

Connecting the Dots: Impact of CLIL on Vocabulary Size and Content Mastery in Higher Education

Maia Al Hajri, Abdo Al Mekhlafi, Thuwayba Al Barwani

Article Info

Article History

Received:
April 17, 2021

Accepted:
September 24, 2021

Keywords :

Content and Language Integrated Learning (CLIL), English language learning, vocabulary size, content learning, EFL, higher education

DOI:

10.5281/zenodo.5527158

Abstract

This study aimed at investigating the impact of Content and Language Integrated Learning (CLIL) on higher education students' vocabulary size and content mastery. The study involved a sample of 58 students from Middle East College in Omanduring the academic year 2019-2020. The experimental group consisted of 28 students, while the control group consisted of 30 students. The research instruments included a vocabulary size test and a content mastery test. The results of this study revealed that there was no statistically significant difference between the control and the experimental groups' scores in the vocabulary post-test and the delayed-test. However, the vocabulary learning gain of the experimental group was higher than that of the control group, which means that CLIL had a positive impact on students' vocabulary size, but this impact was not statistically significant. Moreover, there were differences between the control and the experimental groups' scores in the content mastery test in favor of the experimental group. However, they were not statistically significant.

Introduction

The integration of content and language teaching in education is by no means a contemporary invention. Throughout history, novel languages have been imposed on the colonized and on linguistic minorities within countries. For instance, Latin was originally spoken only in the areas around Rome. However, the growing power of the Roman Republic made Latin the dominant language in Italy and then throughout the Mediterranean basin (Zecca-Naples, 2015). Similarly, immigrant students, who are often members of linguistic minorities, must use a language other than their mother tongue to learn content. In the 1960s, the immersion method was introduced in Canada to teach French. In this context, students were taught school subjects, such as history, music, mathematics, the arts, and the sciences, in French. However, many students never achieved native-level proficiency in French (Lambert, 1977). Immersion was then adopted in Europe as a so-called two-in-one approach, which focused on learning a foreign language while simultaneously studying school subjects in that language. This approach was adopted in various European contexts under the umbrella term content and language integrated learning (CLIL) (Coyle, 2006; Dalton-Puffer, 2008).

Generally, Arabic-speaking learners face many challenges in learning English. A central issue is lack of vocabulary. Vocabulary is an essential language sub-skill for the four main skills of reading, writing, listening, and speaking, and it is an extremely important aspect of second language knowledge (Willis & Ohashi, 2012). However, many Arabic-speaking learners lack the necessary vocabulary to express their meaning, which leads to feelings of insecurity and an inability to continue with interactions as far as necessary (Hussein & Eltayef, 2003; Hurtado, 2002). Yet, for many decades, vocabulary learning was not given the attention that it deserved in foreign language teaching. However, it is no longer the Cinderella skill of the foreign language classroom, as it has come to be considered a fundamental sub-skill for reading, writing, listening, and speaking (Willis & Ohashi, 2012). Lexicon plays a prominent role in linguistics, and in line with Wilkins's (1972) statement, "Without grammar, very little can be conveyed; without vocabulary, nothing can be conveyed" (p. 111-112) (as cited in Ismail, Zaid, Mohamed, & Rouyan, 2017), vocabulary now receives greater emphasis in foreign language teaching (Hurtado, 2002).

The English language is assigned high importance in the Sultanate of Oman, due to its necessity and importance in the study of science and technology, travel to non-Arabic speaking countries, understanding a range of cultures, conducting business, and finding white-collar jobs (Al-Issa, 2006). In Oman, English is a core subject in private and public schools and in colleges and universities. However, a vast majority of students who finish twelfth grade and enter higher education cannot communicate appropriately and effectively in English. The

newly developed Omani school curriculum remains ineffective for producing adequate levels of English among students (Al-Issa & Al-Bulushi, 2012). This means that Omani students are likely to continue to leave school and enter higher education with low proficiency in English.

Al-Hosni (2014) found that Omani students struggle to find appropriate vocabulary items when they speak in English, indicating an insufficient vocabulary reservoir. Al-Seyabi and Tuzlukova (2014) found that vocabulary is a significant challenge to the development of students' writing skills. Many students have difficulty choosing adequate vocabulary to express their thoughts, and conversely, Al-Mahrooqi (2012) found that many students do not concentrate on understanding language: instead, they simply memorize vocabulary. However, all of the participants in that study nevertheless reported that vocabulary is still their main hindrance to reading comprehension.

CLIL, a term that was adopted by the European Network of Administrators, Research and Practitioners in the mid-1990s (Coyle, 2006), is a commonly used approach in Europe. Many studies have found that CLIL effectively promotes English skills (Agrelo, 2010; Delliou&Zafiri, 2016; Pérez Cañado, 2018). CLIL guides students to learn the target language in the same way that native speakers do, resulting in improved language knowledge and communication skills (Klimova, 2012). When compared to non-CLIL students two or three years ahead of them, CLIL students were found to have similar or higher scores on language tests (Lasagabaster, 2008; Navés, 2011). Encouraged by these findings, the present study was designed to investigate the impact of CLIL method on vocabulary size and content mastery of Omani college students.

Literature Review

English Vocabulary and Vocabulary Size

The term 'vocabulary' has received a range of definitions. Alizadeh (2016) described it as the knowledge a language learner has of words in the target language and their meanings. Alqahtani(2015) defined it as the total number of words in a language and their meanings. However, TnanhHuyen and Thi Thu Nga (2003) characterized vocabulary as a language element that links the four language skills (as cited in Asyiah, 2017). Yet, the meaning of words is not the only knowledge that learners acquire in relation to vocabulary. Nation (2001) divided word knowledge into nine aspects, arranged in three levels: the form, meaning, and use of a word. For the form level, learners should consider the word's written form, spoken form (pronunciation), and elements. Nation explains that for the meaning level, attention should center on the form and meaning of a word, its concepts and reference, and its associations. For the use level, the word's grammatical function, its common collocations, and constraints on its use should be understood.

An essential distinction that should be considered in relation to word knowledge is that between the size of one's vocabulary, also called its breadth, and its depth. Vocabulary size refers to the amount of vocabulary a person knows in a language, and depth reflects the amount the learner knows about these words. Thus, vocabulary size is a quantitative measure of word knowledge, and depth is a qualitative measure (Ma, 2009). Milton (2009) indicated that breadth and depth are equally important for a learner's vocabulary to allow comprehension of a text. However, Laufer, Batia, and Paul (2012) showed that breadth of vocabulary is more important than in-depth knowledge of a few words (as cited in Larsson, 2014).

The number of words in English language is continually changing, as new words are being added, old ones are falling into disuse, and new uses for existing words are being developed (Nation & Waring, 1997). It is also very difficult to decide whether compound words, archaic words, proper names, abbreviations, dialect words, and alternative spellings are to be included in such a count. A good way of counting the words in a language might be to count word families in terms of their base words, or "lemmas", as they are called. A lemma is a set of lexical forms that share a stem and belong to the same word class (JanebiEnayat, et al., 2018). Dupay (1974) and Goulden, Nation and Read (1990) investigated the vocabulary represented in Webster's Third International Dictionary (1963) and found that it contained around 54,000 word families, excluding compound words, archaic words, proper names, abbreviations, dialect words, and alternative spellings (as cited in Nation and Waring, 1997).

According to D'Anna, Zechmeister, and Hall (1991), there is also little agreement regarding the size of the vocabulary that an undergraduate student in the United States might have. Accordingly, a study was conducted to measure the number of word families a language user knows. Findings indicated that native speakers can be expected to add 1000 word families per year until maturity, which means that a 5-year-old native speaker knows around 4000 to 5000 word families. Correspondingly, a university graduate would have a vocabulary of around 20,000 word families. However, Nation and Waring (1997) argue that measured vocabulary size of adult non-native speakers is not even close to that of natives, as many adult English learners know much fewer than 5000 word families, even though they have been studying English for many years.

Content and Language Integrated Learning

Definition of CLIL

Content and Language Integrated Learning (CLIL) is an umbrella term, first adopted in the mid-1990s by the European Network of Administrators, Research and Practitioners (Coyle, 2006). It is a dual-focused method, in which a foreign language is used as a means for teaching and learning both a subject content and the language

itself (Coyle, 2006; Marsh, 2013; Education Information Network in the European Union, 2006). Paliwoda-Pękosz and Stal (2015) called CLIL a dynamically developing instructional methodology, given impetus by its dual emphasis on transferring knowledge and enhancing foreign language skills. Here, language is a goal of teaching and learning, but it is not given precedence over other goals (Dalton-Puffer, 2008). Dalton-Puffer (2008) and Coyle (2006) characterized CLIL's goals as follows:

1. Developing intercultural communication skills,
2. Preparing learners for internationalization,
3. Conveying values of tolerance and respect to other cultures,
4. Providing opportunities for studying content through different perspectives,
5. Providing access to subject-related terminology,
6. Improving overall target language competence,
7. Enhancing learners' oral communication skills,
8. Diversifying forms of classroom practice, and
9. Increasing learners' motivation.

Moreover, CLIL incorporates the following features: a focus on language, content, and cognition; use of authentic materials; promotion of active learning strategies; scaffolding; and enhancement of learner autonomy (Hammond, 2001, as cited in Chostelidou&Griva, 2014).

The dual focus of CLIL on content and language leads CLIL programs to be divided into two distinct types: hard/strong CLIL and soft/weak CLIL. Ball (2009) described hard/strong CLIL programs as those that teach academic subjects in English by a non-native subject teacher. In such programs, subject teachers focus on teaching the content of their subject, supplemented with some language instruction, with the result that they teach the content of their subject through the foreign language. In addition, in this approach, assessment and feedback focus more on the subject content than on language (as cited in Ikeda, 2013). In soft/weak CLIL programs, language teachers integrate the content of other, non-language, subjects into their foreign language teaching as a means of presenting and practicing the language. In these programs, assessment and feedback focus on foreign language only (Dale & Tanner, 2012).

The distinction between these types should be accompanied by differentiation between similar programs present along the continuum of the integration of content and language that may closely resemble CLIL. According to Cenoz (2015), on the left end of this continuum, beyond soft CLIL, is content-based instruction (CBI), defined as "the concurrent study of language and subject matter, with the form and sequence of language presentation dictated by content material" (Brinton, Snow, & Wesche, 1989, as quoted in Cenoz, 2015). CBI is an integrated approach to language instruction that uses topics, text, and tasks from content or subject classes but focuses on cognitive and academic language skills (Crandall & Tucker, 1990). CBI instructors are language teachers, and they do not focus on content mastery because the content is only provided as a meaningful context for the language (Dale & Tanner, 2012).

On the far right end of the continuum, in immersion, learners learn all or most of their school subjects in a second language without specific instruction on language skills. The language of instruction is not taught explicitly here but is only the medium through which curricular content is conveyed (Barimani, 2011). In this case, subjects are taught by subject teachers, not language teachers (Dale & Tanner, 2012). Students in such programs are simply expected to reach native speaker proficiency, unlike students in CLIL programs, who are taught predetermined and planned language skills and functions (Lasagabaster& Sierra, 2009).

