

Language and Content Outcomes of CLIL and EMI: A Systematic Review

Resultados de lenguaje y contenido de AICLE y EMI: una revisión sistemática

Resultados de linguagem e conteúdo de AICL e EMI: uma revisão sistemática

19

Keith MATHEW GRAHAM

<http://orcid.org/0000-0001-9277-8589>.

Texas A&M University, United States of America.

kmgraham@tamu.edu

Yunkyeong CHOI

<http://orcid.org/0000-0003-1539-4447>.

Texas A&M University, United States of America.

ychoio421@tamu.edu

Amin DAVOODI

<http://orcid.org/0000-0001-9103-3024>.

Texas A&M University, United States of America.

amin@tamu.edu

Shakiba RAZMEH

<http://orcid.org/0000-0002-1869-6566>.

Texas A&M University, United States of America.

shakiba@tamu.edu

L. Quentin DIXON

<http://orcid.org/0000-0002-9122-8064>.

Texas A&M University, United States of America.

qdixon@tamu.edu

Received: 2018-05-25

Sent for peer review: 2018-08-13

Accepted by peers: 2018-08-23

Approved: 2018-08-23

To reference this article in APA style / Para citar este artículo en APA / Para citar este artigo
Graham, K. M., Choi, Y., Davoodi, A., Razmeh, S. & Dixon, L. Q. (2018). Language and content outcomes of CLIL and EMI: A systematic review. *LACLIL*, 11(1), 19-37. DOI: 10.5294/lacil.2018.11.1.2

ABSTRACT. Around the world, language teachers are shifting to content-based instruction (CBI) as a way to teach English, most commonly in the form of Content and Language Integrated Learning (CLIL) or English-Medium Instruction (EMI). With the spread of CBI around the world, it is important to understand how this shift in teaching has affected student outcomes. Using a systematic literature review approach, this study examines current literature on the effect of CBI on language and content outcomes. Twenty-five articles met the inclusion criteria for this study and were examined. The results show mixed findings on the effectiveness of CBI on student outcomes, with the majority of studies showing either positive or neutral effects for CBI when compared with non-CBI classrooms. However, the study also reveals multiple methodological issues that cause difficulties for any strong conclusions about CBI to be made. In addition, while CLIL in Spain has received a lot of research attention, other countries remain understudied. Therefore, this study concludes with a call for future research of CBI outcomes that examine a variety of countries and account for the methodological flaws identified.

Keywords: Content and language integrated learning; English-medium instruction; language outcomes; content outcomes; systematic review.

RESUMEN. Alrededor del mundo, los profesores de idiomas están cambiando a la instrucción basada en el contenido (CBI) como una forma de enseñar inglés, más comúnmente en la forma de Aprendizaje Integrado de Contenido y Lenguas Extranjeras (AICLE) o inglés como medio de instrucción (EMI). Con la difusión de CBI en todo el mundo, es importante entender cómo este cambio en la enseñanza ha afectado los resultados de los estudiantes. Utilizando un enfoque sistemático de revisión de la literatura, este estudio examina la literatura actual sobre el efecto de CBI en el lenguaje y los resultados del contenido. Veinticinco artículos cumplieron los criterios de inclusión para este estudio y fueron examinados. El estudio arrojó resultados mixtos sobre la efectividad de CBI en los resultados de los estudiantes, en los que la mayoría de los estudios muestran efectos positivos o neutros para CBI en comparación con las aulas no CBI. Sin embargo, el estudio también revela múltiples problemas metodológicos que causan dificultades para que se establezcan conclusiones sólidas sobre CBI. Además, mientras en España CLIL ha recibido mucha atención de la investigación, en otros países sigue sin estudiarse. Por lo tanto, este estudio concluye con un llamado a futuras investigaciones de los resultados CBI que examinen una variedad de países y den cuenta de las fallas metodológicas identificadas.

Palabras clave: aprendizaje integrado de contenidos y lenguas extranjeras; inglés como medio de instrucción; resultados del lenguaje; resultados de contenido; revisión sistemática.

