results of the test show that the Asymptotic Significance (2-tailed) or $p$ value is less than 0.05, which results in the acceptance of the null hypothesis ($H_0$). This result can be interpreted as an indication of no statistically significant gains in the CLIL group’s mean proficiency scores after the CLIL intervention ($Z= -1.41$, $p=.888$).

**Figure 1.** Numeric averages for Pre and Post CEPT test
Table 4. Wilcoxon signed ranks test for CLIL group CEPT pre and post-test scores

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>CLIL Post Test - CLIL Pre test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Z )</td>
<td>(-.141^b) ( Asymp. Sig. (2-tailed) )</td>
</tr>
</tbody>
</table>

a. Wilcoxon Signed Ranks Test  
b. Based on negative ranks.

Source: SPSS 23 output based on study data.

Table 5 shows the outputs of the Wilcoxon signed ranks test for the Non-CLIL group. The results of the test show that the Asymptotic Significance (2-tailed) or \( p \) value is less than 0.05, which results in the acceptance of the null hypothesis (\( H_0 \)) instead of the alternative hypothesis (\( H_1 \)). Therefore, this result can be interpreted as an indication of no statistically significant gains in the non-CLIL group’s participants’ mean proficiency at the end of the intervention (\( Z = -.256^b \) \( p = .798 \)).

Table 5. Wilcoxon signed ranks test for CLIL group CEPT pre and post-test scores

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>Non CLIL Pre-test - Non CLIL Post-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Z )</td>
<td>(-.256^b) ( Asymp. Sig. (2-tailed) )</td>
</tr>
</tbody>
</table>

a. Wilcoxon Signed Ranks Test  
b. Based on positive ranks.

Source: SPSS 23 output based on study data.

To determine whether the apparent difference of 0.42 in the mean CEPT post-test scores for the CLIL and Non-CLIL groups was statistically significant, a final Wilcoxon signed ranks test was performed. As seen in Table 6, the Asymptotic Significance (2-tailed) or \( p \) value is less than 0.05, which results in the acceptance of the null hypothesis (\( H_0 \)). Therefore, there was no statistically significant difference in the mean scores of both groups after the CLIL intervention (\( Z = -.254^b \) \( p = .799 \)).
Table 6. Wilcoxon signed ranks test for CLIL and Non-CLIL group CEPT post-test scores

<table>
<thead>
<tr>
<th>Test Statistics(^a)</th>
<th>Non CLIL Post Test - CLIL Post Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>(-.254)(^b)</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.799</td>
</tr>
</tbody>
</table>

\(^a\) Wilcoxon Signed Ranks Test  
\(^b\) Based on negative ranks.

Source: SPSS 23 output based on study data.

Qualitative analysis

The study also included the analysis of the answers given by the CLIL group of students when completing the surveys. This information was gathered, assessed and classified into general categories based on the words of greatest impact and repetition. Not all items total 100%, since students were afforded the opportunity to choose more than one option. Students were also free to add any additional comments.

General perspectives about CLIL

CLIL methodology represented a new approach to learning English for the students, since only one had heard of it previously. They felt this method offered a practical, interactive and dynamic approach, with content directly related to their curriculum (see Figure 2).

Figure 2. General perceptions of CLIL

Source: Compiled by authors.
Improved skills and perceived degree of difficulty

Students reported feeling that their oral production, as well as their vocabulary, reading and listening comprehension improved in the CLIL classroom. Grammar was the least cited skill among their answers (see Figure 3). In comparing the degree of difficulty when learning English through an ESP vs. CLIL approach, most students considered CLIL to be “easier,” which was revealed to be directly related to the absence of grammar and vocabulary quizzes throughout the semesters and the use of multiple sources instead of a single textbook. During the focus groups conducted after the intervention regarding this aspect, students described the precision required to obtain high grammatical scores (much like for a math test) as a source of great stress and performance anxiety. Thus, with no grammar quizzes to prepare for and grammatical precision constituting only a minor component of their oral evaluations, students noted they felt they could spend more time focusing on the content learned in class, which they agreed required more concentration. The term “easy” also implied a certain freedom from the daily use of a single textbook, which many students found to be tedious. For them, the CLIL classroom was more engaging, since “we work with real life content, and we do not focus on a course book,” as one student indicated. The students who marked CLIL as “more difficult” cited as a reason their self-reported lack of English proficiency, which they viewed as a barrier to quickly understand the content (See Figure 4).