The 4Cs Framework

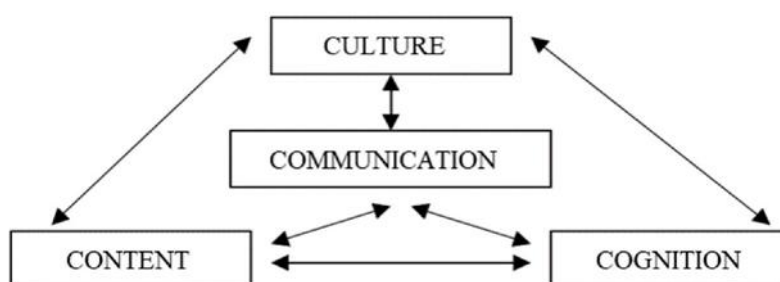
Coyle (1999) asserted that a successful CLIL lesson should combine the elements of the 4Cs curriculum, which consists of four main elements: content, cognition, communication, and culture. The content of CLIL lessons is drawn from content subjects, such as history and science, rather than from everyday life or general content (Sepešiová, 2015). For a CLIL lesson to be successful, it should enable students to progress in their knowledge and skills and develop their understanding of specific elements of a defined curriculum. However, learning subject-related content involves much more than acquiring knowledge and skills, including developing cognitive and thinking skills that enable learners to construct their own knowledge and skills in a way that is relevant and appropriate (Lantolf, 2000; Vygotsky, 1978, as cited in Coyle, 2008).

In addition, in the CLIL classroom, the key to language learning is communication, as in this environment, students are involved in a range of communicative activities in which they use the language as a means of learning the content. In other words, they use the language to learn while they learn to use the language (Coyle, 1999). Mohan (1986) argued that interactions in the learning context are fundamental to learning. In this context, students have the chance to think through instructional materials and make them their own if their instructors provide multiple opportunities for exploratory speech and writing (as cited in Coyle, 1999).

Moreover, CLIL develops students' awareness of other cultures by exposing them to a range of authentic situations and texts. Students are thus exposed to alternative perspectives and new shared understandings, which deepen their social awareness of otherness and of self (Klimova, 2012).

Figure 1

The 4Cs Framework for CLIL (Zydatiβ 2007, 16)



Adapted from “Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe,” by C. Dalton-Puffer, 2008, Na, p. 3. Copyright 2008 by University of Vienna.

According to Coyle (2008), the 4Cs framework is not a theory, but a conceptualization of CLIL, rooted in a philosophical stance that applies to education in general rather than to CLIL in particular. Overall, the 4Cs framework suggests that effective CLIL happens through a progression in knowledge and skills, greater understanding of the content, engagement in related cognitive processing, interaction in a communicative context, development of appropriate language knowledge and skills, and experience of deepening intercultural awareness.

Advantages of Implementing CLIL

Over the course of decades, the literature has addressed many advantages of the CLIL method. According to Coyle (2006), CLIL is effective for improving student proficiency in foreign languages by increasing classroom exposure to a target language. CLIL leads to better-developed metalinguistic competence in learners, which gives them increased insight into understanding their target language and its structures, and this results in a better understanding of abstract concepts through that language (Surmont, Van de Craen, Struys, & Somers, 2014). Most of the studies that have focused on CLIL and language development have found a positive and significant impact of CLIL on developing language competencies (e.g., Fernández-Fontecha, 2014; Pérez Cañado, 2018; Samper, 2015; Xanthou, 2010; Moghadam & Fatemipour, 2014; Sanad & Ahmed, 2017). Furthermore, the test scores of CLIL students are similar to or even higher than those of non-CLIL students who are 2 or 3 years ahead of them in schooling (Lasagabaster, 2008; Navés, 2011).

Delliou and Zafiri (2016) examined the impact of CLIL in developing the speaking skills of sixth-grade students in Greece. Observational assessment and a pretest and posttest for language designed by the Ministry of Education of Greece were administered to assess student performance in speaking skills. The findings of this study indicated that both CLIL and non-CLIL groups improved their oral communication skills, although the CLIL group showed higher levels of performance. However, no improvement was shown in speaking accuracy over the course, as the students were more concerned with completing their tasks successfully. On the posttest, however, the CLIL students showed higher levels of accuracy. Likewise, their oral communication fluency and their use of communication strategies, vocabulary, and cohesion improved remarkably. However, this study was limited by its small sample size, as only 30 students participated.

Studies in Asian contexts have revealed convergent results. Ikeda (2013) studied the impact of CLIL on writing ability in Japanese secondary school students (16- to 17-year-olds). The study began in April 2011 and ended in August 2013, and it involved 80 students (62 females and 18 males). An essay-writing pretest and posttest were administered to measure students’ writing ability. There was a statistically significant difference between students’ results on the pretest and the posttest in favor of the posttest. However, this was due entirely to development in writing fluency (length of writing), while writing accuracy (grammar, usage, and mechanics) did not improve. A teacher who participated in the experiment observed that more students were able to write than before, and each student in general was able to write more. However, the teachers did not correct the students’ errors, following the CLIL principle that language learning should be natural and students should correct their errors. Furthermore, no explicit instructions on how to write or organize paragraphs were given during the course.

The results reported by Ikeda (2013) and Delliou and Zafiri (2016) highlight that language skills and aspects of language learning are not developed equally under CLIL conditions, and some aspects of language knowledge are either unaffected by CLIL or affected in an indefinite way (Dalton-Puffer, 2007). Table 1 shows the learning aspects and language skills that are affected or unaffected by CLIL.

Table 1

Learning Aspects and Language Skills Affected or Unaffected by CLIL (Dalton-Puffer, 2007)

Favorably affected	Unaffected or indefinite effect
Receptive skills	Syntax

Vocabulary Morphology Creativity, risk-taking, fluency, quantity Emotive/affective outcomes	Writing Informal/non-technical language Pronunciation Pragmatics
--	---

Further, Van de Craen et al. (2007) argued that subject matter knowledge is not less developed in CLIL instruction than elsewhere, particularly in primary schools. However, the case here is different than in secondary schools, as the research results vary. Some researchers have argued that knowledge acquisition in a subject is highly affected by many important factors that go beyond instruction as dictated by CLIL, such as age (Paradis, 2004) and inhibition (Bialystok et al., 2005). Van de Craen et al. (2007) referred to the argument that Stohler (2006) made that CLIL should have no positive or negative effects on content learning because language and knowledge are intimately related and cannot be distinguished. Another related belief is that because the medium of learning in the CLIL classroom is less perfectly known than the L1, this could be expected to reduce subject competence due to imperfectly understood content and to teachers' attempts to simplify content to make it more understandable (Hajer, 2000, as cited in Van de Craen et al., 2007).

However, many studies have shown that students who study school subjects in a CLIL context exit with the same content knowledge as their peers who study in their mother tongue (e.g., Xanthou, 2011; Mattheoudakis, Alexiou, & Laskaridou, 2014; Younes, 2016; Seikkula-Leino, 2007; Chostelidou & Griva, 2014). Vollmer et al. (2006) assessed these results and claimed that CLIL students work on their tasks more persistently, which leads to the acquisition of a higher degree of competence in the subject. In other words, linguistic challenges lead more often to intensified mental construction than to task abandonment, resulting in deeper semantic processing and a better understanding of curricular concepts (as cited in Dalton-Puffer, 2008). According to Jabrun (1997), students in CLIL classrooms expend twice the cognitive effort to understand content through a foreign language, which makes them more effective learners (as cited in De Diezmas, 2016). De Diezmas (2016) found that mental construction, that is, the development of lower- and higher-order thinking skills, and intelligence are cornerstones of CLIL methodologies.

Another important effect of CLIL resides in the increase in learner motivation that it provides. Many studies have shown that CLIL students are more motivated than students who are studying a subject in their mother tongue or using other teaching methodologies. The majority of CLIL learners are intrinsically motivated because they already have a good command of the foreign language and are seeking to speak the language on a native-speaker level (Papaja, 2012). However, the type of motivation might differ by age. Doiz, Lasagabaster, and Sierra (2014) studied the motivation of 393 students (12–15 year-olds) from the Basque Autonomous Community in Spain. They found that CLIL students were more motivated than non-CLIL students. In particular, younger CLIL learners were more intrinsically motivated than their non-CLIL counterparts, while older learners were more extrinsically motivated than their non-CLIL counterparts.

Other consequences of learning with the CLIL methodology were highlighted by Klimova (2012):

1. It builds intercultural knowledge and understanding,
2. It prepares learners for internationalization,
3. It allows students to learn languages the same way that native speakers do,
4. It accesses subject-specific target terminology,
5. It prepares students for their future study and/or working life,
6. It enables the use of a diversity of methods and forms of classroom practice,
7. It complements individuals' learning strategies.

Limitations of CLIL

In spite of the number of advantages that CLIL can provide for language learners, certain limitations and concerns should be highlighted. The first is that there is no CLIL model for exporting, although it has been disseminated throughout Europe and to other parts of the world. However, Marsh, Maljers, and Hartiala (2001) found that the only CLIL model available for export is the so-called no-one version, and they indicated that CLIL models should be designed specifically for each context because a single model will not fit all contexts. CLIL models could differ in terms of the five dimensions of CLIL (culture, environment, language, content, and learning), as well as in learner age, socio-linguistic environment, and degree of exposure to CLIL (Ravelo, 2014).

Another limitation that should be considered is the lack of appropriate preparation and training for teachers. CLIL teachers must be professionals who take an active role in adapting their teaching methodologies, consider their students' characteristics, and analyze and personalize the CLIL context (Coyle, Hood, & Marsh, 2010; Ravelo, 2014). The outcome of this is that knowledge of CLIL is a minimum requirement, such that CLIL teachers need professional training. Coyle et al. (2010) asserted that formal language teaching must be part of CLIL. Hence, teachers who can reposition their teaching philosophies in response to new demands can become successful CLIL teachers, so long as they do not fail to recognize the potential of CLIL.

CLIL teachers find it difficult to select appropriate content to achieve target language objectives. Concerns have been raised about the content of CLIL lessons, such as student failure to understand content due to their limited

command of the language (Ravelo, 2014) and teachers' failure to find appropriate authentic materials to promote active learning (Ravelo, 2014; Mehisto, Marsh, & Frigols, 2008). Ravelo (2014) suggested that the former issue be resolved by teaching the students useful learning strategies and that the latter be addressed with translations of authentic texts and student involvement in the comparison of primary and secondary sources.

Related Research

Many empirical studies conducted in European countries have found a positive impact of CLIL education on vocabulary learning and retention in English. CLIL students have high scores on vocabulary tests, and their vocabulary size increases through CLIL study, as does their retention of new words, regardless of age, gender, educational level, or regional background. However, considering the sample of the current study (higher education students) and as far as the researcher is concerned, many of previous studies that focused on the impact of CLIL on vocabulary learning generally or vocabulary size specifically have been conducted in primary and secondary school contexts, and only a few has focused on higher education.

