How Effective is the Assessment Component of a Customized CLIL Program?

¿Qué tan efectivo es el componente de evaluación de un programa AICLE personalizado?

Quão eficaz é o componente de avaliação de um programa CLIL personalizado?

Ali Işık
https://orcid.org/0000-0002-3305-7922
Istinye University, Turkey
ali.isik@istinye.edu.tr

Received: 23/03/2020
Sent to peer review: 08/04/2020
Accepted by peers: 01/07/2020
Approved: 13/07/2020

DOI: 10.5294/laclil.2020.13.2.5

To reference this article (APA) / Para citar este artículo (APA) / Para citar este artigo (APA)

Işık, A. (2020). How effective is the assessment component of a customized CLIL program? *Latin American Journal of Content & Language Integrated Learning*, *13*(2), 241-287. https://doi.org/10.5294/laclil.2020.13.2.5

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

ABSTRACT. Considering the pivotal role of assessment, this study aimed to investigate the attitudes of the students and the teachers towards the assessment component of a customized content and language integrated learning in an English as a foreign language program implemented at the tertiary level in Turkey. It also sought to study its effectiveness as a tool for the integrated assessment of language and content. Data were obtained by a mixed-method research approach from 525 university freshman students and 17 English language teachers via questionnaires and follow-up interviews with the teachers and the students. The results indicated that both the students and the teachers developed positive attitudes towards the assessment component of content and language integrated learning. The assessment component was also found to be an adequate tool for the integrated assessment of content and language.

Keywords (Source: Unesco Thesaurus): Content and language integrated learning; assessment; learner attitude; teacher attitude; teacher training.

RESUMEN. Teniendo en cuenta el papel fundamental de la evaluación, el objetivo de este estudio era investigar las actitudes de los estudiantes y los profesores hacia el componente de evaluación de un aprendizaje integrado de contenido y lengua personalizado en un programa de inglés como lengua extranjera implementado en el nivel terciario en Turquía. También buscaba estudiar su efectividad como herramienta para la evaluación integrada del lenguaje y el contenido. Los datos se obtuvieron mediante un enfoque de investigación de método mixto de 525 estudiantes universitarios de primer año y 17 profesores de inglés a través de cuestionarios y entrevistas de seguimiento con los profesores y los estudiantes. Los resultados indicaron que tanto los estudiantes como los profesores desarrollaron actitudes positivas hacia el componente de evaluación del aprendizaje integrado de contenido y lenguaje. El componente de evaluación también resultó ser una herramienta adecuada para la evaluación integrada del contenido y el lenguaje.

Keywords (Source: Unesco Thesaurus): aprendizaje integrado de contenido y lengua; evaluación; actitud del alumno; actitud del profesor; formación de profesores.

RESUMO. Levando em conta o papel fundamental da avaliação, o objetivo deste estudo foi investigar as atitudes de alunos e professores sobre o componente de avaliação de uma aprendizagem integrada de conteúdo e linguagem em um programa de inglês como língua estrangeira implementado no nível terciário na Turquia. Também buscou estudar sua eficácia como ferramenta de avaliação integrada de linguagem e conteúdo. Os dados foram obtidos usando uma abordagem de pesquisa de método misto de 525 estudantes universitários do primeiro ano e 17 professores de inglês por meio de questionários e entrevistas de acompanhamento com professores e alunos. Os resultados indicaram que alunos e professores desenvolveram atitudes positivas em relação ao componente de avaliação da aprendizagem integrada de conteúdo e linguagem. O componente de avaliação também acabou sendo uma ferramenta adequada para a avaliação integrada de conteúdo e linguagem.

Keywords (Source: Unesco Thesaurus): aprendizagem integrada de conteúdo e língua; avaliação; atitude do aluno; atitude do professor; treinamento de professor.

Introduction

Although content language integrated learning (CLIL)[1] has offered a fresh perspective on language education and led to the reconceptualization of the approach to language and learning, pedagogical aspects of language education, nature of a language program, and the roles of stakeholders in the language education process, assessment in CLIL has not aroused the same amount of interest among researchers (Lo & Fung, 2018; Morgan, 2006). The empirical evidence and theoretical discussions on CLIL assessment (CLIL-A) are far from being proportional to the popularity of CLIL. Hence, CLIL-A represents the underdeveloped aspect of CLIL and lacks a common solid theoretical and empirical basis (Barbero, 2012; Maggi, 2012; Massler, Stotz, & Queisser, 2014; Otto, 2018; Reierstam, 2015).

Since CLIL is bifocal, ultimately, CLIL-A needs to address assessing both language and content and achieve the balance between the two (Massler, 2010; Short, 1993; Tedick & Cammarata, 2006). As assessing both is not a common practice and foreign to many English language teaching (ELT) teachers and subject matter teachers, it is hard to find common ground on how to carry out a sound and valid assessment reflecting the nature of CLIL. Moreover, it is claimed that synchronous integrated assessment of language and content creates a dilemma to calculate the effect of each on students' performance (Douglas, 2010; Maggi, 2012; Massler et al., 2014; Wewer, 2014).

Another, yet related issue, is CLIL-A literacy of teachers, which is another underdeveloped aspect of CLIL (Barrios & Milla-Lara,, 2018; Massler, 2010). CLIL-A extends the basic requirement of ELT teachers and subject matter teachers and requires them to assume new roles. First, CLIL teachers need to possess general assessment literacy skills to design, implement and evaluate effective assessment and customize assessment for their local contexts (Purpura, 2016; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014). Besides, they are to be equipped with the knowledge and skills specific to content-based assessment

[1] CLIL is used as an umbrella term for all versions of the content-oriented language programs, including content-based instruction (CBI).

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

(Maggi, 2012). They need to know and practice how to employ a variety of CLIL-A techniques to integrate language and content and manage a fine balance between them. Moreover, they are to adjust the level of cognitive operations required for each task in terms of content and language proficiency. In other words, CLIL-A is more demanding than a general language assessment practice and requires more specific assessment literacy.

To sum up, the theoretical framework of CLIL-A has been unsettled so far in terms of assessing both the content and language learning of students (Massler et al. 2014; Otto, 2018). Thus, assessing integrated content and language is an intricate issue and poses a rigorous problem. Ultimately, what is essentially needed in CLIL-A is a framework to bring the strands together to assess program objectives fairly, validly, and reliably and to provide feedback to the stakeholders to be exploited for evaluation and amelioration.

Literature review

To provide a framework for CLIL assessment and to train and guide teachers comprehensive projects were launched in Europe. Language in Content Instruction (LICI, 2009) was initiated to address a wide range of issues in CLIL including an assessment grid based on the Common European Framework. In the same vein, the Assessment and Evaluation in CLIL Project (AECLIL, 2013) aimed to provide a perspective focused on effective assessment and evaluation in CLIL. Similarly, CLILA (Massler et al., 2014) proposed an assessment framework to assess both language and content learning in primary schools. Some European countries followed the same path. The Republic of Ireland (2007), Scotland (2010), and Portugal (2016) started assessment projects to provide national guidelines and standards for CLIL assessment. However, no research studying the effectiveness of these projects has been reported, to the knowledge of the researcher.

Researchers also attempted to provide an alternative perspective or approach to assessment in CLIL. In their seminal work on assessment in CLIL, Coyle et al. (2010) outlined the principles of assessment

in CLIL and tried to answer the questions about what is assessed, how it is assessed, when it is assessed, and by whom it is assessed. They also suggested a framework to assess content and language. O'Dwyer and de Boer (2015) reported two case studies from two Japanese universities and concluded that, when learner involvement and collaboration were encouraged in CLIL-A, learners assumed more responsibility to self-regulate and self-assess their learning. It was also found that active learner involvement in the assessment process led to efficient use of language skills when handling both language and content. In another study, where the English language skills and content learning of Portuguese students at the early primary level were assessed, Xavier (2016) reported the lack of a common CLIL assessment framework and the need for teacher training in assessing CLIL. Ultimately based on the findings, a sample assessment framework derived from the learning-oriented approach was proposed as a basis for assessing CLIL and teacher training. In Colombia, Leal (2016) proposed a three-dimensional assessment grid composed of Cognitive Academic Proficiency (CALP) functions, cognitive skills, and content. It was found that the grid helped teachers to balance language and content by considering the language demands and the level of difficulty of test items. Peña (2017) investigated the effect of assessment for learning in CLIL in Spanish primary schools and found that it was beneficial for both content and language learning. Likewise, inspired by the functional view of language, Otto (2018) proposed the Functional Model to assess language in CLIL by emphasizing the essential role of language in academic discourse. Although all these individual initiatives have pinpointed the problems in CLIL-A and presented different perspectives to overcome them, they have failed to spark a movement in CLIL-A on the grand scale.

Researchers also tried to explain CLIL-A through empirical evidence. Serragiotto (2007) surveyed CLIL assessment practices in Italian schools and found that there was no common understanding about the weight of language and content in assessment. It was also indicated that there was no common CLIL-A framework and consequently, no systematic assessment of content and language was observed.

Findings of the CLIL Learner Assessment Project (CLILA) targeting to determine how CLIL assessment was practiced in elementary

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

and pre-elementary education in Germany and Switzerland revealed that there were no clear-cut guidelines for teachers on how to assess and manage the balance between content and language (Massler et al., 2014).