Figure 3. Improved language skills through self-assessment

Source: Compiled by authors.
Advantages and disadvantages of CLIL

Seventy-seven percent of students said they preferred CLIL over the ESP approach they had been exposed to up until the start of the study. While none claimed to prefer the ESP approach over all, two students did suggest a combined method would be ideal (see Figure 5).

When asked about perceived advantages and disadvantages associated with CLIL, students indicated that CLIL offered them more opportunities since contents studied were more closely related to their academic program and future profession. Other advantages cited include the sensation of feeling more self-sufficient when presenting projects, as well as communicating more fluently and using a more specialized vocabulary (see Figure 6).

As for disadvantages, three students cited difficulties in understanding some lessons and greater effort required to comprehend others. One student noted he would have liked to receive additional grammar instruction. It is important to remember that the questionnaire did not include direct questions regarding advantages or disadvantages; rather, it was the students themselves who provided their opinions in the comments sections of the questionnaire.
Discussion

After the two-semester intervention, there were no statistically significant increases or decreases in the proficiency levels of the CLIL group. In addition, the CLIL group did not experience any advantages over the non-CLIL group, a finding that is consistent with past studies comparing the language gains between control and experimental groups exposed to CLIL (Arribas, 2016; Graham et al., 2018; Merino & Lasagabaster, 2018). With an average A2 level at the onset and at the end of the intervention (as measured by the CEPT), both groups were at what is considered an elementary level of language proficiency, which involves using commonly used expressions and speaking in simple terms to accomplish routine activities (Council of Europe, 2019). Despite the widespread appreciation for CLIL as a classroom approach, the stagnant low proficiency among the students before and after exposure to CLIL is consistent with the notion that one of critical elements that can influence the success of a CLIL program is the language level of learners (Bruton, 2013; Paran, 2013).

The role of intensity in CLIL outcomes is also worth highlighting, since some authors have suggested that significant improvements in English as a Foreign Language (EFL) skills through CLIL can be best achieved through high intensity instruction (Merino & Lasagabaster,
In this study, five hours of CLIL instruction per week may not have been sufficient to yield statistically significant results in proficiency as measured by the CEPT, even though all the participants passed the course.

In analyzing the qualitative results, the data showed that the participants saw CLIL as a valuable learning tool that helped them feel more connected to their chosen academic program. In other words, instead of considering English as just one more subject, it instead became a tool to acquire relevant content. This finding is linked to the fact that communicative approaches to language teaching tend to focus on acquiring language skills in preparation for a future need that may or may not arrive; in CLIL environments, the content itself serves as a more solid link between students and the language (Lorenzo, Casal, & Moore, 2010). All participants stated they would take a CLIL course again in the future, if given the opportunity, and that they would support future levels of English instruction with a CLIL approach. They also indicated that the CLIL course required a shift in thinking, listening and working, since the flow of each unit and activities were no longer as predictable as they had been in the ESP textbook. This result is not surprising, since CLIL is known to call on students to use their higher order thinking skills to be able to analyze, synthetize, discuss, evaluate, and interpret content with and through the foreign language (Cummins, 2008). However, this different type of effort did not appear to dampen the students’ appreciation and enthusiasm for the approach. Since ESP tends to prioritize grammatical precision far more than CLIL (Yang, 2016), the CLIL course allowed students a certain latitude for grammatical mistakes as they focused on content. This helped mitigate feelings of anxiety, which is known to increase in classroom environments where grades depend on grammatical accuracy (Horwitz, Horwitz, & Cope, 1986). Thus, oral production was cited as the most improved language skill, with students commenting that they felt more relaxed when preparing and presenting oral projects and giving oral content exams as part of the regular exam period at the university. Given this relaxed attitude, it is understandable why the students viewed CLIL as an “easier” approach than ESP even though mastering the content required greater concentration; their comments suggest they weren’t fully conscious of the higher-order thinking skills involved.
From the perspective of the researchers of the study, who were also the classroom teachers, the creation and preparation of CLIL materials required a greater degree of effort and time, as opposed to using the ESP textbook, which came pre-loaded with all classroom activities, tests, and exercises. However, the switch from ESP to CLIL every unit meant that the teachers also experienced a surge of creativity and flexibility afforded by the realistic content. In contrast with the predicable nature of the ESP course driven by the textbook, the CLIL class offered an entertaining and more authentic environment. Thus, teachers felt the same freedom as the students, despite the greater cognitive workload, which is a well-known characteristic of the approach.