Samper (2015) investigated the impact of CLIL on English language learning in 15 female students (18–40 years old) studying at the University of Alicante, Spain. The study was conducted during the 2011–2012 academic year and lasted for 5 months, during which the participants were taught English in a CLIL environment for a total of 30 2-hour sessions. The researcher administered vocabulary pretests and posttests using the testing software DAILANG, which is a tool that generates tasks and questions at different levels based on candidates' responses in order to provide accurate data on their level. In that study, 11 students out of 15 surpassed their pretest level to a higher one on the posttest in at least one language skill (listening, writing, reading, vocabulary, and/or grammar). Thus, CLIL had a positive impact on students' language development in general including vocabulary development.

Pérez Cañado (2018) conducted a longitudinal study of the impact of CLIL on English language achievement (vocabulary, grammar, reading, listening, and speaking) relative to educational level (primary, secondary, and baccalaureate education). A sample of 1,033 CLIL students and 991 EFL (non-CLIL) learners from 53 public, private, and charter schools across 12 Spanish communities were involved in this study. The tools employed for information-gathering included a motivation, verbal intelligence, and English language tests. In addition, a questionnaire on students' personal information was administered. It was found that CLIL groups outperformed their non-CLIL counterparts in all language skills, including vocabulary knowledge, at all educational levels. In addition, the results indicated an increase in student performance in delayed posttests, which were administered 6 months after the initial posttests. The delayed tests showed significant differences between the scores of the CLIL and non-CLIL groups in favor of CLIL students.

Another longitudinal study was conducted by Sylvén and Ohlander (2015) which aimed at investigating the impact of CLIL on receptive vocabulary. This study involved 240 students from three schools in Sweden who were 15–16 years old when the study began in 2011. Throughout the study, which lasted for 3 years, the participant groups (five CLIL and three non-CLIL groups) and individuals were compared with each other simultaneously and with themselves longitudinally. During these 3 years, data on the students' motivation and knowledge of vocabulary, synonyms, and collocations were collected, along with measures of other skills. To measure students' vocabulary knowledge, the researcher administered a VLT. The results of that test showed that the CLIL groups significantly outperformed the non-CLIL groups. Further, males outperformed females in both CLIL and non-CLIL groups.

Iglesias Diéguez and Matinez-Andian (2017) studied the relationship between receptive vocabulary and general proficiency, comparing the vocabulary size of CLIL students to that of non-CLIL students and to older non-CLIL students. In that study, 55 Basque-Spanish students were assessed. They were divided into four groups based on the type of instruction received and hours of exposure. The groups were as follows: (a) a non-CLIL group of 12-year-olds ($n = 10$), (b) a CLIL group of 12-year-olds with more hours of exposure than group a ($n = 15$), (c) a non-CLIL group of 14-year-olds with similar hours of exposure to group a ($n = 15$), and (d) a CLIL group of 14-year-olds with more hours of exposure than group c ($n = 15$). The researchers administered a language proficiency test together with 1,000- and 2,000-word frequency-band vocabulary tests. The results showed a statistically significant correlation between vocabulary size and language proficiency in all four groups, indicating that vocabulary has an important role to play in language proficiency. The comparison between the CLIL and the non-CLIL groups showed statistically significant differences between students of the same age in vocabulary size, in favor of the CLIL students. Groups b (12-years-old CLIL students) and c (14-year-old non-CLIL students) received a similar amount of exposure to English (1,116 and 1,173 hours, respectively), and the results revealed no statistically significant differences between the groups in vocabulary size. These results could mean that CLIL simply provides learners with more hours of exposure to a foreign language. This exposure results in increased vocabulary size, which is similar to that of other learners who are 2 years above them. However, if the learners' exposure to foreign language was controlled, this study would perhaps show more accurate results of the impact of CLIL.

The same limitation was observed in a study conducted by Catalán and De Zarobe (2009) that measured vocabulary size in 65 female CLIL students and 65 female non-CLIL students from the north of Spain. When

the research data were gathered, the non-CLIL group had received 629 hours of English instruction, and the CLIL group had received 960 hours and was exposed to English as a medium of teaching a science subject for 6 years and an arts and Crafts subject for 4 years. This study revealed that CLIL students had higher achievement in the three vocabulary tests that were administered by the researcher.

Alonso (2015) compared the size of the receptive vocabulary of 79 CLIL sixth-graders (11–12 years old), 199 non-CLIL sixth-graders (11–12 years old), and 132 non-CLIL tenth-graders (15–16 years old). The groups received 944, 629, and 1049 hours of instruction in English, respectively. The non-CLIL groups were exposed to English through the EFL school subject, while the CLIL group received, in addition to the EFL school subject, English input through a CLIL program where English was used as a medium for teaching natural science. The results showed that the vocabulary size of the tenth-grade students was 936 words, and those of CLIL and non-CLIL sixth-graders were 903 words and 601 words, respectively. Moreover, there was no statistically significant differences between the tenth-grade non-CLIL students and sixth-grade CLIL students. However, there was a statistically significant difference between the CLIL and non-CLIL sixth-grade students, showing that the additional exposure that the CLIL students received in the CLIL program enhanced vocabulary acquisition.

Previous studies revealed that participants in CLIL classrooms generally score higher than non-CLIL students of the same age. However, the scores were generally lower than the scores of other learners involved in similar studies and who received similar amounts of instruction in CLIL classes, in the same context (e.g., AgustínLlach & Terrazas Gallego, 2012). The researchers explained the poorer scores by referring to the types of vocabulary input that the students had been exposed to, recommending that teachers revise their teaching methodologies (as cited in Alonso, 2015).

Fernández-Fontecha (2014) focused on gender differences in receptive vocabulary in a group of 55 CLIL fifth-graders (10–11 year-olds). The participants received 839 hours of instruction in EFL classes, 524 hours of English as a medium of instruction, and 315 hours of English in a CLIL medium with instruction in the natural sciences. A 2,000-word frequency-band VLT was adopted from Schmitt et al. (2001) to measure vocabulary size. No statistically significant gender-based difference was found in vocabulary size, with a very slight difference in favor of males, with the means of vocabulary size at 10.93 for males and 10.16 for females (the maximum score was 30). However, no group reached 1,000 words: males had 728 words and females had 677.

Alonso (2015) and Fernández-Fontecha (2014) showed that Spanish students (10- to 12-year-olds) who were being taught using CLIL instruction had fewer than 1,000 of the most frequent English words, and no significant differences were found in vocabulary size between males and females. However, these studies did not involve a comparative group to show the impact of CLIL education relative to other classes that implement other methodologies.

Many of the studies reviewed above have found no differences between males' and females' achievement in vocabulary. The literature suggests that foreign language learning is a largely feminine terrain, with women being more inclined to study foreign languages and outperforming their male counterparts (Cenoz, Genesee, & Gorter, 2013). However, it appears that CLIL eliminates this gap and produces equal language achievement in males and females and equivalent attitudes toward the target language. For example, Agrelo (2010) examined the impact of CLIL on students' language skills. That study involved a sample of 287 students in the fourth year of secondary education from 10 secondary schools in Galicia, Spain. CLIL courses were taken by 154 students (44.8% males and 55.2% females) for two years, and 133 students (54.9% males and 45.1% females) took the regular English-language course. Students' language proficiency was assessed with a skills-based test designed by teachers and advisors based on assessment criteria drawn from the Common European Framework of References (CEFR). The study findings indicated that there were no significant differences in the four language skills between the male and female students who studied using CLIL method. The global scores of males and females were also close (70.55 and 69.28 for males and females, respectively).

These results are explained in the literature through an investigation of motivation as an important determinant for achievement. This body of research suggests that males and females differ in their motivation, regardless of the teaching methodology used. In general, females are more motivated to learn foreign languages (Heinzmann, 2009). However, Carr and Pauwels (2006) demonstrated that this pattern does not hold for all foreign languages: German, French, Italian, and Spanish are learned better by women than Japanese, Chinese, Russian, Modern Greek, and Latin, which are learned better by men. Similarly, the subjects of mathematics and physics are considered masculine (as cited in Heinzmann, 2009). To clarify this point, Schmidt, Boraie, and Kassabgy (1996) conducted a study that found that females are better language learners because they are intrinsically motivated, while male students have more extrinsic motivation. Hence, Roquet, Llopis, and Pérez-Vidal (2016) concluded that CLIL helps eliminate gender-based differences because it helps male students become more extrinsically motivated to learn the foreign language due to its integration with other subjects that motivate them.

However, Roquet, Llopis, and Pérez-Vidal (2016) found that CLIL does not eliminate the gender gap in language learning. They examined the influence of the gender variable on English-language achievement in

Spanish adolescent learners. The study included an experimental group of 50 students who were involved in both a formal English course and a CLIL program, as well as a control group of 50 students, who were involved in the formal English course only. The study addressed writing, reading, listening, grammar, and vocabulary skills. After 1 year of treatment, female students still outperformed males in both research groups. Thus, CLIL may be even more beneficial for female participants in foreign language proficiency.

Concerning vocabulary learning, Heras and Lasagabaster (2015) reported that females continue to outperform males in CLIL classes. They examined the impact of CLIL on affective factors and vocabulary learning, using a sample of 46 students (16-year-olds) from Basque schools. There were 21 non-CLIL students (10 females and 11 males) and 25 CLIL students (12 females and 13 males). Both groups received 1,080 hours of EFL at school, but the CLIL group received an additional 110 sessions of CLIL education in physical education. The researchers administered vocabulary pretests, posttests, and delayed tests in vocabulary to measure the CLIL group's acquisition of subject-related vocabulary. The results indicated that CLIL had a positive impact on learning subject-related vocabulary in both males and females. However, females outperformed males in the three vocabulary tests.

To the best of the researcher's knowledge, CLIL is not widely used in the Middle East. No country in this region has administered CLIL education in public schools or at institutions of higher education, with the exception of some private colleges and schools. Few studies have investigated the impact of CLIL on foreign language learning and vocabulary development in this context. In a nearby location, Xanthou (2010) investigated the impact of three educational trends in second-language learning, including CLIL. That study compared the impact of CLIL, word-list strategies, and teaching in the first language in second-language vocabulary development. The study sample was 60 11-year-old students from Cyprus, who were divided into three groups: a control group ($n = 21$) that was taught geography in five 40-minute sessions in the students' L1, an experimental group ($n = 24$) that was taught geography in five 40-minute sessions in English (CLIL group), and an experimental group ($n = 15$) that was provided with lists of English-language vocabulary and L1 equivalents (word-list group). A pretest and a posttest were administered to measure the students' vocabulary knowledge. The results showed that the CLIL group and the word-list group had significantly higher scores on the posttest than on the pretest, while the non-CLIL group did not show any improvement. In addition, the CLIL group significantly outperformed the other two groups on the posttest. However, the intervention in this study may have been unequal, as there was no indication whether the control group was exposed to the English vocabulary that they were tested on.