Likewise, having studied content-based assessment practices in Finnish primary schools, Wewer (2014) reported that there was no common framework to collect data systematically, and assessment was carried out fortuitously. In another study, Reierstam (2015) observed little to no difference between CLIL and non-CLIL in terms of language-related assessment procedures and pointed out a need for teacher training for a sound content-based assessment. Further similar evidence was obtained in Greece by Zafiri and Zouganeli (2017), who reported that the teachers tried to assess both content and language; however, the assessment practice was not systematic and satisfactory, and there was no assessment framework. In a similar vein, Barrios and Milla-Lara (2018) conducted a survey in Spain to investigate the assessment component of CLIL and found the teachers could not achieve the balance between content knowledge and target language skills. In another study, Lo and Fung (2018) examined the effect of the target language on the performance of content knowledge on CLIL-A in Hong Kong. They found that, for each content knowledge task, there was a certain level of confounding language demand. They also indicated that, as the grade level of the students increased in the education system, so did the cognitive level of the tasks, which were accompanied by an increasing focus on the productive skills in the target language. In a countrywide study, Zhetpisbayeva et al. (2018) conducted research in Kazakh secondary schools to examine the CLIL-A practices in accordance with the new assessment system, which would be implemented in the 2019–2020 academic year. They found that the subject teachers did not pay enough attention to language skills and the collaboration between the subject matter teachers and English teachers was not satisfactory. Finally, they reported the lack of assessment tools and an assessment framework guiding CLIL-A practices.

To sum up, the research stated above portrays the lack of a common framework to guide CLIL-A, which leads to unsystematic and disorderly content-based assessment practice. It is also observed that the balance between language and content is hard to keep and teachers

tend to favor one over the other, depending on their academic background. Finally, the data reveal the need for teacher training in CLIL-A.

The present study

CLIL-A needs to be aligned with the nature and requirements of a CLIL program (Massler et al., 2014; Morgan, 2006). Thus, how assessment is planned, implemented and evaluated in CLIL is to be studied thoroughly to complement CLIL programs (Inbar-Lourie, 2008). However, the review of literature suggests a need for an exemplary framework and model, especially in Turkey, on how to practice CLIL-A. Moreover, the evidence on CLIL-A is scarce and there is no study on CLIL-related assessment in Turkey. Thus, this study aimed to investigate the effectiveness of a CLIL-A practice implemented at the tertiary level in an English as a foreign language (EFL) context in Turkey. Also, it attempted to display how to assess content and language in a balanced manner considering the goals of an EFL program. Besides, it aims to spark an interest in CLIL-A in the Turkish EFL context and contribute to the insufficient yet evolving empirical evidence on it. Specifically, the study aims to answer the following research questions:

1. What are the attitudes of the Turkish ELT teachers towards the assessment component of the CLIL program?
2. What are the attitudes of the Turkish EFL students towards the assessment component of the CLIL program?
3. How effective is the assessment component of the CLIL program to assess the English language development of the Turkish EFL students?
4. How effective is the assessment component of the CLIL program to assess the academic content knowledge of the Turkish EFL students?

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

The context of the study

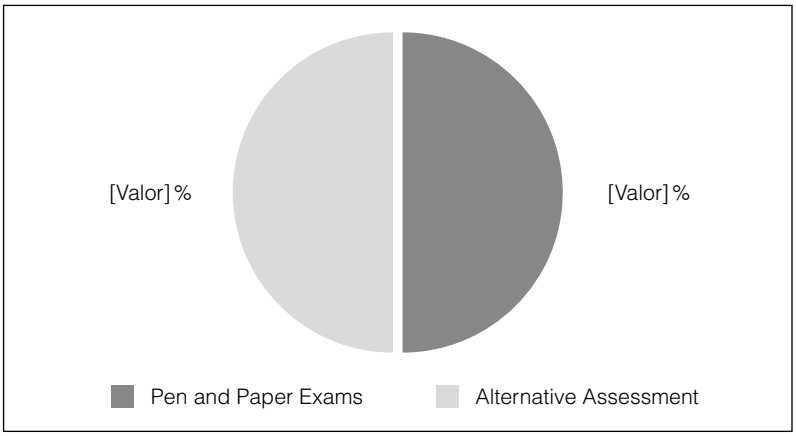
The Program

Integrating CLIL in EFL and, tailored for each academic program at a given university, this particular CLIL program was unique in Turkey, initiated with the slogan of "one foreign language without losing one year." In Turkey, English prep education is considered as a common solution to teach English to students, but it costs a year in the lives of the students, in addition to the economic cost (Isik, 2003, 2008). Unlike the common practice, this particular CLIL program divided the total hours of English education in a regular prep class into four (following the four-year undergraduate program) and distributed them evenly to each year of the four-year academic programs. Each department allotted eight to twelve hours a week for CLIL in its academic program. To realize the goals of the CLIL program and meet the content and English language needs of the students, 17 separate sets of in-house CLIL materials customized for 17 different academic programs were generated and implemented.

CLIL-A

At the tertiary level in Turkey, the common practice is to offer a general EFL Grammar, vocabulary, reading, and, to some extent, writing skills are tested via pen and paper exams. Contrary to the common practice, in this CLIL-based instruction, both process- and product-oriented approaches were adopted to assess language and content. Hence, a customized assessment approach composed of CLIL-based assessment and alternative assessment, each making up 50% of the final grade of learners, was adopted to assess both the knowledge of academic content and English of the students, as depicted in Figure 1.

Figure 1. Overall grade percentages

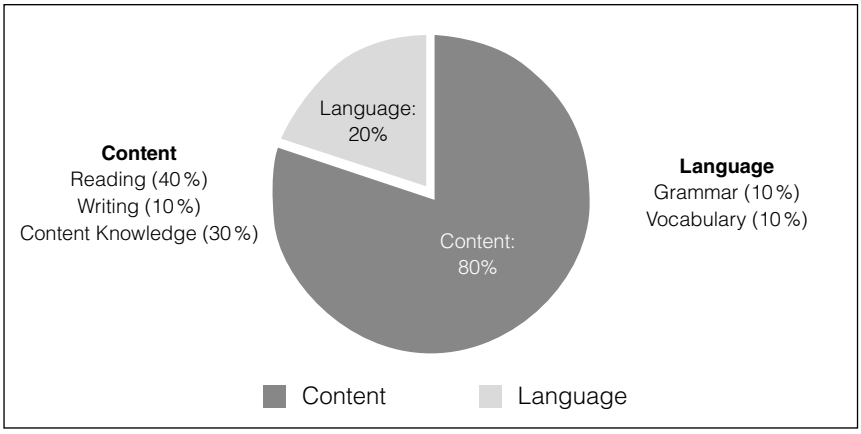


Source: Own elaboration.

Pen and paper exam

The pen and paper exams were implemented as weekly quizzes and monthly exams. A template was developed to systematize assessment procedures and achieve validity (see Appendix 1). The use of language was ingrained in academic content, and they were assessed together. As shown in Figure 2, reading covered 30%, writing 20%, and academic knowledge 30% of the exams. The weight of the grammar and vocabulary was the same, that is, 10% each.

Figure 2. The content of the pen and paper exam



Source: Own elaboration.

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

Alternative assessment

The students were informed by the teachers that assessment was an on-going process and was not limited to merely the pen-paper exams. The teachers exploited a daily self-evaluation chart, individual conference, class observation, and portfolios as the means of the alternative assessment.

Teacher Training

As the EFL teachers participated in the study lacked training and experience in CLIL and CLIL-A, the advisor planned an initial training program for them before the classes started. Having previously designed and implemented CLIL programs and assessment in addition to offering courses in materials development courses and assessment in the ELT departments of major universities in Turkey since 1999, the advisor had enough theoretical knowledge and practice in CLIL. The advisor organized an 80-hour workshop on the basics of CLIL-A, and a template on how to assess was shared with the teachers (see Appendix 1). Consequently, scaffolded by the advisor, the teachers prepared their table of specifications considering the goals of the CLIL courses they would teach. In the meantime, the advisor reminded them that the tasks were to meet the three pillars of assessment, namely, content, language, and cognitive processes (Barbero, 2012). In the training process, the advisor, co-working with each teacher, provided immediate and continuous support. Following the initial training, the teacher training continued throughout the academic year as they started the actual assessment process during the academic year to assess both the language and content knowledge of their students (see Appendix 2).

Student Orientation

Not only the teachers but also the learners needed training, since CLIL-A was new for them. Each class was visited one by one by the program advisor, who briefed the students about the CLIL program and how they were going to be assessed. The teachers also continuously briefed their students on how they would be assessed, stressing the importance of alternative assessment, which the students found to be quite novel.

Methodology

Participants

In this study, a quasi-experimental design was implemented, and the participants, 525 university freshman EFL students and 17 ELT teachers, were selected through convenience sampling. As the CLIL-based program was unique, quasi-experimental design and convenience sampling were appropriate to investigate the effectiveness of the assessment component of the program. For the follow-up interviews, five students from each faculty were selected through random sampling. The students and their faculties were tabulated in Table 1.

Table 1. The participating students and their faculties

Faculties	Health Sciences	Arts and Sciences	Pharmacy	Economics, Administrative and Social Sciences	Fine Arts, Design and Architecture
Number of students	228	45	48	117	87

Source: Own elaboration.

Regarding the teachers, 17 ELT teachers who had no prior training and experience in CLIL took part in the study. Two of the teachers had 10–15 years of teaching experience and 15 of them had 0-5 years of experience. All the teachers were graduates of ELT departments in Turkey, and one of them had a Ph.D. in ELT.

Data collection

A mixed-methods research design was used to collect data with the help of the student and teacher questionnaires, teacher and student follow-up interviews, Oxford Placement Test (OPT), and CLIL-A assessment component.