RESUMO. No mundo todo, os professores de idiomas estão mudando a instrução baseada no conteúdo (CBI) como uma maneira de ensinar inglês, mais comum na forma de Aprendizagem Integrada de Conteúdos e Línguas (AICLE) ou inglês como meio de instrução (EMI). Com a difusão de CBI no mundo todo, é importante entender como esta mudança no ensino afetou os resultados dos estudantes. Por meio de uma abordagem sistemática de revisão da literatura, este estudo examina a literatura atual sobre o efeito de CBI na linguagem e resultados do conteúdo. Vinte e cinco artigos preencheram os critérios de inclusão para este estudo e foram examinados. O estudo mostrou resultados mistos sobre a eficácia de CBI nos resultados dos estudantes, em que a maioria dos estudos mostram efeitos positivos ou neutros para CBI em comparação com as salas de aula no CBI. Contudo, o estudo também revela múltiplos problemas metodológicos que causam dificuldades para que possam estabelecer-se conclusões sólidas sobre CBI. Além do mais, enquanto na Espanha CLIL tem recebido muita atenção da pesquisa, em outros países continua sem ser estudada. Portanto, este estudo conclui com uma chamada a futuras pesquisas de resultados CBI que examinem uma variedade de países e solucionem as falhas metodológicas identificadas.

Palavras chave: aprendizagem integrada de conteúdos e línguas estrangeiras; inglês como meio de instrução; resultados da linguagem; resultados de conteúdo; revisão sistemática.

Introduction

Over the last few decades, the way English is being taught has undergone a massive change internationally, shifting from teaching English as a foreign language to using English as a medium of instruction (Dearden, 2015). This new form of language teaching is known as Content-Based Instruction (CBI), an umbrella-term that describes classrooms where “students are taught academic content in a language they are still learning” (Lightbown, 2014, p. 3). CBI may be practiced in various forms, with the most common being Content and Language Integrated Learning (CLIL) or English-Medium Instruction (EMI) (Brinton & Snow, 2017).

CBI has seen expeditious growth around the world that has generally outpaced research (Eurydice, 2006; Neghina, 2017). Given the massive change in teaching practice, it is important that an assessment be conducted on whether the goals of content and language outcomes have been met. In order to assess this, our study uses a systematic search of the literature to review what we know about CBI student outcomes. Specifically, this paper examines how CBI teaching practices compare with traditional language teaching.

Theoretical Framework

Two theories inform our interpretation of the CBI literature — the input hypothesis and the cognitive load theory. The input hypothesis (Krashen, 1985) suggests that language acquisition can only occur if the input does not exceed the learner’s acquired language level, $i+1$ level higher. Krashen (1985) argues that such input should be authentic and comprehensible, meaning that it should be neither too easy nor too difficult for the learners. Though Krashen’s input hypothesis has been challenged for being vague, over ambitious, and even dangerous for its emphasis on simplified speech (Liu, 2015), the basis of the hypothesis may be helpful in understanding CBI. Zhao and Dixon (2017) extend Krashen’s work to CBI situations by suggesting that the

“i+1” should not only refer to language but also to content, thus saying content should also not exceed students’ content knowledge plus one level higher. When viewed in this way, effective CBI may be precipitated on the premise that language must not only be noticed, but it also, in terms of content, must be comprehensible.

Cognitive load theory (Sweller, 1988) describes how cognitive resources can be overloaded during learning tasks when learners find their focus split between disparate sources of information related to a learning goal. CBI has the potential to take students’ attention away from a learning goal and cause cognitive overload (Piesche, Jonkmann, Fiege, & Keßler, 2016). By introducing both content and language simultaneously, students could potentially find their focus split between trying to understand content and comprehend language. Thus, cognitive load theory would require that, to be successful in CBI, students need both the prerequisite content background knowledge as well as sufficient language ability in order for their cognitive resources to be focused on the learning objective of the class.

Method

For this paper, we reviewed empirical studies on students’ learning outcomes in CBI classrooms published in peer-reviewed journals between 2008 and 2018, as any research beyond ten years may be too outdated to represent current CBI programs. Five academic databases were searched: Educational Resources Information Center (ERIC), Linguistics and Language Behavior Abstracts (LLBA), Scopus, PsychINFO, and Web of Science. Abstracts were searched using the following: “English Medium Instruction” OR “EMI” OR “Content and Language Integrated Learning” OR “CLIL” OR “Content based instruction” OR “CBI” OR “Content Based Language Teaching” OR “CBLT” AND “Teaching” NOT “French.” We found 645 references with 488 remaining after duplicate removal.