Sanad and Ahmed (2017) studied the impact of CLIL on the development of vocabulary skills and retention in 10 students from Majmaah University, Saudi Arabia. The students were taught 10 lessons with the CLIL method, in which science, history, biology, literature, computer science, geography, communication, politics, environment, and business contents were taught. A vocabulary test was designed and administered as a pretest, posttest, and delayed test. The test results showed a statistically significant difference between the mean scores for the pretest and the posttest, in favor of the posttest. There were no statistically significant differences between the mean scores on the posttest and the delayed test. This means that CLIL had a positive influence on learning and retention of vocabulary. However, it should be noted that this study did not include a control group, nor did it indicate students' amount of previous exposure to English language or the period between the administration of the posttest and the delayed test.

Xanthou (2011) also investigated the impact of CLIL on English vocabulary development in 77 students (11-year-olds) from Cyprus. The sample was divided into four groups: two of them were taught content through the L1, and the other two were taught through CLIL, with English as a medium of instruction. The experiment lasted for 3 weeks, and students received three 80-minute science lessons. The researcher administered a vocabulary pretest and posttest and observed some CLIL lessons. The CLIL groups showed statistically significant differences on the results of the vocabulary tests, in favor of the posttest. The CLIL groups were superior to the non-CLIL groups. However, this study had the same limitations as the previous study by Xanthou (2010), in which non-CLIL groups were not exposed to the vocabulary that they were tested in. Further, the researcher reported that the two groups were taught the same subject content and the same teaching methods were used for both. Here, there is no guarantee that the study intervention was valid, because CLIL education is subject to different teaching procedures than teaching in L1. CLIL is built on the assumption that equal emphasis is given to foreign language learning and subject content learning. It not only substitutes the use of a foreign language for the L1 as a medium of instruction, but it must also include the incorporation of several theories and specific strategies that lead to the achievement of these two points of emphasis.

Concerning the second focus of this study, Xanthou (2011) also studied the impact of CLIL on content mastery. The findings indicated statistically significant differences in content knowledge in favor of the posttest in each of its four groups. However, the comparison between the learning in the experimental and the control groups showed non-significant differences, suggesting that all groups learned the subject-matter content at a roughly similar level.

Other studies have found the same results, suggesting that students, who study school subjects with CLIL, develop the same amount of content knowledge as their peers who study them in their mother tongue. Mattheoudakis, Alexiou, and Laskaridou (2014) investigated the impact of the CLIL methodology on content knowledge of geography. The study involved 51 students (11- to 12-year-olds) from Thessaloniki, Greece. The instruction took place over 1 academic year (9 months), during which 26 students were taught geography through the medium of English (CLIL), and 25 were taught the same subject in the students' mother tongue (Greek). To measure student acquisition of subject knowledge, three content tests were administered to the two groups at three different times. Both groups scored similarly on all three tests. The CLIL group scored higher on two of the tests, and in one of the two, the difference was statistically significant. This means that CLIL students were not negatively affected by the method in terms of content learning.

Younes (2016) studied the impact of CLIL on content learning in 23 Arab high school students (15-year-olds) in the UAE. The students were studying at a private school in Al Sharjah and were of a variety of regional backgrounds (Egypt, Syria, Lebanon, and Palestine). The data-collection tools included written assignments and multiple-choice questions focusing on subject content. CLIL had a positive impact on student content learning. However, the teachers reported a struggle to balance content and language. The researcher argued that foreign-language proficiency influences content learning, as the results showed that language proficiency correlated positively with content mastery. This means that students with higher levels of language proficiency tended to achieve higher scores on the content tests.

Seikkula-Leino (2007) investigated how well CLIL students acquired content in mathematics and Finnish (the students' mother tongue). The study examined 217 students from grades 5 and 6 at a Finnish school; 116 were enrolled in CLIL classes. After the completion of the courses, the participants were tested on mathematics and on Finnish. Their results showed no statistically significant differences between CLIL and non-CLIL students in their achievement on the Finnish test, which supports the supposition that mother tongue skills of students learning in a foreign language are not weaker than in those who are learning in their mother tongue. The results showed a statistically significant difference between students' scores on the mathematics test that favored the CLIL students. However, the number of non-CLIL overachievers was greater than the number of CLIL overachievers. This means that students taught in their mother tongue tended to overachieve more often than those in the CLIL class, which does not support the conclusion that students working in a foreign language would have better success in mathematics than students who are studying mathematics in their mother tongue (Sampera, 1994, as cited in Seikkula-Leino, 2007).

Chostelidou and Griva (2014) investigated 270 higher education students' development of content and language related to their academic and occupational pursuits. The researchers measured students' reading skill in terms of content and language processing. A CLIL test was administered to students to assess their performance in content and language related to their academic and occupational pursuits. Findings showed that there were no statistically significant differences between the CLIL and non-CLIL groups at time of pre-intervention measurement. However, at post-intervention measurement, the CLIL students had statistically significantly higher mean scores than non-CLIL students.

The studies reviewed in this section assert the positive impact of CLIL in vocabulary development and content mastery. Many of these studies involved young learners and few focused on students in higher education. Further, some of these studies were severely limited by their design, including absence of a control group, invalidity of the intervention, and lack of proper control of exposure to the English language. While studies in school education seem to indicate more positive results of CLIL methodology and its impact on different language skills and a variety of content areas, a lot more needs to be established with regards to higher education especially in the Middle east context. Therefore this study set out to bridge the knowledge gap that currently exists in the literature. This study addressed the following questions:

1. To what extent does CLIL affect vocabulary size?
2. To what extent does CLIL affect achievement and retention of content knowledge?

The study sought to fill the gap of previous research in the following ways:

1. selecting participants from students in higher education,
2. including a comparison group,
3. implementing a valid intervention and
4. providing equal exposure to the English language for both the experimental and control groups.

Materials and Methods

Research Design

This study adopted a mixed method design that combined elements of quantitative and qualitative research approaches. To answer the research questions, the study involved a control group and an experimental group. A vocabulary size test (VST) and a content mastery test were administered to collect quantitative data. Qualitative data was collected using interviews and observations.

Participants

The population of this study was 108 full-time students enrolled in the Middle East College in Oman who were studying the oral and written communication module in the fall semester of the academic year 2019–2020. This module was selected because it is an English for specific purposes module (ESP) module. These modules are meant to develop students' communication skills in the workplace. Both language skills and knowledge of communication in the workplace are subject content for this course, which made it suitable for implementing the CLIL method. Additionally, the module is taught for 4 hours per week, which gave the researcher sufficient time to conduct the intervention.

The students who participated in this module were distributed over three sections, but the study sample included only two, with a total of 58 students, who were 19–28 years old. To ensure that the students in these sections had homogeneous levels of vocabulary, the Vocabulary Size Test (VST) was administered to all three sections before the intervention was begun. The VST results were used to select the two groups that were more homogeneous to participate in this study as a control group and an experimental group.

To confirm that the control and the experimental groups that were selected were homogeneous in terms of vocabulary size before the intervention, their scores in the vocabulary size pretest were analyzed using an independent-sample t-test. Table 2 shows the vocabulary pretest scores of the experimental and control groups.

Table 2

Independent-Sample T-Test for Vocabulary Size Pretest Scores (n = 58)

	Group	N	M*	SD	T	Df	p
Pretest	Control	30	3323.3	1967.4	.568	56	.573
	Experimental	28	3044.6	1756.6			

*The total score = 14000

Table 2 shows that there were no statistically significant differences between the scores of the control group and the experimental group. This means that the two groups had equivalent vocabulary size at the outset of the study.

The instructor who taught the control group was a male Filipino instructor with a doctorate in education management and a minor in English studies. He had 16 years of teaching experience, 7 at MEC. The experimental group was taught by a female Indian instructor with a doctorate in teaching English to speakers of other languages (TESOL). She had 24 years of teaching experience, 15 at MEC.

The researcher presented the key principles and objectives of CLIL to the instructor of the experimental group, along with key methods. The teacher collaborated with the researcher in designing the lessons implemented during the intervention. She was also provided with a CLIL teaching checklist drawn from Dale and Tanner (2012), which supported her in the implementation. The researcher used the same checklist during the observations to ensure that the intervention was valid.

Research Tools

To answer the research questions, four tools were employed for information gathering: vocabulary size test, content mastery test, interviews and classroom observations.

Vocabulary Size Test

A vocabulary size test (VST) (incorporating a pretest, posttest, and delayed test) was administered to measure the impact of CLIL on vocabulary size. The test was developed by Nation and Beglar (2007). Nation (2012) described parallel versions of the test for use by both native and non-native speakers of the English language. In this study, the 14,000-word monolingual version was used. This test contains 140 multiple-choice items, with 10 items from each level, corresponding to 1000 word families. It is a test of decontextualized receptive knowledge of written vocabulary, the type of vocabulary knowledge required for reading. However, it does not measure listening vocabulary size or vocabulary for speaking and writing. The test measures knowledge of the form of the written word, the form–meaning connection, and to a smaller degree, concept knowledge (Nation, 2012).

Each test item consists of a word followed by a simple, non-defining sentence that contains the word. The non-defining sentence indicates the part of speech, limits the word meaning in case of a homograph or widely different senses, and cues the meaning by presenting an example of use. Moreover, for each item, learners are provided with four answer options. All three distractors provided are of the same part of speech of the correct answer, and mostly around the same frequency block of 1000 words.

To calculate the total receptive vocabulary size for a given learner, the score on the test should be multiplied by 100. Thus, a score of 45 would imply a vocabulary of 4,500 word families. The researcher used the computerized version of the test. This version automatically calculates the size of the learners' vocabulary.

According to Nation (2012), the different versions of this test were piloted in the following ways:

1. Applied linguists who are native speakers of English read and critiqued the test.
2. Target words were replaced with nonsense words and test-wise native speakers were tasked with guessing the originally correct answer. This was done to check whether it was the choices that indicated the correct answer.

3. The “Range” program was used to check word-frequency levels for the words used in the example sentences and choices for each item in the tests. The Range is a software used to analyze the vocabulary load of texts, including how many and what words occur in a particular text.

4. Rasch-based analyses were conducted with nearly 200 students in Japan (Beglar, 2010).

Although the test was determined to be valid, the researcher nevertheless piloted the test for this study to measure its reliability and appropriateness for the context. The pilot study was performed with 12 students at MEC, 3 males and 9 females in their first semester of study in logistics management. They were 19 to 23 years old. The researcher assessed the instrument using test–retest reliability, with 7 days between test and retest. The results were analyzed using SPSS software. Cronbach’s alpha was computed to ascertain reliability, and a value of 0.99 was found, which indicated that the test had high reliability in this context.

Content Mastery Test

A short test was developed by the researchers in collaboration with the course instructor to investigate achievement and retention of the content taught to the students during the intervention. The questions in the test were written bearing in mind the content and objectives of the oral and written communication module, which included the following five lesson units: communication in the workplace, basic patterns of business messages, fundamentals of report writing, other forms of business communication, and content-based reading.