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

The questionnaires

The student questionnaire (see Appendix 3) and the teacher questionnaire (see Appendix 4) developed for the AECLIL Project (2013) funded by the European Commission were used to collect data from the teachers and the students. The questionnaires were administered in the final week of the 35-week academic year as the participants were assumed to form their attitudes about CLIL-A by then. The internal consistency reliability of the questionnaires calculated using Cronbach's alpha was found to be .82.7 and .79.3, respectively.

The follow-up interviews

In addition to the questionnaires, the researcher carried out follow-up interviews with both the teachers and the students to get a deeper understanding of the assessment process. All the teachers and five students from each faculty selected through random sampling took part in the interviews. The teacher follow-up interviews focused on developing, implementing, and evaluating the effectiveness of the assessment component (see Appendix 5). Likewise, the student follow-up interviews attempted to consider the perceptions of the students about how well the assessment component assessed their content knowledge and English development (see Appendix 6). The follow-up interviews with the students were carried out in weeks 26 and 27 of a 28-week academic year. The interviews with the ELT teachers were held in the last week. The researcher talked with one participant at a time and recorded the interview. The recordings were transcribed for analysis.

OPT

To assess the effectiveness of the language component of CLIL-A, OPT was exploited as a benchmark to compare the cumulative CLIL-A scores of the students to their OPT scores.

CLIL-A exams

CLIL-A exams were used to gauge both the academic content knowledge and English development of the students.

Data analysis

SSPSS was used to analyze the data. The percentages and frequencies obtained from the questionnaires were calculated and presented via

descriptive statistics. The data elicited from the interviews were categorized and coded for evaluation. The Pearson correlation coefficient was computed to assess the relationship between the scores of OPT and the language component of CLIL-A.

Results

Teacher questionnaire

The data obtained from the teacher questionnaire indicated that all the teachers who took part in this particular foreign-language education program had no prior experience in CLIL. Although they had no CLIL and CLIL-A background, all the teachers pointed out that they found their CLIL-A experience very effective. Concerning what to assess, all of them reported that they considered both content and language important when preparing their CLIL-A tasks (Table 2).

Table 2. The importance of factors in assessing content knowledge

	Very important		Important		Partially important		Not important	
	F*	P**	F*	P**	F*	P**	F*	P**
Oral and written skills	15	88.2	2	11.8	0	0.0	0	0.0
Content areas/themes	17	100	0	0.0	0	0.0	0	0.0
Use of content obligatory vocabulary	17	100	0	0.0	0	0.0	0	0.0
Use of content compatible vocabulary	17	100	0	0.0	0	0.0	0	0.0
Mastery of various forms of expression and multimodal instruments	10	58.8	7	41.2	0	0.0	0	0.0
Linguistic accuracy	7	41.2	7	41.2	3	17.6	0	0.0
Linguistic complexity	6	35.3	8	47.1	3	17.6	0	0.0
Mastery of a disciplinary written genre	15	88.2	2	11.8	0	0.0	0	0.0
Analytical skills	8	47.1	7	41.2	2	11.8	0	0.0

*Frequency ** Percentage

Source: Own elaboration.

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

Considering only "very important" and "important" options, all the teachers gave importance to oral and written skills, content, the use of content-obligatory and content-compatible vocabulary, the mastery of various forms of expressions via different instruments, mastery of a disciplinary written genre. The majority of the teachers considered linguistic accuracy and complexity, as well as the analytical skills, important.

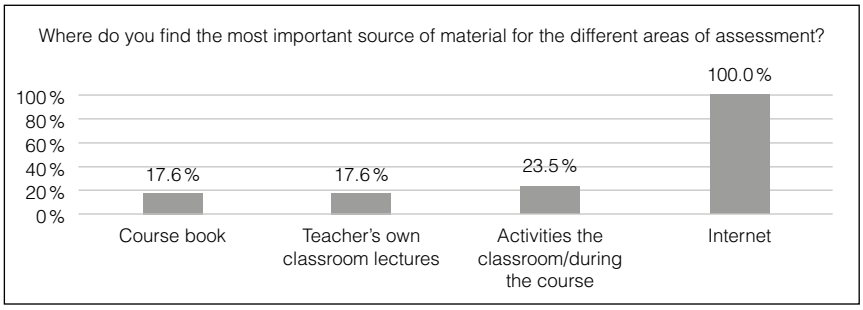
Table 3. Techniques considered to be important in terms of effective assessment of student performance

	Very important		Important		Partially important		Not important	
	F	P	F	P	F	P	F	P
Language portfolio	17	100	0	0.0	0	0.0	0	0.0
Observations of another person	13	76.5	4	23.5	0	0.0	0	0.0
Pupil peer assessment	11	64.7	6	35.3	0	0.0	0	0.0
Pupils' self-assessment or reflection	17	100	0	0.0	0	0.0	0	0.0
Simulations	17	100	0	0.0	0	0.0	0	0.0
Essays or compositions	12	70.6	5	29.4	0	0.0	0	0.0
Presentations (e.g., feedback forms)	11	64.7	6	35.3	0	0.0	0	0.0
Dialogues, interaction	14	82.4	3	17.6	0	0.0	0	0.0
Oral tests or interviews	10	58.8	7	41.2	0	0.0	0	0.0
Written tests or test sections	17	100	0	0.0	0	0.0	0	0.0
Teacher observation	17	100	0	0.0	0	0.0	0	0.0

Source: Own elaboration.

All the teachers considered all the techniques important to assess student performance effectively. They indicated that language portfolios, students' self-assessment or reflection, simulations, written tests or test sections, and teacher observation were very important (Figure 3).

Figure 3. The sources of materials for CLIL-A tasks



Source: Own elaboration.

All the teachers indicated that they exploited the internet for developing their CLIL-A materials. About a quarter of them stated that the activities in the classroom during the course formed the basis of their CLIL-A tasks. About one-sixth of the teachers reported that they made use of the coursebook and classroom lectures to develop CLIL-A tasks.

As for giving feedback to their students, all the teachers stated that they always provided feedback to their students about both their language performance and mastery of the content knowledge. When answering the question about the methods, they provided information to their students on their language development and content learning, and they all reported that they used school-year reports, class discussion or mutual feedback, and oral and written feedback to their students.

The problems the teachers encountered were summarized in Table 4.

In terms of options "always or very often" and "often," all the teachers indicated that combining content and language, adapting the cognitive level of the tasks to the level of the students, and not having enough content knowledge were all demanding. Likewise, the majority of the teachers indicated "adapting the language of the tasks to the level of the students" as a major problem.

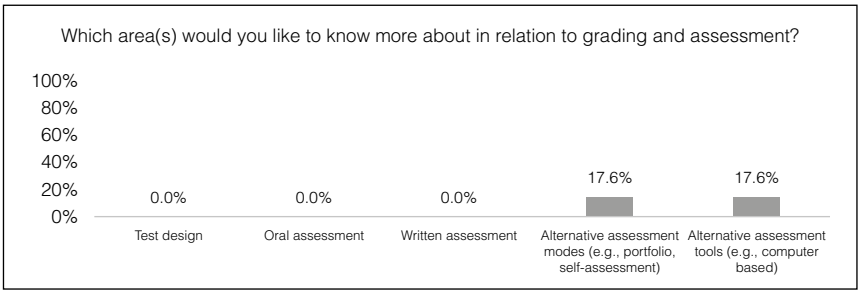
LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

Table 4. Problems encountered

	Always or very often		Often		Sometimes		Seldom/ never	
	F	P	F	P	F	P	F	P
Adapting language	9	52.9	5	29.4	3	17.6	0	0.0
Combining content and language	11	64.7	6	35.3	0	0.0	0	0.0
Adapting the cognitive level of the tasks	13	76.5	4	23.5	0	0.0	0	0.0
Not having enough knowledge of the content	16	94.1	1	5.9	0	0.0	0	0.0

Source: Own elaboration.

Figure 4. The areas teachers need to know more about

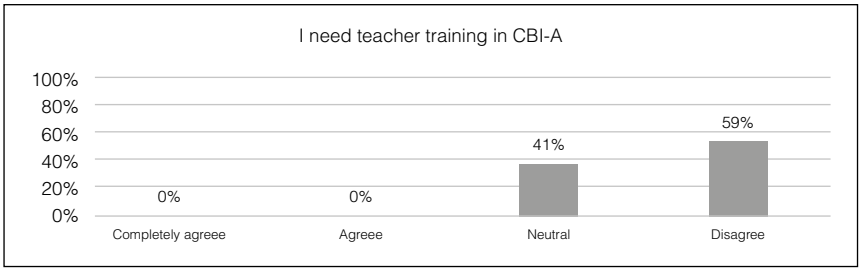


Source: Own elaboration.

All the teachers stated that they knew enough about test design, oral and written assessment. About one-sixth of the teachers said that they would like to know more about alternative assessment modes and tools.

None of the teachers marked the "completely agree" and "agree" option regarding the need for teacher training in CLIL-A.

Figure 5. The need for teacher training



Source: Own elaboration.

Teacher Interviews

Academic background

The ELT teachers graduated from ELT departments and reported that CLIL just came out as one of the syllabus types. Hence, they mentioned that they had no experience in CLIL-A.

The perception of CLIL-A

The teachers all agreed that CLIL-A was challenging. Since there was a need to reflect the bifocal instruction onto assessment, they needed to reflect the weight of the content and language mastered in the syllabus proportionally onto assessment. The second challenge they pointed out was finding the texts on the subject matter covered in the program that were linguistically appropriate to the level of the students. They also needed to consider the cognitive level of the tasks that were prepared considering Bloom's Taxonomy and design an array of tasks requiring lower-order and higher-order operations. Fourteen of the teachers express that they never felt comfortable with CLIL-A because they were not entirely sure about whether the assessment they designed fully covered both the content and language goals. Thirteen of them stated that, despite the common grading framework used to grade alternative assessment, they were never sure about how fair and systematic their grades were. All the teachers said that CLIL-A was extremely time consuming and, aside from offering a CLIL course that already drained their time, planning, implementing, and evaluating CLIL-A created a lot of time pressure on them.