Conclusions and Recommendations

The CLIL course changed the way the students had been accustomed to learning English in the ESP class, where there was great emphasis on acquiring specific vocabulary and grammar to accomplish a given task in each unit (usually by memorization). By making the content the core focus, students exceeded their own expectations and experienced little to no anxiety despite the fact that the content was far deeper and more detailed than the ESP textbook (which covers content superficially). Thus, it is not surprising they suggested keeping CLIL in the Tourism program. However, with the average A2 level the students started with at the start of the study remaining unchanged, moving from an ESP to a CLIL-based instruction as a means of improving proficiency may have been premature.

Despite the stagnant scores, the warm enthusiasm and self-reported increases in confidence among the participants warrants continued research on the implementation of CLIL that measures these variables and takes into account proficiency levels and number of hours of instruction. Given the neutral effect of CLIL in this particular study context, widespread implementation of this approach may be more suitable for a series of intensive upper level courses that can be taken by students approaching the end of their academic program, when they have had the opportunity to participate in Tourism-related
site visits to boost their Cognitive Academic Language Proficiency. To help ensure students are able to cope, it also recommended that English language instruction, either in the form of ESP or EFL (English as Foreign Language), continue to form part of the Tourism program in lower semesters with increased hours of instruction.

**Limitations**

Among the limitations of the study is that the CEPT did not include a component to measure oral and written production, nor did it measure Tourism-related content. While the study participants felt that their oral production skills had improved with the CLIL approach, this was based on self-assessment and could not be verified empirically. Therefore, future research should include oral and written production skills evaluation, which have been frequently cited as the areas that most reflect the benefits of CLIL. To shed further light on the role of language proficiency, further research could also include a comparison of language gains achieved in a higher education CLIL classroom comprised of students with a lower language level versus those with a higher proficiency level.

**References**


Burger, S., Weinberg, A., & Wesche, M. (2013). Immersion studies at the University of Ottawa: From the 1980s to the present. Cahiers de L’ilob, 6, 21-43. DOI: https://doi.org/10.18192/olbiwp.v6i0.1130


Ravelo, L. (2014). Demystifying some possible limitations of CLIL (content and language integrated learning) in the EFL classroom. *LACLIL, 7*(2), 71-82. DOI: http://doi.org/10.5294/laclil.2014.7.2.4


Appendix A – Sample classroom materials

**Business Scenario Unit 1 – Customer Service in the Airline Industry**

**PART I. Read the following case. Discuss it with your classmates, analyze the problem.**

In April 2017, United Airlines personnel and security officers at Chicago O’Hare International Airport violently removed Dr. David Dao from
flight 3411. The reason? The flight was oversold, and United wanted the seats for its own employees.

The backlash came almost immediately. People raced to attack the airline. One of the security officers was placed on administrative leave of absence. United’s CEO was under fire, and forced to release an apology calling the event “truly horrific.” New rules and regulations regarding oversold flights were put into place.

This incident is an isolated and extreme example, yet it seems to confirm what many people think about air travel today — that the airlines would rather make easy money than care about customers. It has become very clear to observers that customer service should matter to organizations in all industries. An organization simply does not exist without customers, and customer service represents the active role organizations play in connecting with those customers. Good customer service might be a differentiator for companies that provide similar products, and superior customer services might justify premium pricing in some industries.

How Airlines Can Improve Customer Service

Airlines may never have perfect customer service, because their operations are so complex and there are many conditions, such as the weather, that are outside their control but nonetheless affect passengers. However, airlines can improve by training their people to focus on the following:

1. **Clear communication:** Air travel can be a confusing experience, and a customer service representative must be able to give information in a way that’s timely and useful. The airlines have been making some progress in this through technology: Airline apps can inform people when gates have been changed or boarding times have been delayed.