The test consisted of two parts, with a maximum total of 20 marks. The first part (15 marks) included 15 multiple-choice questions, and the second part (5 marks) featured a fill-in-the gaps question. The test was administered after the intervention (posttest) and again at 8 weeks after the posttest (delayed test).

To confirm the validity of the test, it was checked by six reviewers as a validation jury to determine whether it addressed the objective of the study and of the module. The jury was made up of expert instructors from MEC who teach or have taught oral and written communication module. The jury provided comments on the clarity of the items and difficulty levels. Following these comments, the test was modified.

The researchers performed a pilot study of the test to measure its reliability. The pilot study involved 13 students at MEC, 2 males and 11 females in their first semester of logistics management. They were 18–23 years old. A test–retest reliability was done, in which the same students took the same test twice at a 2-week interval. The results of the test were analyzed using SPSS. Cronbach’s alpha was computed to ascertain test reliability, and it was found to be 0.962, which indicates a high level of reliability.

Interviews

To gain better understanding of the research results, the instructor of the experimental group was interviewed after the intervention. The interview questions focused on the administration of instruments, implementation of the intervention, students’ behaviors, and their motivation and enthusiasm. The interview questions were as follows:

1. Did you observe any students cheating while taking the tests?
2. Based on your observations, what do you think about students’ honesty while taking the tests?
3. Did you face any difficulties while planning and teaching through CLIL? If yes, what were they?
4. Many students have been noticed to be unmotivated to learn, as many of them were not participating or even attending the sessions. What do you think about that?

In addition, some students were interviewed to recheck the data collection process in order to address possible honesty issues that perhaps affected the results of the study. It is very important that students complete the questionnaires and take the tests honestly, if the study is to have value. Hence, two students from the control group and three students from the experimental group were asked the following questions:

1. Did you observe any students cheating while taking the tests?
2. What do you think about your classmates’ honesty while taking the tests?

Besides, to investigate students’ motivation during the intervention, the three students from the experimental group were asked the following question: “Many students have been noticed to be unmotivated to learn, as many of them were not participating or even attending the sessions. What do you think about that?”

Observations

Observations were conducted during the experimental group’s sessions to insure the validity of the intervention. The control group was also observed during five sessions of the oral and written communication module. Notes about teaching methods and techniques and students’ behaviors were taken during observations.

Procedures

During the experiment, the control group received no intervention. Participants were taught the content and skills of the module in the usual way. This included lecturing, questions by the instructor addressed to the students, and in-class activities that offered limited opportunity for collaboration. Likewise, very little or no attention was given to language use or vocabulary learning. For each class session, the students received reading materials through the “Moodle” platform a few days before the class sessions to enable them to prepare. Moodle is a learning management system designed to allow educators and administrators to create online learning courses. In class, the instructor presented the session topic and elicited from the students what they knew about

it. For some sessions, students received activities on the session topic, but these activities were largely done individually and corrected in whole-class discussions.

Students in the experimental group, by contrast, were taught using the CLIL method, where equal attention was given to content learning and language learning. The procedures incorporated a focus on the CLIL session components, as presented by Dale and Tanner (2012), namely, activating, guiding understanding, focus on language, focus on speaking, focus on writing, assessment, review, and feedback. Additionally, intensive use of graphic organizers, communicative activities, and vocabulary learning strategies was adopted.

In the activating phase of the CLIL method, the instructor used graphic organizers, visuals (including pictures and videos), and games to elicit what students already know about the session's topic and any language related to the topic or to review content and language from the previous session. In the guiding understanding phase, the instructor used a range of different input—slideshows, videos, graphic organizers, and text—to present the information and skills that students were expected to acquire. In this phase, students mainly practiced reading and listening skills while learning the content. During the third phase, where the focus turned to language, the instructor presented items on language use and strategies for language learning. The instructor thus helped the students recycle vocabulary, notice sentence structures, use a vocabulary file or glossary, and apply strategies to learn vocabulary.

For the phases of focus on speaking and focus on writing, students performed pair and group work, in which they communicated actively to solve problems related to the topic and filled each other's information gaps. The activities in these sections included completing or making graphic organizers, fill-in-gaps activities, discussions, group writing, and presentations. Sometimes, these two phases were combined and activities that focused on both speaking and writing were implemented. In addition, in the assessment, review, and feedback phase, the instructor assessed student learning by implementing short oral or written quizzes. She also gave feedback on students' spoken and written language.

Data Analysis

The data obtained from the vocabulary pretest, posttest, and delayed test were analyzed using SPSS statistics software version 22. An independent-sample t-test was used to compare the pretest, posttest, and delayed-test scores for the control and the experimental groups. A paired-sample t-test was also used to compare the pretest, posttest, and delayed-test scores for each group separately.

Similarly, the data obtained from the content mastery test were analyzed using SPSS statistics software version 22. An independent-sample t-test was used to compare the scores for the posttest and the delayed test in the control and experimental groups. A paired-sample t-test was used to compare the posttest and the delayed-test scores for each group separately. In addition, the data collected from the interviews and observations were assessed to support and interpret the results of the other instruments used in the study.

Results

Impact of CLIL on Vocabulary Size

To address the question on the impact of CLIL on vocabulary size, the VST developed by Nation and Beglar (2007) was administered to the control group and the experimental group before and after the intervention, as well as 8 weeks after intervention. Table 3 presents the results of an independent-sample t-test for the scores of the experimental and control groups on the vocabulary pretest. The table shows that there were no statistically significant differences in the pretest scores for the control group and the experimental group, $t(56) = .568$, $p = .573$, (two-tailed).

Table 3

Independent-Sample T-Test for Vocabulary Size Pretest Scores (n = 58)

	Group	N	M*	SD	T	Df	p value
Pretest	Control	30	3323.3	1967.4	.568	56	.573
	Experimental	28	3044.6	1756.6			

*The total score = 14000

Table 4 presents the results of an independent-sample t-test for the scores of the experimental and control groups on the vocabulary posttest. The table shows that there were no statistically significant differences in the posttest scores of the control group and the experimental group, $t(56) = -1.157$, $p = .252$ (two-tailed).

Table 4

Independent-Sample T-Test for Vocabulary Size Posttest Scores (n = 58)

	Group	N	M*	SD	T	Df	p value
Posttest	Control	30	3673.3	1590.5	-1.157	56	.252
	Experimental	28	4225	2029			

*The total score = 14000

Table 5 presents the paired-sample t-test for the scores for the experimental and the control groups in the pretest and posttest of the VST.

Table 5

Paired-Sample T-Test for Vocabulary Size Pretest and Posttest Scores (n = 58)

Group		N	M*	SD	T	Df	p value
Control	Pretest	30	3323.3	1967.4	-2.228	29	.034
	Posttest	30	3673.3	1590.5			
Experimental	Pretest	28	3044.6	1756.6	-5.639	27	.000
	Posttest	28	4225	2029			

*The total score = 14000

Based on table 5, the results of the test showed a statistically significant increase in the control group's scores on the VST after intervention, $t(29) = -2.228$, $p < .05$, (two-tailed). As well, the results also showed a statistically significant increase in the experimental group's scores in the VST after receiving the intervention, $t(27) = -5.639$, $p < .001$, (two-tailed).

The results of the independent-sample t-test and the paired-sample t-test shown in tables 3, 4 and 5 indicated that the experimental and control groups scored significantly higher in the posttest. The control group's mean score increased by 350 word families, and the experimental group's mean score increased by 1180.4 word families. Although the learning gain of experimental group was higher than that of the control group, this difference was not statistically significant.

To explore the impact of CLIL on students' vocabulary size further, the VST was administered for the third time 8 weeks after the posttest. Table 6 presents the results of the paired-sample t-test for the scores of the experimental and control groups in the vocabulary posttest and delayed test.

Table 6

Paired-Sample T-Test for Vocabulary Size Posttest and Delayed-Test Scores (n = 58)

Group		N	M*	SD	T	Df	p value
Control	Posttest	30	3673.3	1590.5	-1.567	29	.128
	Delayed test	30	3856.6	1349.2			
Experimental	Posttest	28	4225	2029	-1.718	27	.097
	Delayed test	28	4425	1794			

*The total score= 14000

Table 6 shows that there were no statistically significant differences in the control group's scores in the VST 8 weeks after the intervention, $t(29) = -1.567$, $p = .128$, (two-tailed). There were also no statistically significant differences in the scores of the experimental group in the VST 8 weeks after receiving the intervention, $t(27) = -1.718$, $p = .097$, (two-tailed).

Table 7

Independent-Sample T-Test for Vocabulary Size Delayed-Test Scores (n = 58)

Group		N	M*	SD	T	Df	p value
Delayed test	Control	30	3856.6	1349.2	-1.369	56	.176
	Experimental	28	4425	1794.1			

*The total score = 14000

Table 7 presents the results of the independent-sample t-test for the scores of the experimental and control groups on the delayed test. There were no statistically significant differences between the scores of the control and experimental groups in the VST 8 weeks after the intervention, $t(56) = -1.369$, $p = .176$, (two-tailed).

Overall, the results of the VST showed that there were no statistically significant differences between the scores of the control and the experimental groups in the pretest, posttest, and delayed test. These do not support the conclusions of previous research, which has indicated that CLIL enhances vocabulary learning and increases vocabulary size (e.g., Alonso, 2015; Catalán & De Zarobe, 2009; Fernández-Fontecha, 2014; Sylvén & Ohlander, 2015; Iglesias Diéguez & Martínez-Andian, 2017; Pérez Cañado, 2018; Samper, 2015; Xanthou, 2010; Moghadam & Fatemipour, 2014; Sanad & Ahmed, 2017; Heras & Lasagabaster, 2015). However, the results did provide a good indicator of the positive impact of CLIL on vocabulary size, which was shown in the bigger learning gain of the experimental group. The learning gain of the experimental group between the pretest and the posttest is 1180.4 word families, while that of the control group is only 350 word families.

The data in the tables above indicate that the control group became more homogeneous during the period of the experiment, as the standard deviation for the group's scores decreased from 1967.4 in the pretest to 1590.5 in the posttest and then to 1349.2 in the delayed test. On the other hand, the standard deviation for the experimental group's score increased from 1756.6 in the pretest to 2029 in the posttest and then decreased to 1794.1 in the delayed-test. This indicates that the group became less homogeneous. The fluctuation of the experimental group's homogeneity is one of the reasons for the non-significant difference between the control and the experimental groups' scores on the VST. The increased variability in the experimental group led to greater pooled variance, which resulted in lower t values and higher p values. Hence, to decrease the pooled variance

and the p value of the independent-sample t-test results, the researcher omitted some of the subjects of the study to obtain more proximate levels of variability in the control group and the experimental group. After several attempts, the researchers located some outliers in the data and omitted the results of seven participants in the control group and four participants in the experimental group.