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

Qualifications for CLIL-A

They all reported that the initial training on CLIL-A, the ongoing training via real assessment, continuous training, and scaffolding provided by the advisor made them assess effectively. They evolved their knowledge and skills by engaging in actual practice in real-life contexts. Five of the teachers indicated that they needed to improve their skills in evaluating and scoring alternative assessment.

Effectiveness of CLIL-A

All the teachers stated that CLIL-A was quite effective to assess both the program objectives and student performance. Both the language and content covered in the program were reflected proportionally well enough to develop a fair and valid assessment. They also indicated that weekly individual conferences with the students were found to be very fruitful to evaluate their weekly performance and daily self-assessment reports. Hence, CLIL-A turned into an ongoing process through which the teachers provided immediate feedback to their students to improve their learning process. It was also exploited by the teachers to revise and improve their teaching practices. On the other hand, six of the teachers pointed out that, especially within the first month, some students experienced problems with the alternative assessment. Since it was quite novel for them, they either did not know what to do or underestimated its role. Similarly, eight teachers indicated that few students did not grasp the need for self-assessment reports and filled them out of obligation and fear of being evaluated negatively by their teachers.

Washback

All the teachers pointed out that, as the students knew that they would be assessed on content, the students felt they were obliged to pay attention to content, not only language forms. In short, the teachers reported that CLIL-A had a positive impact on the language program. On the other hand, two of the teachers indicated the dual assessment focus was quite new for the students and some students failed to adapt themselves to that novel practice, and CLIL-A doubled the burden for some students. To ease the burden, some students referred to resources in Turkish, their mother tongue, to better understand the

content while preparing for exams and to obtain a better score on the exam. In that sense, some students discovered a loophole to learn content, which might have undervalued the value of covering content in English. Finally, all the teachers reported that the students saw the value of what they did in English as pleasure in their free time was taken into consideration while evaluating their performance, which fueled more free-time activities in English.

The resources exploited for CLIL-A

All the teachers stated that the CLIL program they taught was customized and they had to develop their assessment tasks. They referred to the internet to find the texts and adapted them in terms of the academic and linguistic content regarding the levels of their students. Six of the teachers reported that, as they were not native speakers of English, they felt the need to have their texts edited by a native speaker. Seven of the teachers said that they also exploited the materials provided by the lecturers from the faculty for which they prepared CLIL materials. However, they needed to adapt them as well to make them appropriate for the level of their students.

The problems encountered

The teachers found CLIL-A massively demanding, as it required them to assess both language and content, which was quite new for them. Assessing academic content, which was not their expertise, was particularly challenging. In the same vein, nine of the teachers pointed out that they felt a rigorous time pressure in developing assessment tasks for both language and content. Likewise, all the teachers reported that CLIL-A was quite a new practice for the students, and it took some time for the students to get used to such an assessment type that concurrently focused on content. Regarding the alternative assessment, eleven teachers made it clear that it was completely new for the majority of the students, who could not believe that it would affect their final performance grades. Some of the students who got high scores on the pen and paper component of CLIL-A failed because they paid lip service to the alternative assessment or did not fulfill their required tasks in the first semester. Those students objected to their final grades, stating that they got high grades on the pen and paper

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

exams, but they still failed. Although they knew the grading scheme, they did not change their traditional conception of assessment limited only to the pen and paper exams. Such a problem was not experienced in the second semester, since the students realized the importance of the alternative assessment in determining their final grades.

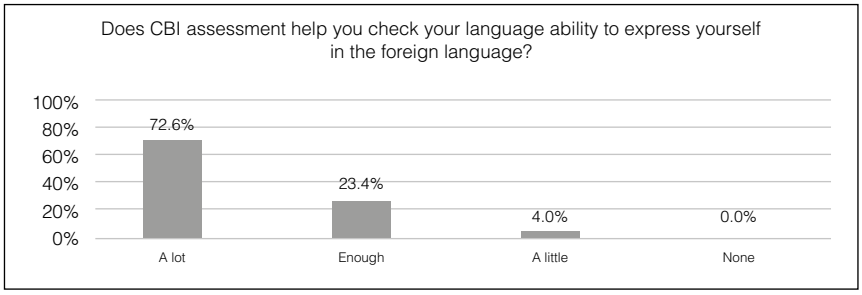
The need for teacher training

The teachers stated that they did not need teacher training. Three of the teachers said that they felt qualified enough for CLIL-A and suggested that they could collaborate with the advisor to train the teachers that would be hired in the following academic year.

Student questionnaire

The data obtained from the student questionnaire are summarized in Figure 6 below:

Figure 6. The effectiveness of CLIL-A in assessing foreign language

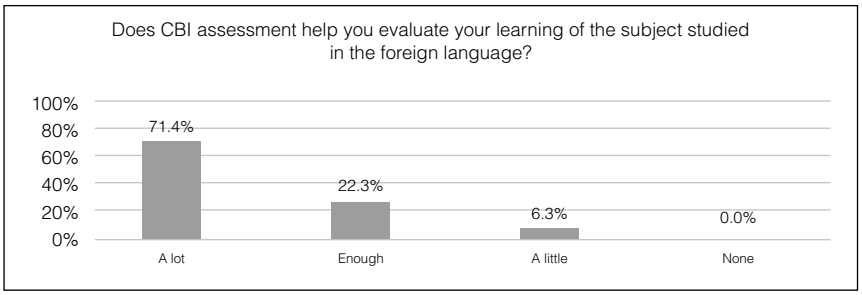


Source: Own elaboration.

For the effectiveness of CLIL-A in assessing language performance, about a quarter of the students found CLIL-A efficient (Table 7).

Regarding the effectiveness of CLIL-A in assessing content, about seven-tenths of the students found CLIL-A efficient (Table 5).

Figure 7. The effectiveness of CLIL-A in assessing content knowledge



Source: Own elaboration.

Table 5. How safe did the students feel about assessment?

	Very comfortable		Comfortable		Partially comfortable		Not comfortable	
	F	P	F	P	F	P	F	P
Oral interchange	21	12.0	90	51.4	36	20.6	28	16.0
Presentations	19	10.9	98	56.0	41	23.4	17	9.7
Pupil peer assessment	35	20.0	79	45.1	44	25.1	17	9.7
Pupils' self-assessment or reflection	83	47.4	89	50.9	3	1.7	0	0.0
Simulations	27	15.4	93	53.1	41	23.4	14	8.0
Written tests or test sections	114	65.1	61	34.9	0	0.0	0	0.0
Essays or compositions	119	68.0	56	32.0	0	0.0	0	0.0
Language portfolio	130	74.3	45	25.7	0	0.0	0	0.0

Source: Own elaboration.

When the findings are presented considering only the "very comfortable" and "comfortable" options, about two-thirds of the students reported that they were comfortable with oral interchange, oral presentations, and simulations. Almost all the students felt safe with self-assessment, written tests and essays, and language portfolio. Regarding the importance assigned to language elements summarized in Table 6, all the students rated pronunciation, knowledge of vocabulary,

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

knowledge of content, and clarity of expression as "important" and "very important." About half of the students went for "important" and "very important" options for grammatical correctness.

Table 6. The importance assigned to language elements during task performance

	Very important		Important		Partially important		Not important	
	F	P	F	P	F	P	F	P
The correct pronunciation of words	76	43.4	99	56.6	0	0.0	0	0.0
Grammatical correctness	30	17.1	69	39.4	67	38.3	9	5.1
Knowledge of vocabulary	127	72.6	48	27.4	0	0.0	0	0.0
Knowledge of the contents	129	73.7	46	26.3	0	0.0	0	0.0
Clarity of expression	132	75.4	43	24.6	0	0.0	0	0.0

Source: Own elaboration.

Regarding only the "very important" and "important" options, all the students perceived teacher oral observation helpful to assess their performance. As can be seen in Table 7, the overwhelming majority of the students found the language portfolio, self-assessment, dialogues and interaction, and written tests as techniques useful to reveal their performance. The majority of the students marked observation by another person, peer-assessment, simulations, presentations, and oral tests and interviews as useful tools reflecting their performance.

Concerning feedback, all the students indicated that receiving feedback about their mastery of content and language growth was very important (Table 8).

In terms of feedback, both on the mastery of content and language, all the students reported that they received feedback via teachers, tests, self-assessment, school reports, school assignments, and portfolios. Regarding the preferred feedback means, teacher feedback and self-assessment were the most popular ones.

Table 7. Helpful techniques to assess student performance

	Very important		Important		Partially important		Not important	
	F	P	F	P	F	P	F	P
Language portfolio	127	72.6	39	22.3	9	5.1	0	0.0
Observations of another person	24	13.7	91	52.0	43	24.6	17	9.7
Pupil peer assessment	41	23.4	82	46.9	38	21.7	14	8.0
Pupils' self-assessment or reflection	75	42.9	93	53.1	7	4.0	0	0.0
Simulations	39	22.3	85	48.6	44	25.1	7	4.0
Essays or compositions	79	45.1	86	49.1	10	5.7	0	0.0
Presentations (e.g., feedback forms)	27	15.4	89	50.9	51	29.1	8	4.6
Dialogues, interaction	47	26.9	91	52.0	24	13.7	13	7.4
Oral tests or interviews	31	17.7	75	42.9	42	24.0	27	15.4
Written tests or test sections	97	55.4	67	38.3	11	6.3	0	0.0
Teacher observation	163	93.1	12	6.9	0	0.0	0	0.0

Source: Own elaboration.