Inclusion and exclusion criteria were used to further narrow the literature. A study was included if (1) the course instructional language was English and the majority of the population’s L1 was not English (i.e., EFL setting); (2) it was designated/entitled/described as

teaching content through English; and (3) it directly measured students' learning outcomes in CBI and non-CBI settings. On the other hand, we excluded book chapters, systematic reviews, meta-analyses, or commentaries and also articles addressing English for academic purposes (EAP) or English for specific purposes (ESP). After considering inclusion and exclusion criteria, 25 of the 488 articles remained and were included in the study.

Results

The twenty-five articles included in the study can be found in Table 1. Twenty-two of the studies were reported as CLIL programs, balancing language and content objectives, whereas three were reported as EMI, typically with a sole focus on content objectives. The included studies come from only two continents, Europe (N=23) and Asia (N=2). Of the European studies, most were from Spain (N=17), along with single studies from Austria, Belgium, Cyprus, the Czech Republic, and Germany. There was one additional study from Europe that examined Germany, Italy, and the Netherlands together. In Asia, one study came from Hong Kong and another from Taiwan. Seven studies were conducted in primary schools, twelve studies were in secondary schools, and an additional two studies had participants from both primary and secondary schools. Additionally, four studies were conducted in tertiary contexts. Nineteen studies examined language outcomes and six examined content outcomes. The following sections detail the findings.

Language outcomes

CBI has been proposed as a replacement for traditional language teaching by many of its supporters, so naturally much research has been conducted to examine whether CBI can produce better language outcomes.

Overall language proficiency

The findings from five studies looking at general proficiency levels show that CBI has had mixed results for language proficiency outcomes for

Table 1. Overview of included articles

Author (Year)	CBI Type/Level	Country	# of Participants	Focus	Result
Agustín-Llach (2016)	CLIL primary	Spain	129	Vocabulary	No difference
Agustín-Llach (2017)	CLIL primary	Spain	140	Vocabulary and writing	No difference
Arribas (2016)	CLIL secondary	Spain	403	Vocabulary	No difference
Basterrechea and del Pilar García Mayo (2014)	CLIL primary	Spain	116	Third-person singular production	No difference
Binterová, Petrášková, and Komínková (2014)	CLIL primary	Czech Republic	39	Mathematics	CLIL > Non-CLIL
Canga-Alonso (2015a)	CLIL Primary and secondary	Spain	410	Vocabulary	CLIL > Non-CLIL
Canga-Alonso (2015b)	CLIL primary	Spain	255	Vocabulary	CLIL > Non-CLIL
Coral, Lleixà, and Ventura (2018)	CLIL primary	Spain	85	General language proficiency	CLIL > Non-CLIL
Dafouz, Camacho, and Urquía (2014)	EMI tertiary	Spain	316	Accounting, finance, and history	No difference
Fung and Yip (2014)	EMI secondary	Hong Kong	199	Physics	Non-EMI > EMI
Gené-Gil, Juan-Garau, and Salazar-Noguera (2015)	CLIL secondary	Spain	50	Writing	Non-CLIL > CLIL (lexical variety)
Gierlinger and Wagner (2016)	CLIL secondary	Austria	87	Vocabulary	No difference
Goris, Denessen, and Verhoeven (2013)	CLIL secondary	Germany, Italy, and The Netherlands	263	General language proficiency	CLIL > Non-CLIL
Hernandez-Nanclares and Jimenez-Munoz (2017)	EMI tertiary	Spain	654	World economic history and world economy	No difference
Lazaro-Ibarrola (2012)	CLIL secondary	Spain	26	Speaking – Morphosyntactic development	CLIL > Non-CLIL
Lorenzo, Casal, and Moore (2010)	CLIL primary and secondary	Spain	448	General language proficiency	CLIL > Non-CLIL
Manzano-Vázquez (2014)	CLIL secondary	Spain	36	General language proficiency and writing lexical errors	No difference
Maxwell-Reid (2010)	CLIL secondary	Spain	24	Writing	CLIL more characteristic of English writing
Mesquida and Juan-Garau (2013)	CLIL secondary	Spain	42	Speaking – Negotiation strategies	CLIL wider variety of strategies
Moore (2011)	CLIL secondary	Spain	158	Speaking – Turn-taking	CLIL more collaborative turns / Non-CLIL more overall