Table 8

Independent-Sample T-Test for Vocabulary Size Pretest, Posttest, and Delayed-Test Scores after Omitting Outliers (n = 47)

	Group	N	M*	SD	T	Df	p. value
Pretest	Control	23	2982.6	1335.7	1.368	45	.178
	Experimental	24	2500	1073.4			
Posttest	Control	23	3434.7	1211	-.356	45	.723
	Experimental	24	3562.5	1244.4			
Delayed test	Control	23	3617.4	1034	-.682	45	.499
	Experimental	24	3825	1050.9			

*The total score = 14000

Table 8 shows the independent-sample t-test for the scores of the experimental and the control groups in the pretest, posttest, and delayed test after omitting outliers. Outliers were omitted to decrease variability in the experimental group in order to test if levels of significance can be achieved through this adjustment. Results persistently indicated no statistically significant difference between the scores of the control group and the experimental group in the pretest, posttest, or delayed test, even after removing outliers. Yet, variability did increase in the experimental group's scores in the posttest but by a lower degree than in the data obtained before omitting the outliers.

Impact of CLIL on Content Achievement and Retention

To address the question of the impact of CLIL on students' achievement and retention of content knowledge, a content mastery test was administered to the control group and the experimental group immediately after the intervention and 8 weeks after the intervention.

Table 9

Independent-Sample T-Test for Content Mastery Posttest Scores (n = 58)

	Group	N	M*	SD	T	Df	p value
Posttest	Control	30	8.7	3.4	-.285	56	.777
	Experimental	28	9	3.73			

*Total possible score = 20

Table 9 presents the results of the independent-sample t-test for the experimental and the control groups' scores in the content-mastery posttest. The table shows that there were no statistically significant differences between the posttest scores of the control group and the experimental group, $t(56) = -.285$, $p = .777$, (two-tailed).

Table 10

Independent-Sample T-Test for Content Mastery Delayed-Test Scores (n = 58)

	Group	N	M*	SD	T	Df	p value
Delayed test	Control	30	8.5	3.26	-.396	56	.693
	Experimental	28	8.8	2.88			

*Total possible score = 20

Table 10 presents the results of the independent-sample t-test for the experimental and the control groups' scores in the content-mastery delayed-test. The table shows that there were no statistically significant differences between the delayed-test scores in the control group and the experimental group, $t(56) = -.396$, $p = .693$, (two-tailed).

Table 11

Paired-Sample T-Test for Content Mastery Posttest And Delayed Test (n = 58)

	Group	N	M*	SD	T	Df	p. value
Control	Posttest	30	8.7	3.4	.705	29	.487
	Delayed test	30	8.5	3.26			
Experimental	Posttest	28	9	3.73	.464	27	.647
	Delayed test	28	8.8	2.88			

*Total possible score = 20

Table 11 presents the results of the paired-sample t-test for the scores of the experimental and control groups in the content-mastery posttest and delayed test. There were no statistically significant decreases in either group's scores on the content mastery test 8 weeks after the intervention.

The results of the independent-sample t-test and the paired-sample t-test in tables 9, 10 and 11 showed slight differences between the scores for the control and experimental groups in the content mastery posttest and delayed test. There was a slight decrease in the scores for the two groups in the delayed test. However, these differences were not statistically significant. This means that the control group and the experimental group

acquired similar amount of content knowledge during the experiment. Further, students in both groups were able to retain the knowledge that they had acquired during the intervention. Thus, CLIL had no statistically significant impact on achievement or retention of content knowledge.

These results were shared with the instructor of the experimental group to determine possible factors for these results. The instructor responded that similar results were found by her and the other instructor of the other group in the oral and written communication module. Table 12 shows the mean scores of the control group and the experimental group for the module assessments over the semester.

Table 12

Mean Scores for Module Assessments over the Semester

Group	Assessment 1	Assessment 2	Assessment 3	Assessment 4	Total mean
Control	15.3	25.95	8.07	21.79	71.11
Experimental	13.33	29.45	7.09	20.36	70.23

Table 12 shows that there was only a slight difference in the mean scores for the two groups, which means that the control group and the experimental group reached the same level of achievement in the oral and written communication module assessments. These results strongly support the findings of the content mastery test administered by the researchers.

Discussion

The results of the VST and the content mastery test showed that CLIL had no statistically significant impact on vocabulary size and content achievement and retention. While the results of previous studies promoted the effectiveness of CLIL, this study does not support this conclusion.

To explain the difference in the results of this study, factors related to the research sample, the application of the intervention, and the teaching method must be considered. First, the researcher observed that many students did not attend their classes punctually. The majority submitted their classwork, homework, and projects late. Some students did not speak in class, even when they had completed the written work. Further, some students did not like to participate in most sessions. This could be interpreted to mean that the students were not highly motivated to learn.

According to the literature, students taking CLIL courses are expected to develop more positive attitudes toward the modules and to have increased motivation to learn. Papaja (2012) found that CLIL students were more motivated than students who study in their mother tongue or with another teaching methodology. Motivation is not the only contributor to achievement, but it is also an important outcome in itself. According to Dale and Tanner (2012), CLIL instructors should encourage learners to interact and should adopt activities that promote independent thinking, speaking, and writing. They also argued that CLIL learners are more motivated to learn and can develop a strong sense of achievement as they notice their own rapid progress in the foreign language.

To follow up on these considerations, three students from the experimental group were interviewed on their observed behaviors and motivation. Students asserted that even though that the majority of students might seem unmotivated, they actually enjoyed the sessions. However, one student noted that others were more enthusiastic at the beginning of the semester, but then their enthusiasm diminished. The students argued that they were stressed by the amount of material and the number of assignments in this module and their other modules. Nonetheless, they stated that they still felt motivated to learn and improve their language level. One student reported that some of her classmates worked hard all semester, while others were careless and only wanted to pass. The teacher of the experimental group also reported that a high percentage of the students were very motivated and enthusiastic.

Concerning students' honesty while taking the tests, four students reported during the interviews that they answered all of the test questions to the best of their ability. However, one student admitted that she was not reading all of the questions carefully because she did not consider the tests to be important. The last student stated that not all of the students in his class completed the test honestly, as they felt bored while taking them. He believed that about 70% of students were responding to the questions as best they could.

Another possible explanation for the study results could be the unrealistic key principle of CLIL. The researcher observed that the instructor of the experimental group faced a challenge in encouraging the students to participate and interact, perhaps because this was her first time teaching with CLIL. To shift from being a language teacher to be a CLIL teacher is not easy. According to Dale and Tanner (2012), CLIL gives teachers new roles and responsibilities, such as activating their students' prior knowledge and encouraging them to participate and interact. New CLIL teachers also face many challenges, including the need to focus on both language and content teaching and to assess learners' progress on both tracks.

The instructor for the experimental group noted that planning teaching materials and activities that focus on both language and content was a challenge. Although the content of the module was predetermined, the language that students would use could not be predicted. Hence, she found that she needed to brainstorm all of the possible language items that the students might need for each session and then predict the kind of errors that they would be likely to commit. She also reported that the most difficult task was recognizing whether the students had achieved both the language learning objectives and the module objectives. This was judged based on their

achievement of assigned tasks. However, it was not always possible to decide when to focus on students' mistakes related to the content and those related to the language, keeping in mind that too much feedback might frustrate the students.

CLIL is defined as a dual-focused method, in which a foreign language is a means of teaching and learning both content and language (Coyle, 2006; Marsh, 2013). A key principle of CLIL is that equal focus should be given to language learning and content learning. However, while this sounds perfect on paper, in practice, it is a difficult issue. According to Cammatara and Tedick (2012), it is impossible and unrealistic to balance content instruction and language instruction. Hence, if instructors do attempt to balance these aspects, both may become problematic (as cited in Cenoz, Genesee, & Gorter, 2013).

What is more, no single blueprint can be used for content and language integration in different countries and learning situations (Ravelo, 2014). Marsh, Maljers, and Hartiala (2001) reported that the only CLIL model that is available for export is the so-called "no-one version" of CLIL. In other words, CLIL models should be designed specifically for each context, because no model can fit other contexts (Ravelo, 2014). This necessary flexibility of CLIL could be both a strength and a weakness. On the one hand, CLIL programs can cater directly to their students' learning styles, levels of competence, and needs, as well as to their culture and environment (Samper, 2015). On the other hand, this entails much difficult work for teachers, as their knowledge of CLIL may not be sufficient to guarantee their ability to apply it well. Additionally, students might find it difficult to advance in either language, content, or both (Coonan, 2007 as cited in Bruton, 2011).

Consideration of the limitations of previous studies may also provide possible explanations of the research results. Many authors have argued that CLIL is more selective than it is beneficial. According to Bruton (2011), many of the studies that have found CLIL to be effective and beneficial have involved numerous anomalies in research and analyses, which cast doubt on their conclusions. Bruton reviewed several studies (e.g., Lasagabaster, 2008; Lasagabaster & Sierra, 2009) and found that CLIL students were advantaged by factors that went beyond CLIL instruction itself, such as motivation, self-esteem, attitudes toward learning English and hours of exposure to English. In many studies, the students began the CLIL program with higher motivation, more positive attitudes toward English, more hours of instruction in English, and higher starting proficiency. This implies that supposed benefits may be due to student selection. Structural selectivity appeared to have a greater impact on students' achievement than CLIL itself (Breidbach & Viebrock, 2012). However, Bruton (2011) cautioned that this finding does not imply that CLIL is not beneficial, as positive results have been observed that are not questionable in this way, but the effectiveness of CLIL for the average student should be investigated, as well as its impact on students who are not selected in a selective scenario.

In this vein, in some studies that have focused on vocabulary size (e.g., Alonso, 2015; Catalán & De Zarobe, 2009; Fernández-Fontecha, 2014; Sylvén & Ohlander, 2015; Iglesias Diéguez & Matinez-Andian, 2017); many researchers did not sufficiently control the amount of exposure to English in their studies. According to Catalán and De Zarobe (2009), CLIL instruction usually comes hand-in-hand with more hours of instruction, which implies more exposure to the target language. Hence, they argued that it cannot be clear whether the results of their study and of similar studies are due to the use of CLIL or to the increase in hours of instruction.

The amount of exposure has a large impact on students' vocabulary size. Longer exposure to English leads to increasing vocabulary. Empirical evidence has been found for the positive effects of increasing the hours of instruction on second-language and foreign-language acquisition and development (Catalán and De Zarobe, 2009). Hence, the vocabulary results obtained by these studies (Alonso, 2015; Catalán & De Zarobe, 2009; Fernández-Fontecha, 2014; Sylvén & Ohlander, 2015; Iglesias Diéguez & Matinez-Andian, 2017) might be a result of the hours of exposure that CLIL students received not of the method itself. However, this is mere speculation because, to the researchers' knowledge, no studies have yet shed light on whether CLIL instruction is intrinsically beneficial for increasing learners' vocabulary apart from the greater exposure it provides to the English language.

This study came to fill this gap in the literature by controlling the variable of the exposure to English language. First of all, the vocabulary pretest showed that students in the control and experimental group had similar levels of vocabulary before the intervention. Besides, during the experimental period, students in both groups took the same four college modules in English, including the module in which the intervention was implemented. This means that the control and the experimental groups received the same amount of exposure to English in their studies during the period of the intervention. Hence, any differences in the resulting changes in vocabulary on the posttest and delayed test would be attributed to the intervention. However, no statistically significant differences were found between the two groups. Yet, the vocabulary learning gain of the experimental group was higher than that of the control group, which may indicate that CLIL had a positive impact on students' vocabulary size.