Table 8. Feedback means and preferred feedback means

	Teacher feedback, oral		Tests		Self-assessment		Peer assessment		Feedback by other school adults		School report		Monthly report or comparable		Teacher-parent discussion		Coping with school assignments		Portfolio or comparable	
	F	P	F	P	F	P	F	P	F	P	F	P	F	P	F	P	F	P	F	P
Feedback on English	175	100	175	100	175	100	113	64.6	43	24.6	175	100	0	0.0	0	0.0	175	100	175	100
Feedback on content mastery	175	100	175	100	175	100	113	64.6	43	24.6	175	100	0	0.0	0	0.0	175	100	175	100
Preferred feedback	175	100	0	0.0	172	98.3	79	45.1	0	0.0	0	0.0	0	0.0	0	0.0	-	-	-	-

Source: Own elaboration.

As illustrated in Table 9, the overwhelming majority of the students did not report any serious problems regarding the language of the content, content, cognitive load, and instructions of the texts and

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

tasks. Finally, the students' overall evaluation for CLIL-A was positive and they all found it very effective.

Table 9. Problems encountered

	Always or very often		Often		Sometimes		Seldom	
	F	P	F	P	F	P	F	P
The language of the content was too difficult	0	0.0	39	22.3	47	26.9	89	50.9
The instructions were too difficult	0	0.0	14	8.0	33	18.9	128	73.1
The content was too difficult	0	0.0	9	5.1	16	9.1	150	85.7
The cognitive level of the tasks was too difficult	0	0.0	20	11.4	46	26.3	109	62.3

Source: Own elaboration.

Student Interviews

Perception of CB-A

The students considered CLIL-A a 360-degree assessment, all-encompassing. Covering both content and language and all the activities in and out of the educational context made it a useful tool for them. However, seven students indicated that assessing content in a language program seemed bizarre for them, since their academic disciplines assumed the main responsibility to assess their academic content knowledge.

Effectiveness of CLIL-A

The overwhelming majority (96%) of the students reported that they felt they were being assessed effectively trough CLIL-A. They found it comprehensive enough to cover all their English-related activities, including what they did in their free time. The students all considered CLIL-A as a means of both performance awareness and self-evaluation. On the other hand, 4% of the students had doubts about the use of alternative assessment.

Validity

All the students pointed out that CLIL-A reflected what was covered in the program. The tasks and the content through which they were presented were familiar. Furthermore, 88% of the students indicated that the inclusion of the activities they performed in English in their free time was fair. On the other hand, 16% of the students raised their concerns about the objectivity of alternative assessment.

The effect of CLIL-A on EFL learning

They also added that it affected their language learning activities positively. The assessment of the academic content made them pay attention to the content covered in the materials. Furthermore, 80% of the students stated that they felt motivated to get engaged in activities in English in their free-time, knowing that they were being evaluated.

Challenges

All the students said that CLIL-A was quite new and challenging for them. However, they all found it quite useful, since it attempted to test content and language related to their academic disciplines. Similarly, all the students reported that alternative assessment was another new practice, and they were uncertain about what they were required to do at the very early stages. Hence, they all reported that it took some time to get used to CLIL-A.

Comparison of the Language Component of the CLIL-A and OPT

To see how well the English Language section of the CLIL-A assessed the English Level of the students, the relationship between the cumulative results they obtained from the CLIL-A language component and the scores they got on the OPT was assessed. Table 10 presents the relationship between the CLIL-A language component and OPT.

Table 10 indicates that there was a strong positive correlation between the two variables (r= .990, n = 43, p= .000).

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

Table 10. The relationship between OPT and CLIL-A

		OPT	CLIL-A
OPT	Pearson correlation	1	.990*
	Sig. (2-tailed)		.000
	N	43	43
CLIL-A	Pearson correlation	.990*	1
	Sig. (2-tailed)	.000	
	N	43	43

*Correlation is significant at the 0.01 level (2-tailed).

Source: Own elaboration.

Content Assessment Scores

The content presented in the CLIL materials was also tested. Table 11 depicts the data obtained from the content assessment.

Table 11. The scores of the students obtained on the exams

	N	Mid-term 1		Final 1		Mid-term 2		Final 1		Average	
		Mean*	%	Mean*	%	Mean*	%	Mean*	%	Mean*	%
Faculty of Health Sciences	76	21.30	98	23.30	100	26.30	100	28.30	100	24.5	99.5
Faculty of Arts and Sciences	15	24.30	100	23.30	100	20.30	97	21.30	96	22	98.3
Faculty of Pharmacy	16	21.30	98	22.30	100	24.30	98	24.30	100	22.8	99
Faculty of Economics, Administrative and Social Sciences	39	27.30	100	25.30	100	26.30	100	27.30	100	26.3	100
Faculty of Fine Arts, Design and Architecture	29	23.100	100	21.30	97	19.30	94	27(30	100	22.5	97.8

* Over 30 points.

Source: Own elaboration.

Considering the average content scores and percentages, the students from all the faculties performed well on the exams assessing content knowledge.

Discussion

Overall, the results indicated that the attitudes of both the students and the teachers were positive towards CLIL-A. It was also observed that CLIL-A was an effective tool to assess content and language. More specifically, the evaluation of the CLIL program manifests positive evidence for the first research question on the attitudes of the EFL teachers towards CLIL-A. Given that none of the teachers had any training or experience in the CLIL-A, they found it quite challenging. They were uncertain and doubtful when they first started practicing CLIL-A, but they gradually got adapted to it and got increasingly more apt and secure as they kept practicing it. They effectively managed to assess both content and language to gauge if the program goals were met regarding language and content. This finding is in the same line with those of Leal (2016), Peña (2017), Serragiotto (2007), and Zafiri and Zouganeli (2017), who also indicated that teachers managed to assess both language and content in CLIL programs. However, this finding contradicts that of Massler et al. (2014) and Zhetpisbayeva, et al. (2018), who reported that teachers failed to maintain the balance between content and language.

The teachers primarily emphasized content knowledge, content-specific genre, and students' self-expressions effectively using content-related vocabulary in their oral and written production. In other words, they underlined the basic required factors to carry out tasks in their academic discipline. They also gave importance to linguistic accuracy and complexity, but not as much as the other factors stated above. In other words, concerning the language component of the program, they tended to emphasize vocabulary more, considering it essential to go over basic academic-specific terminology to cover academic content. They prioritized the comprehension of the content provided in the texts by their students. They also favored fluency over

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

accuracy. These findings conflict with those of Barrios and Milla-Lara (2018), and Serragiotto (2007).

To obtain accurate and enough data about the language growth and content mastery of their students, the teachers favored utilizing a wide range of means, as indicated by Barrios and Milla-Lara (2018), and O'Dwyer and de Boer (2015). In terms of informing the students about their language development and content learning, the teachers provided continuous feedback to their students by using any means of summative and formative assessment. This finding did not support the finding of Wewer (2014), and Zafiri and Zouganeli (2017), who pointed out the problem in providing continuous and systematic feedback to students. Regarding the difficulties encountered, the teachers experienced difficulty in adapting the language and cognitive difficulty of the tasks to the current level of the students. Combining content and language and balancing their weight was another type of problem they had to overcome. Finally, as they were not the experts in the academic discipline for which they were planning and developing assessment tasks, they were functioning in unfamiliar territory and they were unsure of the thematic focus. As they were not native speakers of English, they felt the need to have their tasks proofread by native speakers who also taught English in their department. Thus, they were in continuous need of consulting the subject area experts (lecturers) and native speakers. When self-assessing their qualifications for designing, developing, implementing, and evaluating CLIL-A, excluding alternative assessment, they felt themselves well-qualified in CLIL-A at the end of the academic year and did not report any need for training in assessment, which contradicts the finding of Reierstam (2015) and Xavier (2016). Since alternative assessment is open-ended and comparatively difficult to devise a fair and fixed assessment scheme, a few teachers wanted training in alternative assessment.

Almost all the teachers indicated that the CLIL-A component of the CLIL program was effective in assessing both English and content knowledge of the students, which contradicts the findings of Serragiotto (2007), and Zafiri and Zouganeli (2017). They also believed that their students evaluated CLIL-A positively and found it effective to elicit their performance in English and gauge their content learning. Nevertheless, since the students also felt the novelty effect and went

through an orientation process to understand what CLIL-A was and what they were required to do, the teachers needed to be patient and orient their students accordingly. Furthermore, another problem the teachers experienced had to do with the attitudes of some students towards the CLIL-A component of the language program. They did not realize the role of alternative assessment and gave more validation to pen-and-paper exams. Similarly, they did not take the alternative assessment seriously. Hence, throughout the first semester, the teachers had to try to explain the assessment system to these students and keep their motivation and attendance high.

The teachers thought that they made enormous progress in applying CLIL-A. In general, they were quite positive about the CLIL-A both in terms of their assessment literacy skills and its effectiveness to assess the bifocal goal of the CLIL program.

The findings obtained from the student questionnaire provided a positive answer to the second research question. It showed that the students found the CLIL-A practice very efficacious in assessing both English development and content learning. They were satisfied with the wide range of techniques employed to collect as much data as possible about their performance and felt that they were assessed fairly. However, some students were doubtful about the use of alternative assessment. Concerning the techniques used in CLIL-A, they did not mention any serious problems; however, they felt more comfortable with the tasks requiring written production and alternative assessment. When the tasks required oral production, they felt less comfortable. They also found self-evaluation safer than peer-evaluation. In the same vein, they were more satisfied with their performance in these techniques. The students believed that correct pronunciation, knowledge of vocabulary and content, and clarity of expression were important to express their ideas. They assigned less importance to grammatical accuracy in comparison to other linguistic elements in fulfilling tasks in English.