Author (Year)	CBI Type/Level	Country	# of Participants	Focus	Result
Ouazizi (2016)	CLIL tertiary	Belgium	31	Mathematics	CLIL > Non-CLIL
Piesche, Jonkmann, Fiege, and Keßler (2016)	CLIL primary	Germany	722	Science	Non-CLIL > CLIL
Rallo-Fabra and Juan-Garau (2011)	CLIL secondary	Spain	121	Speaking – Pronunciation	CLIL more intelligible and less accented
Xanthou (2011)	CLIL primary	Cyprus	77	Vocabulary	CLIL > non-CLIL
Yang (2015)	CLIL tertiary	Taiwan	29	General language proficiency	CLIL > non-CLIL (Receptive skills only)

Source: Own elaboration.

students. In Spain, Lorenzo, Casal, and Moore (2010) randomly selected 61 schools out of a pool of 403 and administered an experimental language diagnostic test. The diagnostic test measured all four language skills (reading, writing, listening, and speaking) and found that CLIL students far outperformed non-CLIL students in all areas, with overall averages of 62 percent and 38 percent, respectively. Coral, Lleixà, and Ventura (2018) also found primary CLIL students performed better on a state language test measuring reading and listening comprehension, although only slightly. In secondary schools, Goris, Denessen, and Verhoeven (2013) reported that CLIL students in three European countries outperformed their non-CLIL counterparts on a variety of measures. One study at the tertiary level conducted by Yang (2015) explored general language outcomes also reported positive language outcomes, although not to the same degree. Yang (2015) studied the language proficiency of tertiary students enrolled in an English Tourism degree program. Using scores from a national English proficiency test, it was found that students enrolled in the program scored higher than their non-EMI classmates on receptive skills but not higher than the national average, and no difference was found for productive skills. Contrary to the above findings favoring CBI, Manzano-Vázquez (2014) found secondary students enrolled in a CLIL program did not score significantly differently than non-CLIL students on an English placement test despite the CLIL group having had more instructional hours of English at the time of testing.

Receptive language skills

In an attempt to find more specific answers on CBI's effect on student language outcomes, researchers have conducted studies specifically examining receptive language skills. Much of the research on receptive outcomes of CBI has focused on receptive vocabulary. Six studies reported mixed findings on CBI's effectiveness for developing receptive vocabulary with three finding significant differences in favor of CBI (Canga-Alonso, 2015a, 2015b; Xanthou, 2011) and three finding no significant differences between CBI and non-CBI groups (Agustín-Llach, 2017; Arribas, 2016; Gierlinger & Wagner, 2016). The three studies showing positive effects of CBI on receptive vocabulary all took place with primary school students, whereas those showing no difference studied both secondary (N=2) and primary (N=1) schools.

Productive language skills

CBI research on productive skills is much more varied than receptive skills in terms of research focus. From a broad view of productive skills, six studies focused on writing (Agustín-Llach, 2016, 2017; Basterrechea & del Pilar García Mayo, 2014; Gené-Gil, Juan-Garau, & Salazar-Noguera, 2015; Manzano-Vázquez, 2014; Maxwell-Reid, 2010) and four focused on speaking (Rallo-Fabra & Juan-Garau, 2011; Lazaro-Ibarrola, 2012; Mesquida & Juan-Garau, 2013; Moore, 2011). Current research provides little evidence that CBI students enjoy an advantage when it comes to writing. Maxwell-Reid (2010) provides some evidence of the advantages of CBI for writing, finding that the writing of CLIL students in the study seem to display more of the characteristics of English writing, whereas non-CLIL students tended to display characteristics more associated with Spanish writing. However, Basterrechea and del Pilar García Mayo (2014) and Gené-Gil et al. (2015) found very few significant differences in writing between secondary CLIL and non-CLIL students. Basterrechea and del Pilar García Mayo (2014) found no significant difference in the ability of CLIL secondary students to produce the third-person -s, and Gené-Gil et al. (2015), measuring complexity, accuracy, and fluency of writing, found the only significant difference to be in terms of lexical variety, in favor of the non-CLIL group. Like Gené-Gil et al. (2015), other

researchers have also explored the lexical side of writing. Two studies conducted by Agustín-Llach (2016, 2017) showed similar lexical production in the writing of CLIL and non-CLIL primary students. In terms of lexical errors, Manzano-Vázquez (2014) found no differences in the frequency of lexical errors between CLIL and non-CLIL secondary students after accounting for the outliers in the non-CLIL group.