Another possible explanation for the unexpected results of this study could be found in the age of the sample. According to Marsh (2002), following Krashen's hypothesis, CLIL is concerned with language acquisition, not language learning. The key principle of CLIL is not that the content of the syllabus is learned through the language, nor that the language is learned through the content; rather, there is a reciprocal process, in which

acquisition of the target language occurs at the same time as the syllabus content. According to the critical period hypothesis, the first few years of human life is the most crucial period for acquiring language, and after puberty, the brain gradually loses its plasticity of procedural memory (Paradis, 2004), and achieving full command of a language becomes impossible (Diaz, 2006).

In many previous studies, study samples included young English language learners at the age of puberty, just below it, or just above it (e.g., 15- to 16-year-olds in Sylvén and Ohlander, 2015; 12- to 14-year-olds in Iglesias Diéguez and Matinez-Andian, 2017; 11- to 12-year-olds in Alonso, 2015; 10- to 11-year-olds in Fernández-Fontecha, 2014; 16-year-olds in Heras and Lasagabaster, 2015; 11-year-olds in Xanthou, 2011; 11- to 12-year-olds in Mattheoudakis, Alexiou, and Laskaridou, 2014; and 15-year-olds in Younes, 2016). These studies have produced results that support the use of CLIL to enhance language and content learning. However, this study involved older learners, 19–28 years old. Learners at this age do not acquire language in

the way that younger learners do, so they must resort to altogether different processing mechanisms (Paradis, 2004). This might mean that because CLIL is based on the principle of acquisition and the implicit learning of language, it might not be suitable for older learners, such as college students.

In addition, the intervention in the current study lasted for one semester only, specifically 60 hours of instruction. This duration is not short, but it may not have been to show statistically significant impact of the CLIL method on students' vocabulary size and content learning and retention. On the other hand, previous studies that considered implementing CLIL education for longer periods have showed statistically significant results that support the benefit of CLIL on enhancing learners' vocabulary size and content mastery (e.g., Mattheoudakis, Alexiou & Laskaridou, 2014; Sylvén & Ohlander, 2015; Heras & Lasagabaster, 2015; Younes, 2016). In these studies, the minimum duration of CLIL education was 1 year, and in some longitudinal studies, it was implemented for 3–4 years. Hence, the positive but insignificant results of the current study might be due to the short duration of the intervention.

An important point to highlight here is that the mean scores obtained by the control and the experimental groups were lower than 5000 word families. The control and experimental groups' pretest mean scores were 3323.3 and 3044.6, respectively. Their scores increased to reach 3856.6 and 4425 word families, respectively, in the delayed-test. These scores are close to the vocabulary size benchmark of non-native speakers of English language that Nation and Waring (1997) reported. Many adult English learners know much fewer than 5000 word families, even though they have been studying English for many years (Nation & Waring, 1997). In comparison, a graduated native English speaker knows about 17,000 word families, while a first year college native student knows about 12,000 word families (Cervatiuc, 2008).

The second focus of this study was the impact of CLIL on students' achievement and retention of content knowledge. The results of the content-mastery test showed slight differences between the control and the experimental groups' scores in the posttest and delayed test, but they were not statistically significant. As well, there was a slight decrease in the scores of the two groups in the delayed test, but it was not statistically significant. This means that the control group and the experimental group acquired and retained the same amount of the content knowledge taught to them during the intervention.

These results are in line with previous studies (e.g., Xanthou, 2011; Mattheoudakis, Alexiou & Laskaridou, 2014; Seikkula-Leino, 2007), which found that CLIL does not negatively affect content learning and that CLIL and non-CLIL students acquire the same amount of content knowledge. However, those studies differed from the current study in the type of non-CLIL group involved. They investigated the impact of CLIL on content mastery by comparing a CLIL group that was taught content and language with English as a medium of instruction with a non-CLIL group that was taught the same content through students' mother tongue. Hence, because the medium of learning in CLIL classrooms is less perfectly known than in the L1, it might be anticipated that this would lead to reduced subject competence, as a result of an imperfect understanding of the content (Hajer, 2000, as cited in Van de Craen et al., 2007). However, it has been found that CLIL students spend twice the cognitive effort to learn the content through a second language, which makes them more effective learners (Jabrun, 1997, as cited in De Diezmas, 2016). Previous studies have also shown both cognitive and academic gains when learners are taught through CLIL (Mattheoudakis, Alexiou, & Laskaridou, 2014). CLIL instruction results in deeper semantic processing and better understanding of curricular concepts. As a result, L2 processing is considered to have strong potential for the study of subject-specific concepts rather than being a hindrance (Vollmer et al., 2006, as cited in Dalton-Puffer, 2011). Hence, in content tests, CLIL students do not receive lower scores than non-CLIL students who are taught the same subject in their mother tongue.

However, the results of the current study cannot easily be compared to those of the previous studies mentioned, and the interpretations given cannot be used to explain the results of this study. In this study, a non-CLIL group was taught subject content through English as a medium of instruction, which means that both the control group and the experimental groups were taught in English, but the methods of teaching were very distinct. The control group was taught through lectures with few individual activities, but the experimental group was exposed to a range of strategies, materials, and communicative activities to enhance students' learning of the content.

Van de Craen et al. (2007) found that subject matter knowledge is not less developed in CLIL contexts than elsewhere, particularly in primary schools. However, they argued that this is not so in secondary schools. Although CLIL is described as a new stage of CLT, which is effective for language development in secondary school students and college students, it does not emphasize teaching non-language content and skills. Hence, it is not guaranteed that the principles and strategies borrowed from CLT, such as information-gap activities, would be more effective than other methods for enhancing learners' acquisition of knowledge.

Conclusion

Overall, this study aimed to investigate the impact of CLIL method on higher education students' vocabulary size and content learning and retention. Studies conducted in Europe and the Middle East have revealed some findings that support the implementation of CLIL in education. The current study sought to limit structural errors in previous studies to determine whether CLIL is intrinsically beneficial. Its results showed that CLIL had a positive impact on students' vocabulary size, but this impact was not statistically significant. The vocabulary learning gain of the CLIL group was slightly higher in the posttest and the delayed-test compared to the non-CLIL group. Similar findings were revealed concerning students' content mastery, which means that CLIL is not more beneficial than the pre-existing means of teaching (lecturing) in terms of content learning. In short, this study shed the light on some benefits of CLIL method and it proposes the possibility of implementing it successfully in higher education contexts. Therefore, future research should consider studying the implementation of CLIL in similar educational conditions in order to further explore the impact of this method. However, the researchers propose extending the duration of the intervention to one academic year or longer. In addition, it is recommended to study the impact of CLIL on other language skills and sub-skills. Researchers should also focus on the gender variable with regards to the impact of CLIL on both language and content learning.

Conflict of interests

The authors declare that they have no conflict of interest.

References

- Agrelo, X. S. I. (2010). An insight into Galician CLIL: Provision and results. In D. Lasagabaster and Y. Ruiz de Zarobe (Ed.), *CLIL in Spain: Implementation, results and teacher training* (pp. 55-97). Cambridge Scholars Publishing.
- Al-Hosni, S. (2014). Speaking difficulties encountered by young EFL learners. *International Journal on Studies in English Language and Literature (IJSELL)*, 2(6), 22-30.
- Al-Issa, A., & Al-Bulushi, A. (2012). English language teaching reform in Sultanate of Oman: The case of theory and practice disparity. *Educational Research for Policy and Practice*, 11(2), 141-176. <https://doi.org/10.1007/s10671-011-9110-0>
- Al-Issa, A. (2006). Language problems facing Omani learners of English. *ORTESOL*, 24, 19-26.
- Alizadeh, I. (2016). Vocabulary teaching techniques: a review of common practices. *International Journal of Research in English Education*, 1(1), 22-30.
- Al-Mahrooqi, R. (2012). A student perspective on low English proficiency in Oman. *International Education Studies*, 5(6), 263-271. <https://doi.org/10.5539/ies.v5n6p263>
- Alonso, A. C. (2015). Receptive vocabulary of CLIL and non-CLIL primary and secondary school learners. *Complutense Journal of English Studies*, 23, 59. https://doi.org/10.5209/rev_CJES.2015.v23.51301
- Alqahtani, M. (2015). The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching and Education*, 3(3), 21-34. <https://doi.org/10.20472/TE.2015.3.3.002>
- Al-Seyabi, F., & Tuzlukova, V. (2014). Writing problems and strategies: An investigative study in the Omani school and university context. *Asian Journal of Social Sciences & Humanities*, 3(4), 37-48.
- Asyiah, D. N. (2017). The vocabulary teaching and vocabulary learning: perception, strategies, and influences on students' vocabulary mastery. *Jurnal Bahasa Lingua Scientia*, 9(2), 293-318. <https://doi.org/10.21274/ls.2017.9.2.293-318>
- Barimani, S. (2011). Immersion program: the Indian context. *Language in India*, 11(4).
- Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language testing*, 27(1), 101-118. <https://doi.org/10.1177/0265532209340194>
- Bialystok, E., Craik, F. I., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage*, 24(1), 40-49. <https://doi.org/10.1016/j.neuroimage.2004.09.044>
- Breidbach, S., & Viebrock, B. (2012). CLIL in Germany: Results from recent research in a contested field of education. *International CLIL Research Journal*, 1(4), 5-16.
- Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39(4), 523-532. <https://doi.org/10.1016/j.system.2011.08.002>