The students believed that receiving feedback on their English and content mastery was important. They felt that they were informed enough about their performance in the CLIL program, which supports the findings of O'Dwyer and de Boer (2015), but contradicts those of Wewer (2014), and Zafiri and Zouganeli (2017). They appreciated any

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

means informing them about how well they were doing in the CLIL program; nevertheless, they preferred obtaining feedback from their teacher or via self-assessment, school report, assignments, and alternative assessment. They wanted to get insights into evaluation from their friends and other adults from the university, but they did not want their parents to get involved in the assessment process.

In terms of the problems they encountered in CLIL-A, they did not experience any major problems impeding their performance on the CLIL-A tasks. The language and content of the CLIL tasks did not create any serious problems for the students. The cognitive difficulty of the tasks did not prevent the students from reflecting on both their language and content knowledge. Nevertheless, about half of the students thought that the language of the content and the cognitive level of the tasks created a minor challenge for them, which confirms the finding of Lo and Fung (2018). The overall attitude of the student for CLIL-A was quite positive, and they found it very effective to reflect their performance in the CLIL-A program.

The results provided positive evidence for the third research question investigating the effectiveness of CLIL-A on assessing the English development of the students. Both the teachers and the students agreed that CLIL-A also assessed English language development satisfactorily. Moreover, the high correlation between the cumulative CLIL-A language scores and OPT may be interpreted as evidence about the assessment power of CLIL-A. This finding may not be considered in line with that of Reierstam (2015), who found no difference between CLIL and non-CLIL language assessment practice, and Barrios and Milla-Lara (2018), and Zhetpisbayeva et al. (2018), who indicated that teachers could not manage the balance between language and content. Likewise, the results provided a definitive answer for the fourth research question aiming to provide information about how well CLIL-A assessed the content learning. CLIL-A was found to be useful in terms of the system it suggested and its usefulness to assess content, which contradicts the findings of Massler et al. (2014), Wewer (2014), Zafiri and Zouganeli (2017) and Zhetpisbayeva et al. (2018) who reported the lack of systematic assessment of content. In short, the scores the students obtained on the CLIL-A and the impressions of the teachers and the students revealed that CLIL-A was an effective means to

assess what the students had mastered about academic content and language.

To sum up, the study is likely to offer solutions to the problems in CLIL-A pointed out by Barrios and Milla-Lara (2018), Lo and Fung (2018), Massler et al. (2014), Reierstam (2015), Serragiotto (2007), Wewer (2014), Zafiri and Zouganeli (2017), and Zhetpisbayeva et al. (2018). It offers a balanced system to assess both language and content, fosters active involvement of the students in the assessment process, and handles assessment as an ongoing, all-encompassing process embracing all the activities students perform in English in and out of the education context.

Conclusion

As the findings showed, the assessment component of the CLIL program was perceived positively and found to be effective in assessing content and language. Although it was the first time such an assessment component had been put into practice, partaking in it was highly valued. The scope of the CLIL-A was found to be adequate to address both content and language assessment effectively. The way the CLIL-A was planned, designed, implemented, and evaluated resulted in a fair and valid assessment. The variety of techniques used, including alternative assessment, provided a wider perspective to count in whatever the students did in and outside the classroom. Therefore, CLIL-A, practiced in an ongoing fashion, also functioned as a holistic means of gathering data about student involvement in content and language learning. Thus, it can be concluded that CLIL-A was efficient to assess both their English and content knowledge simultaneously, thereby deeming it a valuable tool for the assessment for learning.

After receiving training and implementing such a specific assessment type for the first time, the teachers also thought CLIL-A aided their professional development. They managed to assess not only language, but also content, by employing alternative means of assessment in addition to traditional ones. In other words, teachers developed a new and wider perspective of assessment by conceptualizing assessment

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

as an ongoing process and employing alternative means to assess language. This study suggests that CLIL-A is a vigorous tool for the integrated assessment of language and content. Since the topics and the tasks addressed were chosen from among the academic disciplines of the students, the stakeholders found it relevant. Moreover, the project was a success. First of all, it was designed and implemented successfully. The CLIL-A was developed and used by the academic needs of the students and they were highly appreciated by all the stakeholders. Besides, the teacher training also worked very well. Teachers who were trained to perform CLIL-A carried out these tasks effectively. In the same vein, the resemblance between the teacher and student answers on the similar or same items revealed that the orientation program for the students worked well and that they were informed enough of the value and requirements of the CLIL-A. In short, it was proven that CLIL-A which was customized exclusively for this particular CLIL program, was designed and implemented effectively.

Finally, it was the first time, in Turkey, that such a CLIL-A program was implemented institution-wide. The program was also one of a kind in nature regarding its design, implementation, and assessment. The 50% weight of alternative assessment in determining the overall performance of the learners and allotting 30% of the CLIL-based exam directly to content were quite new. In other words, university-wide official recognition of alternative assessment as a means to evaluate the performance of the students and the inclusion of content in assessment were revolutionary in EFL education in Turkey. Furthermore, the CLIL program was designed by the university staff with no outsourcing. The way they received training before and during the program was exclusive for the program and it worked well. Also, to the knowledge of the researchers, it was the first time in Turkey that EFL teachers developed their content-based assessment tailored for each academic program university-wide. Such a variety of assessment tasks catered to different academic disciplines was exceptionally successful. In short, the CLIL-A component was proven to be quite effective. It is hoped that it will pioneer similar programs in Turkey and other EFL contexts.

Implications

CLIL-A is likely to provide data-driven information to all stakeholders about how well students perform tasks requiring both language and content knowledge. Once progress in both language and content is made transparent to the students, they are likely to perform better, deeming the program evaluation more accessible. Besides, it needs to present evidence to check if students possess the required knowledge and skills for their subsequent education and career. Hence, it needs to be regarded beyond its traditional role as a grading tool.

Keeping in mind that assessment is also a means of learning and not a procedure ensuing teaching to test the quality of the end-product, it is suggested that a supportive and facilitative context be created during the assessment process to incite student effort and instigate learning. The study especially implied that alternative assessment is a powerful tool where learners are allowed to use language for genuine communicative purposes. It provides a context to use language purposefully to develop their communicative competence by engaging in experiential learning. Moreover, during the assessment process, the students have enough opportunity to self-evaluate and gain awareness of their learning to make informed decisions about their progress and revise their language learning strategies. Another consideration is familiarity. To increase validity and reduce the intervening role of content in assessment, the language assessment tasks need to be contextualized and presented in the subject matter students have already studied and are acquainted with. Having enough background knowledge about the content in familiar tasks and contexts decreases the demand for content knowledge and makes students mainly deal with the linguistic aspect of a task.

Employing a multifaceted assessment approach to offer a solution for the aforementioned problems in CLIL-A CLIL (Barrios & Milla-Lara, 2018; Otto, 2017; Tedick & Cammarata, 2006) is another implication of the study. In that sense, alternative assessment can be employed to complement the traditional assessment practices, as it favors ongoing assessment, encompasses what students do inside and outside of the classroom, and actively involves the stakeholders in the education process (Short, 1993). Alternative assessment is likely to provide

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

diversified, contextualized and meaningful task-based holistic assessment, which covers both product and process (Linfield & Posavac, 2018; Purpura, 2016; Tsagari, 2016).

Moreover, teacher training is a must for CLIL-A. Teachers need to be aware of the content and language needs of students and design assessment accordingly. In addition to what needs to be assessed, they need to be resourceful enough about how to assess and use appropriate techniques to observe the most typical and actual performance of their students. Teachers need to design tasks that are thematically, cognitively and linguistically appropriate to their current levels. Similarly, as students are the active participants of the assessment process, they need to be informed and oriented about the assessment process to make them believe in the value of the process and take it seriously.

For the limitations of the study, the lack of data obtained from the same participants about a general EFL assessment component could be pointed out. The limitations were likely to yield a comparison between the perceptions of the same group of students on two different types of assessment. However, this particular CLIL-A was not designed as research but a real practice. Therefore, it was impossible to implement a general assessment. Moreover, it would have been better to pilot the CLIL-A component before it was put into practice; however, there was no chance for piloting, as the CLIL-A had to be implemented full scale right away. Another limitation has to do with the student's attitude for alternative assessment. The fact that they would be graded via alternative assessment was quite a new experience for some students and they did not completely comprehend how it would be practiced and evaluated. Moreover, some paid lip service to alternative assessment and did the tasks for the sake of doing them. Hence, the data on alternative assessment might not present the whole perspective.

For further research, it is suggested to gather data from two groups of students, one assessed through CLIL-A and one through a general assessment to elicit the assessment-related perceptions of both students and teachers. Moreover, before launching such a completely new project and study, learners need to receive a thorough orientation.