Research on CBI outcomes for speaking is as equally varied as writing, focusing on pronunciation (Rallo-Fabra & Juan-Garau, 2011), negotiation strategies for comprehensibility (Mesquida & Juan-Garau, 2013), turn-taking (Moore, 2011), and morphosyntactic development (Lazaro Ibarrola, 2012). Pronunciation and negotiation strategies are skills that may help second language users increase their comprehensibility in communication. Research suggests CBI students may have an advantage in both of these skills. Rallo Fabra and Juan-Garau (2011) found that CLIL secondary students produced more intelligible and less accented speech than those in a traditional classroom. While Mesquida and Juan-Garau (2013) found no significant differences between CLIL and non-CLIL secondary students in terms of the amount of negotiation strategies used, the authors reported that CLIL students used a wider variety of strategies, possibly meaning these students could negotiate meaning in different manners and situations. Collaborative turn-taking is another technique that may lead to improved communication. Moore (2011) found that, while mainstream foreign language students took more turns overall, the CLIL students took more collaborative turns that facilitated mutual interaction and linguistic/affective support for their conversation partner. Finally, Lazaro-Ibarrola (2012) suggests that CLIL learners show faster growth in morphosyntactic development, leading toward more advanced communication.

Content outcomes

In many CBI classrooms, the purpose is twofold: both “acquiring subject knowledge and competences as well as skills and competences in the foreign language” (Georgiou, 2012, p. 495). In other words, not only students’ acquisition of English skills but also their level of understanding of the subject matter should be a critical criterion for evalu-

ating the effectiveness of this teaching approach. The following studies examine the academic performance of students across various academic disciplines such as math and science in primary or secondary schools as well as tertiary-level courses such as business, accounting, finance, and history.

Mathematics outcomes

Two studies examined the effect of mathematics CLIL instruction at the secondary level (Binterová, Petrášková, & Komínková, 2014; Ouazizi, 2016). In both studies, the students who received CLIL instruction performed better on the mathematics test compared to those who received instruction in their mother tongue. Binterová et al. (2014) assessed students' word problem solving skills and used two mathematics didactic tests (one in English and the other one in the students' mother tongue). Since word problems in mathematics require comprehension of the language, the authors concluded that the CLIL method is more effective in improving students' language skills as well as problem solving skills in math. Ouazizi (2016) used a mathematical test of quadratic equations to measure students' mathematics knowledge. Both the CLIL and non-CLIL group achieved relatively high scores on the test, but the CLIL group scored slightly higher, though it should be acknowledged that the CLIL group had increased instruction time compared to the non-CLIL group, making conclusions difficult.

Science outcomes

Two studies (Fung & Yip, 2014; Piesche et al., 2016) have examined science learning, with both studies finding negative effects for CBI on students' acquisition of science concepts. In Piesche et al. (2016), sixth-grade students in a CLIL science class obtained lower scores than their counterparts in the non-CLIL class on both an immediate post-test and a follow-up test. The CLIL students also did not show strong long-term retention of the science content, contrary to the prominent hypothesis that a bilingual environment would foster long-term memory of the content (Piesche et al., 2016). However, the study participants were new to CLIL and more time may have been needed for them to get used to the CLIL environment.

Similarly, content instruction in L2 seems to be less effective for low-achievers. For example, Fung and Yip (2014) explored the effect of a tenth-grade physics intervention in EMI and CMI (Chinese, L1, as the medium of instruction) classrooms. Among the three physics ability groups, the low-ability EMI students attained a lower level of achievement compared to the low-ability CMI students, while high-ability EMI students performed better than their high-achieving peers in CMI classrooms. Therefore, it seems that various learner factors such as their L2 proficiency, prior experience in CBI, and their prior knowledge of the content area could have a significant influence on the effectiveness of CBI on content learning.

Tertiary-level content outcomes

At the tertiary-level, we found two CBI studies (Dafouz, Camacho, & Urquia, 2014; Hernandez-Nanclares & Jimenez-Munoz, 2017) that examined the academic outcomes of students. Both studies found slightly better performance in the CBI group, but not to a statistically significant degree. For example, Hernandez-Nanclares and Jimenez-Munoz (2017) examined the effect of CBI instruction in a World Economy and World Economy History course. The results of the written examination showed that the CLIL group performed slightly better than the non-CLIL group, although the average grade on the final exam was similar.