2. **Empathy:** Customer service people need to be able to relate to any passenger, even one that is having a terrible day because of a delayed flight or missed connection. Airlines must convey understanding when dealing with these customers.

3. **Patience:** A customer service representative should never turn his or her attention to the next customer until the current one has been completely satisfied.
4. **Social media savvy:** In the past, when customers had a complaint, they’d speak to a gate agent or call the airline. Today, they’re more likely to post about their problem on social media, including Twitter and Facebook — which means the world is watching to see how the airlines react. Because of this, airlines need to have highly responsible social media teams that can offer quick, helpful, understanding responses. If companies do not give assistance, or even respond in a timely matter, their reputation can be hurt.

**PART II. Answer the following questions about the United Airlines case.**
Where appropriate, use modal verbs like should, have to, must, etc.

1. What is your opinion on how United Airlines managed the situation?
2. Could have United Airlines managed the situation in a different way? What do you think they could do better?
3. What is your group’s opinion on the practice of overbooking (over-selling) flights in general?
4. The practice of overbooking is legal and will likely stay the same. What do you think airlines should do in similar cases if they have an overbooked flight?

**WORKSHEET: Problem solving scenarios**

**PART I. Read the following case. Talk with your group members and make recommendations on how you think the airline’s customer service staff should respond based on the previous article. Research online. Present your recommendations using modal verbs like should, have to, must, etc.**

National Airlines had an unhappy customer. Genna Morris flew from Washington, DC, to Los Angeles. The flight stopped at Denver International Airport, where she got off the plane for 30 min. When she returned to her seat, her $500 prescription reading glasses were gone. She asked the flight attendant where the glasses were, and the attendant said the glasses were probably thrown away since the cleaning crew had come in with big bags and put everything in them.
Ms. Morris tried to find the glasses through the airline’s lost-and-found service, but without success. Then she wrote a letter to the airline asking for compensation. Note: Flight attendants usually announce that the plane is continuing to another city and that passengers who are returning should take their things with them. Cabin cleaning services go through planes quickly to take away newspapers, magazines, leftover foods, and trash. Airlines usually feel no responsibility for personal items left in cabins.

**Appendix B – Student questionnaire**

**Cuestionario de opinión - AICLE**

Las siguientes preguntas tienen como objetivo conocer su opinión acerca de las actividades de instrucción basada en contenidos que se realizan en clase. Por favor responda a todas las preguntas con sinceridad.

¿Había usted recibido instrucción basada en contenidos antes de este curso? De ser así, indique en dónde.

SI [ ]  NO [ ]

¿Le parecieron más difíciles o más fáciles las clases de materia en inglés que las clases regulares enfocadas en gramática? Explique:

___________________________________________________________________
___________________________________________________________________

¿Qué prefiere: aprender inglés a través de contenidos/materia o por medio de gramática? Explique:

___________________________________________________________________
___________________________________________________________________

¿Cree que aprender inglés a través de contenidos/materia le ayuda en su aprendizaje? ¿Tomaría un curso similar si tuviera la oportunidad? Explique:

___________________________________________________________________
___________________________________________________________________
¿Usted cree que aprendió más con las clases de inglés basadas en contenidos/materia que las de gramática? ¿O igual? Explique:

___________________________________________________________________
___________________________________________________________________

¿En general, usted cree que la instrucción basada en contenidos/materia puede ayudar a mejorar algunas de las siguientes destrezas? (marque todas las que apliquen)

<table>
<thead>
<tr>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expresión oral</td>
</tr>
<tr>
<td>Vocabulario</td>
</tr>
<tr>
<td>Estructura gramatical</td>
</tr>
<tr>
<td>Lectura</td>
</tr>
<tr>
<td>Destreza auditiva</td>
</tr>
</tbody>
</table>

Otra: ___________________________________________________________________
___________________________________________________________________

Preguntas para grupo focal

1. En su opinión, ¿cuáles son las principales ventajas y desventajas de estudiar materia en inglés?
2. ¿Qué es lo que más le gusto/disgustó de sus clases AICLE?
3. ¿Qué comentarios adicionales tiene acerca del método AICLE?