References

AECLIL. (2013). *Assessment and evaluation in CLIL*. https://aeclil.altervista.org/Sito/wp-content/uploads/2013/02/AECLIL-Assessment-and-evaluation-in-CLIL.pdf

Barbero, T. (2012). Assessment tools and practices in CLIL. In F. Quartapelle (Ed.), *Assessment and Evaluation in CLIL* (pp. 38–56). European Commission: AECLIL-EACEA. https://aeclil.altervista.org/Sito/wp-content/uploads/2013/02/AECLIL-Assessment-and-evaluation-in-CLIL.pdf

Barrios, E., & Milla-Lara, M. D. (2018). CLIL methodology, materials and resources, and assessment in a monolingual context: An analysis of stakeholders' perceptions in Andalusia. *The Language Learning Journal*, *48*(1), 60–80. https://doi.org/10.1080/09571736.2018.1544269

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

Douglas, D. (2010). *Understanding language testing*. Hoddler Education.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385–402. https://doi.org/10.1177/0265532208090158

Isik, A. (2003). Is ESP a success in the Turkish EFL context? *ITL Review of Applied Linguistics, 141*(1), 255–287. https://doi.org/10.2143/ITL.141.0.2003190

Isik, A. (2008). Linguistic imperialism and foreign language teaching. *The Journal of Asia TEFL*, 5(1), 123–144. https://www.researchgate.net/profile/Ali_Isik6/publication/323958594_Linguistic_Imperialism_and_Foreign_Language_Teaching/links/5ab4afecaca27217100444e1/Linguistic-Imperialism-and-Foreign-Language-Teaching.pdf

Leal, J. P. (2016). Assessment in CLIL: Test development at content and language for teaching natural science in English as a foreign language. *Latin American Journal of Content and Language Integrated Learning, 9*(2), 293–317. http://dx.doi.org/10.5294/laclil.2016.9.2.3

LICI. (2009). *Language in content instruction*. http://lici.utu.fi.

Linfield, K. J., & Posavac, E. J. (2018). *Program evaluation: Methods and case studies*. Routledge.

Lo, Y. Y., & Fung, D. (2018). Assessments in CLIL: The interplay between cognitive and linguistic demands and their progression in secondary

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

education. *International Journal of Bilingual Education and Bilingualism*, 1–19. https://doi.org/10.1080/13670050.2018.1436519

Maggi, F. (2012). Evaluation in CLIL. In F. Quartapelle (Ed.), *Assessment and Evaluation in CLIL* (pp. 57–74). http://aeclil.altervista.org/Sito/wp-content/uploads/2013/02/AECLIL-Assessment-and-evaluation-in-CLIL.pdf

Massler, U. (2010). Assessment in CLIL learning. In S. Ioannou-Georgiou & P. Pavlou (Eds.), *Guidelines for CLIL implementation in primary and pre-primary education* (pp. 114–136). Cyprus Pedagogical Institute.

Massler, U., Stotz, D., & Queisser, C. (2014). Assessment instruments for primary CLIL: The conceptualisation and evaluation of test tasks. *The Language Learning Journal*, *42*(2), 137–150. https://doi.org/10.1080/09571736.2014.891371

Morgan, C. (2006). Appropriate language assessment in content and language integrated learning. *Language Learning Journal*, *33*(1), 59–67. https://doi.org/10.1080/09571730685200121

O'Dwyer, F., & De Boer, M. (2015). Approaches to assessment in CLIL classrooms: Two case studies. *Language Learning in Higher Education*, *5*(2), 397–421. https://doi.org/10.1515/cercles-2015-0019

Otto, A. (2017). Assessment issues in CLIL: What you've been wondering but were afraid to ask. *Revista de Investigación*, (1), 1–13. https://app.rdim.es/archivos/publicacion/303dce018b15282742f1b7545667f026.pdf

Otto, A. (2018). Assessing language in CLIL: A review of the literature towards a functional model. *LACLIL, 11*(2), 308–325. http://dx.doi.org/10.5294/laclil.2018.11.2.6

Peña, I. P. (2017). *Assessment for learning in primary CLIL classrooms and its co-construction in classroom discourse* (Doctoral dissertation). Universidad Autónoma de Madrid, Madrid, Spain. https://pdfs.semanticscholar.org/c0b9/660532606c95161d635d2163b22824eb2ea4.pdf

Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, *100*, 191–208. https://doi.org/10.1111/modl.12308

Reierstam, H. (2015). *Assessing language or content? A comparative study of the assessment practices in three Swedish upper secondary CLIL schools* (Licentiate thesis). University of Gothenburg, Gothenburg, Sweden. https://gupea.ub.gu.se/bitstream/2077/40701/4/gupea_2077_40701_4.pdf

Serragiotto, G. (2007). Assessment and evaluation in CLIL. In D. Marsh & D. Wolff (Eds). *Diverse contexts – Converging goals. CLIL in Europe* (pp. 271–283). Peter Lang.

Short, D. J. (1993). Assessing integrated language and content instruction. *TESOL Quarterly, 27*(4), 627–656. https://doi.org/10.2307/3587399

Tedick, D., & Cammarata, L. (2006). Integrated performance assessment: Adapting the model for CBI. *Minneapolis: Center for Advanced Research on Language Acquisition (CARLA)*. https://www.carla.umn.edu/cobaltt/modules/assessment/ipa/index.html

Tsagari, D. (Ed.). (2016). *Classroom-based Assessment in L2 Contexts*. Cambridge Scholars Publishing.

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, *6*(1), 41–63. https://arts.unimelb.edu.au/__data/assets/pdf_file/0010/2349928/6_1_SI3TsagariVogt.pdf

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374–402. https://doi.org/10.1080/15434303.2014.960046

Wewer, T. (2014). *Assessment of young learners' English proficiency in bilingual content instruction CLIL* (Doctoral dissertation). The University of Turku, Turku, Finland.

Xavier, A. (2016). *Assessment for learning in EBE/CLIL: A learning-oriented approach to assessing English language skills and curriculum content at early primary level* (Doctoral dissertation). Nova University Lisbon, Lisbon, Portugal. https://run.unl.pt/bitstream/10362/17973/1/ProjectWork_ELT_AX_OE_FINAL_28_10_2015.pdf

Zafiri, M., & Zouganeli, K. (2017). Toward an understanding of Content and Language Integrated Learning Assessment (CLILA) in primary school classes: A case study. *Research Papers in Language Teaching and Learning*, *8*(1), 88–109. https://rpltl.eap.gr/images/2017/08-01-088-Zafiri-Zouganeli.pdf

Zhetpisbayeva, B. A., Kitibayeva, A. K, Kazimova, D. A., Akbayeva, G. N., & Zatyneiko, M. A. (2018). Assessment issues in Content and Language Integrated Learning (CLIL). *Journal of Advanced Pharmacy Education & Research, 8*(4), 32–38. https://japer.in/storage/models/article/JOBbrVa32ILzNvUY8Vj3Ysq9Op4YlJdySeGkY2nS9B04dAgB6Nxw5kO3QsfQ/assessment-issues-in-content-and-language-integrated-learning-clil.pdf

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

Appendices

Appendix 1. Assessment Guidelines

Compatibility with the Europass Language Passport
Consider the three pillars of assessment, namely, content, language, and cognitive processes

1. **Overall Grade Percentages:** Alternative Assessment: 50% CBI-based assessment: 50%

2. **CBI-based assessment task weight:**
   Content (reading: 40% writing: 10% content knowledge: 30%)
   Language (grammar: 10% vocabulary: 10%)

3. **Steps to be followed**
   a. General aim, subgoals, unit objectives
   b. List of vocab & structures
   c. Table of specification
   d. Preparing the first exam draft
   e. Piloting
   f. Revision
   g. Final draft

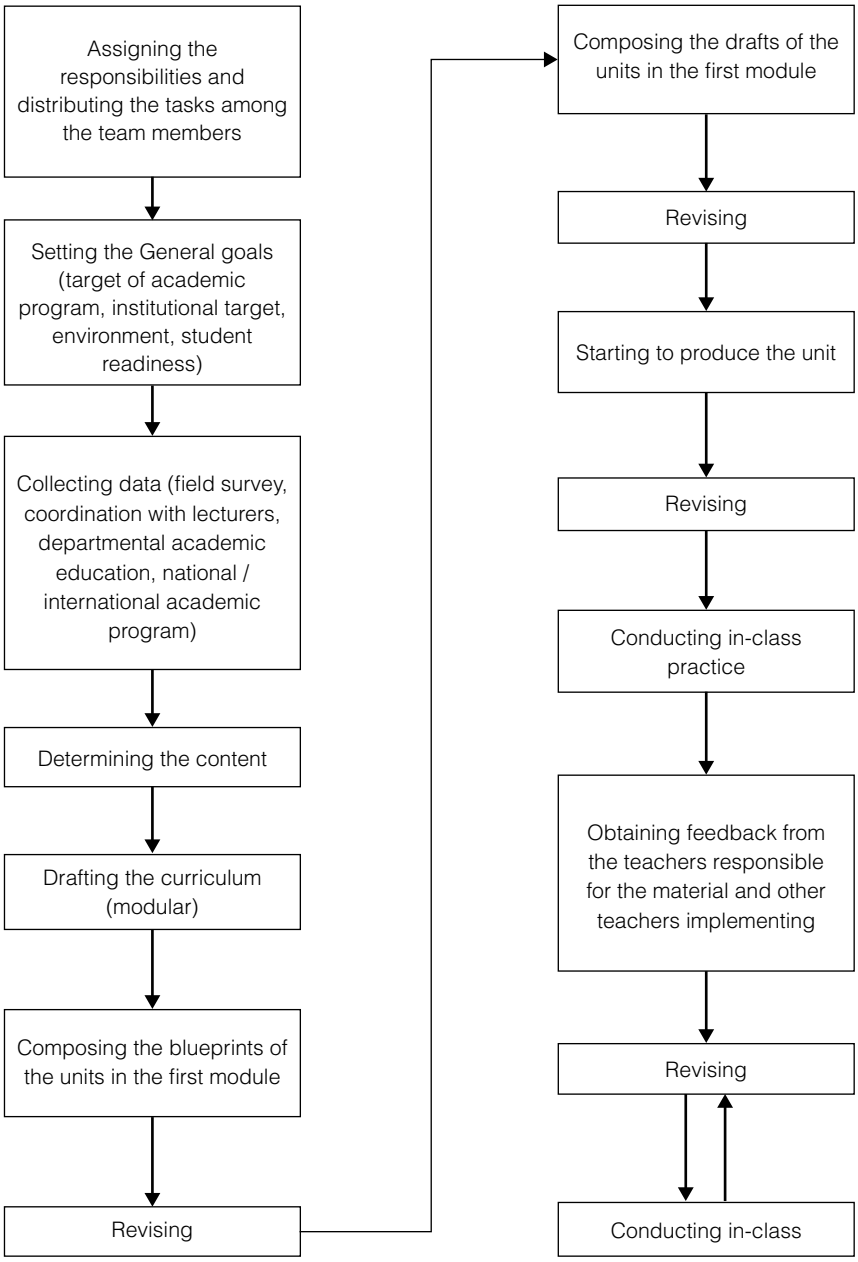
4. **Validity:**
   a. Topics are to be similar to the ones covered in the class.
   b. Question types are to be the same as the ones in the materials
   c. The weight of the topics and question types in the materials are to be reflected directly on the exam.