Similarly, Dafouz et al. (2014) examined the effect of EMI on Spanish undergraduate students' academic performance in accounting, finance, and history, and compared their performance across these three disciplinary subjects through coursework and final grades. Both groups (EMI and non-EMI) obtained very similar results in the three subjects regardless of the language of instruction. Moreover, the students in the history course obtained slightly higher results than the other two subjects in spite of the higher verbal demands. However, since the instructors for these courses were different, the scoring of each instructor could vary, which is also related to an issue of homogeneity of evaluation criteria. In addition, although the authors claimed that these three groups of students were comparable, their English proficiency was not measured. Due to these methodological issues, it may be difficult to conclude CBI advantages in tertiary education.

Discussion

The mixed results found in the literature make it difficult to arrive at any conclusions about the effectiveness of CBI. While there have been positive findings, there are many issues that call the results of the studies into question. Bruton (2011) outlines many of these issues. For one, the CBI programs in most of the research reviewed were elective programs that may naturally attract students with higher aptitudes or motivation toward foreign language learning. Second, in many of the studies reviewed, CBI courses were reported as having extra instruction time. Given this, it is completely feasible that the gains are simply a result of more instruction. Additionally, the research instruments used, particularly in the receptive skill studies, were designed to measure general everyday language proficiency, or basic interpersonal communicative skills (BICS) (Cummins, 1984). However, much of the learning that occurs in a CBI classroom targets cognitive academic language proficiency (CALP). With instruments only measuring BICS, the CALP language gains were likely not accounted for.

Putting aside these limitations, most of the studies exploring language outcomes have found CBI programs to do as well or better than non-CBI programs. Krashen's (1985) input hypothesis can possibly explain why this may be the case, even in the face of seemingly more difficult content. The input hypothesis requires two conditions for language acquisition to occur: (1) a lot of language input, and (2) input that does not exceed one level higher than the learner's current language level. CBI potentially meets both of these conditions, possibly better than traditional language teaching. For the first condition, CBI provides a lot of natural language input in the classroom, such as opportunities for input through teacher lectures, student conversation, content-related videos, textbooks, and other sources. This flood of language through a variety of resources provides a rich language environment full of input opportunities.

However, input alone does not guarantee language acquisition; the second condition of comprehensible input must be met, an area that CBI once again may have an advantage. As one might expect in a content classroom, CBI classrooms may be rich in content media

such as videos, pictures, diagrams, and other visual representations that help make language in the CBI classroom comprehensible. Additionally, content learned in CBI classrooms is often content students have already learned, or at least have the prerequisite knowledge for, in their L1. When students connect their L1 content knowledge to input they receive in a CBI classroom, the connection can help make the language of the content area comprehensible. Finally, the teaching of content provides opportunities to experience similar language repetitiously throughout units, therefore providing certain language input multiple times. For example, in a science classroom, vocabulary such as *experiment* or *mass* will likely be used across units, allowing for multiple encounters. This is in contrast to a traditional language classroom where the scope is broader and similar language may not be used across units. Though many study limitations limit determinations that can be made about the advantages or disadvantages of CBI for language outcomes, when viewing the conditions of CBI classrooms through the lens of the input hypothesis, CBI seems promising for encouraging language acquisition.

Similar to language outcomes, studies which examined students' understanding of the content have shown contradictory results depending on various research contexts, such as type of subject matter, participants' English proficiency, and measurement of the students' content comprehension. Overall, CBI seems to have positive effects on students' content comprehension, but mostly not to a statistically significant degree.

Caution is needed when interpreting results, though, because of several methodological issues. For example, some studies (e.g., Dafouz et al., 2014; Ouazizi, 2016) did not conduct a pre-test to ensure that the two groups had comparable content knowledge prior to the intervention. Moreover, in some studies (e.g., Dafouz et al., 2014), the homogeneity of evaluation criteria was not met when comparing the content outcomes of CBI and non-CBI courses. There was also an issue of lack of consistency in the treatment due to a wide variety of CBI instruction in different schools. These issues make it difficult to support claims of effectiveness of CBI on students' content learning.