- Catalán, R. M. J., & De Zarobe, Y. R. (2009). The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction. In Ruiz de Zarobe, Y., & Jiménez Catalán, R. M. (2009), *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 81-92). Bristol: Multilingual Matters. <https://doi.org/10.21832/9781847691675-008>
- Cenoz, J. (2015). Content-based instruction and content and language integrated learning: the same or different?. *Language, Culture and Curriculum*, 28(1), 8-24. <https://doi.org/10.1080/07908318.2014.1000921>
- Cenoz, J., Genesee, F., & Gorter, D. (2013). Critical analysis of CLIL: taking stock and looking forward. *Applied Linguistics*, 35(3), 243-262. <https://doi.org/10.1093/applin/amt011>
- Cervatiuc, A. (2008). ESL vocabulary acquisition: Target and approach. *The Internet TESL Journal*, 14(1).
- Chostelidou, D., & Griva, E. (2014). Measuring the effect of implementing CLIL in higher education: an experimental research project. *Procedia-Social and Behavioral Sciences*, 116, 2169-2174. <https://doi.org/10.1016/j.sbspro.2014.01.538>
- Coyle, D. (2008). CLIL—A pedagogical approach from the European perspective. *Encyclopedia of language and education*, 1200-1214. https://doi.org/10.1007/978-0-387-30424-3_92
- Coyle, D. (2006). *Content and language integrated learning: motivating learners and teachers*. Scottish Languages Review.
- Coyle, D. (1999). Theory and planning for effective classrooms: Supporting students in content and language integrated learning contexts. *Learning through a foreign language*, 46-62.
- Coyle, D., Hood, P., & Marsh, D. (2010). *Content and language integrated learning*. Ernst Klett Sprachen.
- Crandall, J., & Tucker, G. R. (1990). Content-based language instruction in second and foreign languages. *ERIC*, 82-96.
- Dale, L., & Tanner, R. (2012). *CLIL activities with CD-ROM: a resource for subject and language teachers*. Cambridge University Press.
- Dalton-Puffer, C. (2008). *Outcomes and processes in content and language integrated learning (CLIL): current research from Europe*. Na. <https://doi.org/10.1075/llt.20>
- Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms*. Amsterdam: Benjamins. <https://doi.org/10.1075/llt.20>
- D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior*, 23(1), 109-122. <https://doi.org/10.1080/10862969109547729>
- De Diezmas, E. N. M. (2016). The impact of CLIL on the acquisition of L2 competences and skills in primary education. *International Journal of English Studies*, 16(2), 81-101. <https://doi.org/10.6018/ijes/2016/2/239611>
- Delliou, A., & Zafiri, M. (2016). Developing the speaking skills of students through CLIL. *ELIC*. <https://doi.org/10.18638/eiic.2016.5.1.517>
- Diaz, C. (2006). *The critical period hypothesis*. FASTA University.
- Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). CLIL and motivation: The effect of individual and contextual variables. *The Language Learning Journal*, 42(2), 209-224. <https://doi.org/10.1080/09571736.2014.889508>
- Education Information Network in the European Union. (2006). *Content and language integrated learning (CLIL) at school in Europe*. Eurydice.
- Fernández-Fontecha, A. (2014). Motivation and gender effect in receptive vocabulary learning: an exploratory analysis in CLIL primary education. *Latin American Journal Of Content & Language Integrated Learning*, 7(2), 27-49. <https://doi.org/10.5294/laclil.2014.7.2.2>
- Heinzmann, S. (2009). Girls are better at language learning than boys: do stereotypic beliefs about language learning contribute to girls' higher motivation to learn English in primary school? *Bulletin VALS-ASLA*, 89, 19-36.
- Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, 19(1), 70-88. <https://doi.org/10.1177/1362168814541736>
- Hurtado, C. J. (2002). Lexical hierarchies as a strategy of teaching vocabulary. In W. Gewehr (Ed.), *Aspects of modern language teaching in Europe* (pp. 176-189). New York, NY: Routledge.
- Hussein, N. O., & Elttayef, A. I. (2003). Arab learners' problems in learning English language: A teacher perspective. *Tech. Rep*.
- Iglesias Diéguez, K. & Martínez-Andian, M. (2017). The influence of CLIL on receptive vocabulary: a preliminary study. <https://doi.org/10.18172/jes.3210>
- Ikeda, M. (2013). Does CLIL work for Japanese secondary school students? Potential for the “weak” version of CLIL. *International CLIL Research Journal*, 2 (1), 31-42
- Ismail, N. S., Zaid, S. B., Mohamed, M. H., & Rouyan, N. M. (2017). Vocabulary Teaching and Learning Principles in Classroom Practices. *Arab World English Journal (AWEJ)*, 8(3), 119-134. <https://doi.org/10.24093/awej/vol8no3.9>

- Janebi Enayat, M., Amirian, S. M. R., Zareian, G., & Ghaniabadi, S. (2018). Reliable measure of written receptive vocabulary size: using the 12 depth of vocabulary knowledge as a yardstick. *SAGE Open*, 8(1), 2158244017752221. <https://doi.org/10.1177/2158244017752221>
- Klimova, B. F. (2012). CLIL and the teaching of foreign languages. *Procedia-Social and Behavioral Sciences*, 47, 572-576. <https://doi.org/10.1016/j.sbspro.2012.06.698>
- Lambert, W. E. (1977). Cognitive consequences of bilingual schooling. *Linguistics*, 18(3), 13-133.
- Larsson, T. (2014). Vocabulary learning: a study of students and teachers' attitudes towards English vocabulary learning in lower secondary school.
- Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, 1(1). <https://doi.org/10.2174/1874913500801010030>
- Lasagabaster, D., & Sierra, J. M. (2009). Immersion and CLIL in English: more differences than similarities. *ELT Journal*, 64(4), 367-375. <https://doi.org/10.1093/elt/ccp082>
- Ma, Q. (2009). *Second language vocabulary acquisition* (Vol. 79). Peter Lang.
- Marsh, D. (2013). *Content and language integrated learning (CLIL). A development trajectory*.
- Marsh, D. (2002). CLIL/EMILE-The European dimension: Actions, trends and foresight potential.
- Marsh, D., Maljers, A., & Hartiala, A. K. (2001). *Profiling European CLIL classrooms: languages open doors, European platform for Dutch education*.
- Mattheoudakis, M., Alexiou, T., & Laskaridou, C. (2014). To CLIL or not to CLIL? The case of the 3rd experimental primary school in Evosmos. In *Major Trends in Theoretical and Applied Linguistics 3* (pp. 215-234). Sciendo Migration. <https://doi.org/10.2478/9788376560915.p13>
- Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL content and language integrated learning in bilingual and multilingual education*. Macmillan.
- Milton, J. (2009). *Measuring second language vocabulary acquisition* (Vol. 45). Multilingual Matters. <https://doi.org/10.21832/9781847692092>
- Moghadam, N. Z., & Fatemipour, H. (2014). The effect of CLIL on vocabulary development by Iranian secondary school EFL learners. *Procedia-Social and Behavioral Sciences*, 98, 2004-2009. <https://doi.org/10.1016/j.sbspro.2014.03.635>
- Nation, I. S. P. (2012). The vocabulary size test: Information and specifications.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Ernst Klett Sprachen. <https://doi.org/10.1017/CBO9781139524759>
- Nation, I. S. P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, Acquisition and Pedagogy*, 14, 6-19.
- Navés, T. (2011). How promising are the results of integrating content and language for EFL writing and overall EFL proficiency. *Content and Foreign Language Integrated Learning*, 103-128.
- Paliwoda-Pękosz, G., & Stal, J. (2015). ICT in supporting content and language integrated learning: experience from Poland. *Information Technology for Development*, 21(3), 403-425. <https://doi.org/10.1080/02681102.2014.1003521>
- Papaja, K. L. (2012). The impact of students' attitudes on CLIL. *Latin American Journal of Content & Language Integrated Learning*, 5(2), 28-56. <https://doi.org/10.5294/laclil.2012.5.2.10>
- Paradis, M. (2004). *A neurolinguistic theory of bilingualism* (Vol. 18). John Benjamins Publishing. <https://doi.org/10.1075/sibil.18>
- Pérez Cañado, M. L. (2018). CLIL and educational level: A longitudinal study on the impact of CLIL on language outcomes. *Porta Linguarum: Revista Internacional De Didáctica De Las Lenguas Extranjeras*, (29), 51-70. <https://doi.org/10.30827/Digibug.54022>
- Ravelo, L. C. (2014). Demystifying some possible limitations of CLIL (content and language integrated learning) in the EFL classroom. *Latin American Journal Of Content & Language Integrated Learning*, 7(2), 71-82. <https://doi.org/10.5294/laclil.2014.7.2.4>
- Roquet, H., Llopis, J., & Pérez-Vidal, C. (2016). Does gender have an impact on the potential benefits learners may achieve in two contexts compared: formal instruction and formal instruction+ content and language integrated learning?. *International Journal of Bilingual Education and Bilingualism*, 19(4), 370-386. <https://doi.org/10.1080/13670050.2014.992389>
- Samper, M. C. (2015). *Evaluation of the implementation of CLIL (Content and Language Integrated Learning) methodology in the didactics of the English language in preschool education course taught in the preschool education teacher undergraduate program at the University of Alicante* (Doctoral Dissertation).
- Sanad, H. A., & Ahmed, M. G. (2017). Using content and language integrated learning (CLIL) to develop EFL reading comprehension skills, vocabulary skills and retention among college students. *Journal of Research in Curriculum, Instruction & Educational Technology*, 3(4), 101-131. <https://doi.org/10.12816/0041909>

- Schmidt, R., Boraie, D., & Kassabgy, O. (1996). Foreign language motivation: internal structure and external connections. *University Of Hawai'i Working Papers in English as A Second Language*, 14(2).
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the vocabulary levels test. *Language Testing*, 18(1), 55-88. <https://doi.org/10.1177/026553220101800103>
- Seikkula-Leino, J. (2007). CLIL learning: Achievement levels and affective factors. *Language and Education*, 21(4), 328-341. <https://doi.org/10.2167/le635.0>
- Sepšiová, M. (2015). CLIL lesson planning. In S. Pokrivčáková, et al. (Ed.), *CLIL in Foreign Language Education: E-textbook for foreign language teachers* (pp. 131-152). Constantine the Philosopher University in Nitra. <https://doi.org/10.17846/CLIL.2015.131-152>
- Surmont, J., Van de Craen, P., Struys, E., & Somers, T. (2014). Evaluating a CLIL student: Where to find the CLIL advantage. In *Integration of theory and practice in CLIL* (pp. 55-72). Brill Rodopi. https://doi.org/10.1163/9789401210614_005
- Sylvén, L. K., & Ohlander, S. (2015). The CLISS Project: receptive vocabulary in CLIL versus non-CLIL groups. *Moderna Språk*, 108(2), 80-114.
- Van De Craen, P., Mondt, K., Allain, L., & Gao, Y. (2007). Why and how CLIL works. An outline for a CLIL theory. *Views*, 16(3), 70-78.
- Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *The Language Learning Journal*, 40(1), 125-137. <https://doi.org/10.1080/09571736.2012.658232>
- Xanthou, M. (2011). The impact of CLIL on L2 vocabulary development and content knowledge. *English Teaching: Practice and Critique*, 10(4), 116-126.
- Xanthou, M. (2010). Current trends in L2 vocabulary learning and instruction: is CLIL the right approach. *Advances in research on language acquisition and teaching: selected papers, Thessaloniki, Greece: Greek Applied Linguistics Association (Gala)*, 459-471.
- Younes, M. I. (2016). *The impact of CLIL on Arabic, English and content learning of Arab high school students in the UAE* (Doctoral Dissertation).
- Zecca-Naples, N. (2015). How Latin became the language of the Roman empire: Unravel Magazine. Retrieved from <https://unravellingmag.com/articles/latin/>

Author Information

Maia Al Hajri

School Teacher Ministry of Education
Muscat Sultanate of Oman

Abdo Al Mekhlafi (Corresponding Author)

Associate Professor Sultan Qaboos University
College of Education Curriculum and Instruction
Department Muscat Sultanate of Oman

Thuwayba Al Barwani

Professor Sultan Qaboos University
College of Education Curriculum and Instruction
Department Muscat Sultanate of Oman