5. **General Guidelines**
   a. One to one correspondence between what is taught and tested
   b. Familiarize students with the test (practice effect). Do not trick students
   c. A holistic and integrated approach
   d. Right cognitive and linguistic level
   e. Both process- and product-oriented

f. Student-centered (Students should also be responsible for evaluating their performance)
g. Test what they know not what they do not know
h. Questions for all students at every level (questions everybody can answer, and the ones only the high achievers can answer)
i. Time spent on specific skills, techniques, and topics in class determines the weight of test items.
j. Informative cover (school, class, units, date, time, name, score box. etc.)
k. Practicality
l. Face validity
m. Clear instructions
n. Give enough time
o. One task cannot be prerequisite for another
p. One correct answer
q. Other items, texts or questions cannot suggest the answer to a question
r. Avoid, always, never, only, etc. in the stem or options
s. The difficulty should be in the question (task) itself, not in the form.
t. Do not test world knowledge (They mus not answer the questions without reading or listening to text)
u. Avoid bias (gender, religion, etc.)
v. Do not split the items, texts, and questions onto another page

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

Appendix 2. Assessment Process



Appendix 3. Student Questionnaire

You are asked to fill this questionnaire about content-based assessment so that we can know your opinion on the CLIL experience carried out this year. Indicate your responses with a "√". Thank you for your cooperation.

1. **Does CBI assessment help you check your language ability to express yourself in the foreign language?**
   ☐ A lot ☐ Enough ☐ A little ☐ None

2. **Does CBI assessment help you evaluate your learning of the subject studied in the foreign language?**
   ☐ A lot ☐ Enough ☐ A little ☐ None

3. **How safe/comfortable do you feel in the following situations:**

	Very comfortable	Comfortable	Partially comfortable	Not comfortable
a. oral interchange				
b. presentations				
c. pupil peer assessment				
d. pupils' self-assessment or reflection				
e. simulations				
f. written tests or test sections				
g. essays or compositions				
h. others:				

4. **What do you consider important when you perform a task in a foreign language?**

	Very important	Important	Partially important	Not important
a. the correct pronunciation of words				
b. grammatical correctness				

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

	Very important	Important	Partially important	Not important
c. knowledge of vocabulary				
d. knowledge of the contents				
e. clarity of expression				
f. others:				

5. Which techniques are more helpful?

	Very important	Important	Partially important	Not important
a. language portfolio				
b. observations of another person				
c. pupil peer assessment				
d. pupils' self-assessment or reflection				
e. simulations				
f. essays or compositions				
g. presentations (e.g. feed back forms)				
h. dialogues, interaction				
i. oral tests or interviews				
j. written tests or test sections				
k. teacher observation				

6. How important is it to you to receive information on your English skills in CLIL subjects?

- Very important
- Important
- Partially important
- Not important

7. **How important is it to you to receive information on your mastery of CLIL subjects**
   - Very important
   - Important
   - Partially important
   - Not important

8. **In which ways do you receive feedback on your English skills at school in lessons other than English?**
   - Teacher feedback, oral
   - Tests
   - Self-assessment
   - Peer assessment
   - Feedback by other school adults
   - School report
   - Monthly report or comparable
   - Teacher-parent discussion
   - Coping with school assignments
   - Portfolio or comparable
   - Other

9. **In which ways do you received feedback on your content mastery at school in lessons other than English?**
   - Teacher feedback, oral
   - Tests
   - Self-assessment
   - Peer assessment
   - Feedback by other school adults
   - School report
   - Monthly report or comparable
   - Teacher-parent discussion
   - Coping with school assignments
   - Portfolio or comparable
   - Other

10. **Which means of feedback would you prefer?**
    - Teacher feedback, oral
    - Tests
    - Self-assessment
    - Peer assessment
    - Feedback by other school adults
    - School report
    - Monthly report or comparable
    - Teacher-parent discussion
    - Other

11. **What problems do you have?**

	Always or very often	**Often**	**Sometimes**	**Seldom**
a. The language of the content was too difficult.				
b. The instructions were too difficult.				
c. The content was too difficult				
d. The cognitive level of the tasks was too difficult.				

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

12. Do you think content-based is effective?

⛁ Very effective ⛁ Effective ⛁ Partially effective ⛁ Not effective

Appendix 4. Teacher Self-evaluation Questionnaire

You are asked to fill in this questionnaire so that we can know your opinion on the content-based assessment experience carried out this year. Indicate your responses with a "√". Thank you for your cooperation.

1. **Which are your previous experiences in CLIL teaching?**

   ⛁ None ⛁ A few ⛁ Some ⛁ Many

2. **How do you consider your experience of content-based assessment?**

   ⛁ Very effective ⛁ Effective ⛁ Partially effective ⛁ Ineffective

3. **What is included in the basis for assessment?**

   ⛁ Content only ⛁ Language only ⛁ Both language and content
   ⛁ It depends on the type of task

4. **Which factors weigh the most when assessing students' knowledge in your subject?**

	Very important	Important	Partially important	Not important
a. Oral and written skills				
b. Content areas/themes (e.g. knowledge about eras, systems, theories)				
c. Use of content obligatory/subject specific vocabulary				
d. Use of content compatible/general academic vocabulary and fluency				
e. Mastery of various forms of expression and multimodal instruments				

	Very important	Important	Partially important	Not important
f. Linguistic accuracy				
g. Linguistic complexity				
h. Mastery of a disciplinary written genre				
i. Analytical skills				

5. Which techniques are more important?

	Very important	Important	Partially important	Not important
a. language portfolio				
b. observations of another person				
c. pupil peer assessment				
d. pupils' self-assessment or reflection				
e. simulations				
f. essays or compositions				
g. presentations (e.g. feed back forms)				
h. dialogues, interaction				
i. oral tests or interviews				
j. written tests or test sections				
k. teacher observation				

6. Where do you find the most important source of material for the different areas of assessment?

- Coursebook
- Teacher's own classroom lectures
- Activities in the classroom/during the course
- Internet

Other:

7. How often do you give feedback to the pupils on their language proficiency or its progress?

rarely sometimes often always

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287

8. **How often do you give feedback to the pupils on their mastery of the content knowledge?**
   - rarely
   - sometimes
   - often
   - always

9. **Which methods do you use when providing students assessment information on their level of language proficiency?**
   - school year report
   - class discussion or mutual feedback
   - written feedback (e.g. in connection with tests)
   - oral feedback
   - other, which?

10. **Which methods do you use when providing students assessment information on their level of language proficiency?**
    - school year report
    - class discussion or mutual feedback
    - written feedback (e.g. in connection with tests)
    - oral feedback
    - other, which?

11. **What problems do you have?**

	Always or very often	Often	Sometimes	Seldom / never
a. Adapting the language to the students level.				
b. Combining content and language.				
c. Adapting the cognitive level of the tasks to the students level.				
d. Nothing having enough knowledge of the content.				
e. Others:				

12. Which area(s) would you like to know more about in relation to grading and assessment? Interpretation of national knowledge requirements

- Test design
- Oral assessment
- Written assessment
- Alternative assessment modes (e.g. portfolio, self-assessment)
- Alternative assessment tools (e.g. computer based)
- Other

If you responded "other" in the previous question, please elaborate here:

13. I do not feel a need for training regarding grading and assessment

completely agree agree neutral disagree

Appendix 5. Teacher Follow-up Interview Guiding Questions

1. Could you explain what CBI and CBI-A are?
2. Have you ever participated in a CBI program?
3. Has CBI-A helped you assess your students' content knowledge? If yes, please explain.
4. Has CBI-A helped you assess your students' English development? If yes, please explain.
5. How has CBI-A influenced the EFL program?
6. What problems have you experienced with CBI-A?
7. Do you need further training in CBI-A?
8. What else would you lie to say about CBI-A?

Appendix 6. Student Follow-up Interview Guiding Questions

1. What do you think about CBI-A?
2. Has CBI-A been effective to assess your performance?
3. Do you think CBI-A has assessed your performance adequately?
4. How does CBI-A affect your EFL learning?
5. What have been the challenges with CBI-A for you?
6. What else would you lie to say about CBI-A?

LACLIL ISSN: 2011-6721 e-ISSN: 2322-9721 VOL. 13, No. 2, JULY-DECEMBER 2020 DOI: 10.5294/laclil.2020.13.2.5 PP. 241-287