Cognitive load theory (Sweller, 1988) may help explain the findings on CBI content outcomes. In some cognitively challenging disciplines,

such as physics, CBI seems to have a negative effect on students' acquisition of content, particularly for students with little background knowledge or limited English proficiency (Fung & Yip, 2014; Piesche et al., 2016). This suggests that learning content in an L2 could demand a high amount of working memory capacity, making it difficult for students to acquire the content knowledge due to the need to simultaneously process new content and the L2. On the other hand, higher ability students performed better in CBI classrooms than in non-CBI classroom, possibly due to a more reasonable cognitive load (Fung & Yip, 2014). Thus, it seems that CBI might be more suitable for higher level students while L1 instruction may be more beneficial for low-achieving students. However, we caution that policy should not be made on this basis. That is, lower-level L2 students should not, at this point, be denied access to CBI courses based on some partially supported research evidence. The effect of CBI on various levels of students is still questionable because of contradictory findings. In Hernandez-Nanclares and Jimenez-Munoz's (2017) study, none of the students in the CBI classroom were able to reach the highest score while some of the non-CBI students did. Therefore, they concluded that the L1 instruction is more effective for advanced level students at the tertiary-level. On the other hand, a fewer number of students in the CBI classroom failed the course compared to the non-CBI classroom, which could be explained by the positive effect of CBI instruction on students' motivation and attitude toward the class. Due to these conflicting findings, more studies are needed to examine the effect of CBI on different learners, considering their language proficiency, background knowledge in content, as well as their motivation and attitudes toward the class.

Conclusion

This literature review explored current literature on CBI student outcomes, revealing mixed findings for both language and content outcomes. Generally, these studies report that CBI either exceeds non-CBI courses or there is no significant difference. However, caution should be exercised

when interpreting these results due to a large amount of methodological flaws found throughout the literature.

In addition, this current literature review comes with a few limitations. For one, the review was limited to a select few databases. These databases have their own individual criteria for which articles are indexed, meaning articles about CBI that have not been indexed by the selected databases were not included. Second, this review was limited to articles written in English. A few articles were excluded and many more may not have been found due to language. Literature that could have influenced the findings of the review may exist in other languages. Finally, literature from CLIL classrooms in Spain is well represented, while other countries remain understudied. With such an uneven study of contexts, this systematic review may not represent the true realities of CBI globally.

Acknowledging these limitations, this review has revealed issues within CBI literature that deserve further attention. Future CBI research must address the methodological flaws highlighted in this review. Without addressing these issues, the research community will remain unclear about the effect of CBI on student outcomes. Additionally, there is a dire need for studies from other countries in Europe, Asia, and around the world. Also, other varieties of CBI, such as EMI, need more attention. While CBI as a language teaching approach seems promising, much more work is needed before CBI's true potential can be understood.

References

- Agustín-Llach, M. P. (2016). Age and type of instruction (CLIC vs. Traditional EFL) in lexical development. *International Journal of English Studies*, 16(1), 75–96.
- Agustín-Llach, M. P. (2017). The effects of the CLIL approach in young foreign language learners' lexical profiles. *International Journal of Bilingual Education and Bilingualism*, 20(5), 557–573. doi:10.1080/13670050.2015.1103208

- Arribas, M. (2016). Analysing a whole CLIL school: Students' attitudes, motivation, and receptive vocabulary outcomes. *Latin American Journal of Content & Language Integrated Learning*, 9(2), 267–292. doi:10.5294/laclil.2016.9.2.2
- Basterrechea, M., & del Pilar García Mayo, M. (2014). Dictogloss and the production of the English third person “-s” by CLIL and mainstream EFL learners: A comparative study. *International Journal of English Studies*, 14(2), 77–98.
- Binterová, H., Petrášková, V., & Komínková, O. (2014). The CLIL method versus pupils' results in solving mathematical word problems. *The New Educational Review*, 38(4), 238–249.
- Brinton, D. M., & Snow, M. A. (2017). The evolving architecture of CBI. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom* (pp. 2–20). Ann Arbor, MI: University of Michigan Press.
- Bruton, A. (2011). Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010). *Applied Linguistics*, 32(2), 236–241. doi:10.1093/applin/amr007
- Canga-Alonso, A. (2015a). Receptive vocabulary of CLIL and Non-CLIL primary and secondary school learners. *Complutense Journal of English Studies*, 23, 59–77. doi:10.5209/rev_CJES.2015.v23.51301
- Canga-Alonso, A. (2015b). The receptive vocabulary size of Spanish 5th grade primary school students in CLIL and Non-CLIL instruction. *ES Revista de Filología Inglesa*, 36, 63–85.
- Coral, J., Lleixà, T., & Ventura, C. (2018). Foreign language competence and content and language integrated learning in multilingual schools in Catalonia: An ex post facto study analysing the results of state key competences testing. *International Journal of Bilingual Education and Bilingualism*, 21(2), 139–150. doi:10.1080/13670050.2016.1143445
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon, UK: Multilingual Matters.
- Dafouz, E., Camacho, M., & Urquia, E. (2014). ‘Surely they can’t do as well’: A comparison of business students' academic performance in English-medium and Spanish-as-first-language-medium programmes. *Language and Education*, 18(3), 223–236. doi:10.1080/09500782.2013.808661
- Dearden, J. (2015). *English as a medium of instruction—a growing global phenomenon*. Retrieved from <https://ora.ox.ac.uk/objects/uuid:4f72cdf8-b2eb-4d41-a785-4a283bf6caaa>

- Eurydice.** (2006). *Content and language integrated learning (CLIL) at school in Europe*. Brussels: Eurydice. Retrieved from http://www.indire.it/luca-bas/lkmw_file/eurydice/CLIL_EN.pdf
- Fung, D., & Yip, V.** (2014). The effects of the medium of instruction in certificate-level physics on achievement and motivation to learn. *Journal of Research in Science Teaching*, 51(10), 1219–1245. doi:10.1002/tea.21174
- Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J.** (2015). Development of EFL writing over three years in secondary education: CLIL and non-CLIL settings. *The Language Learning Journal*, 43(3), 286–303. doi:10.1080/09571736.2015.1053278
- Georgiou, S. I.** (2012). Reviewing the puzzle of CLIL. *ELT Journal*, 66(4), 495–504.
- Gierlinger, E., & Wagner, T.** (2016). The more the merrier—Revisiting CLIL-based vocabulary growth in secondary education. *Latin American Journal of Content & Language Integrated Learning*, 9(1), 37–63. doi:10.5294/laclil.2016.9.1.3
- Goris, J., Denessen, E., & Verhoeven, L.** (2013). Effects of the Content and Language Integrated Learning approach to EFL teaching: A comparative study. *Written Language and Literacy*, 16(2), 186–207. doi:10.1075/wll.16.2.03gor
- Hernandez-Nanclares, N., & Jimenez-Munoz, A.** (2017). English as a medium of instruction: Evidence for language and content targets in bilingual education in economics. *International Journal of Bilingual Education and Bilingualism*, 20(7), 883–896. doi:10.1080/13670050.2015.1125847
- Krashen, S. D.** (1985). *The input hypothesis: Issues and implications*. New York, NY: Longman.
- Lazaro-Ibarrola, A.** (2012). Faster and further morphosyntactic development of CLIL vs. EFL Basque-Spanish bilinguals learning English in high-school. *International Journal of English Studies*, 12(1), 79–96.
- Lightbown, P. M.** (2014). *Focus on content-based language teaching*. Oxford, UK: Oxford.
- Liu, D.** (2015). A critical review of Krashen’s Input Hypothesis: Three major arguments. *Journal of Education and Human Development*, 4(4), 139–146. doi:10.15640/jehd.v4n4a16
- Lorenzo, F., Casal, S., & Moore, P.** (2010). The effects of Content and Language Integrated Learning in European education: Key findings from the Andalusian Bilingual Sections Evaluation Project. *Applied Linguistics*, 31(3), 418–442. doi:10.1093/applin/amp041

- Manzano-Vázquez, B. (2014). Lexical transfer in the written production of a CLIL group and a non-CLIL group. *International Journal of English Studies*, 14(2), 57–76.
- Maxwell-Reid, C. (2010). Content and Language Integrated Learning (CLIL): The influence of studying through English on Spanish students' first-language written discourse. *Text & Talk*, 30(6), 679–699. doi:10.1515/TEXT.2010.033
- Mesquida, F., & Juan-Garau, M. (2013). CLIL instruction and its effects on the development of negotiation strategies. *Miscelánea: A Journal of English and American Studies*, 47, 125–144.
- Moore, P. (2011). Collaborative interaction in turn-taking: A comparative study of European bilingual (CLIL) and mainstream (MS) foreign language learners in early secondary education. *International Journal of Bilingual Education and Bilingualism*, 14(5), 531–549. doi:10.1080/13670050.2010.537741
- Neghina, C. (2017). *Asia Rising*. Retrieved from <http://www.studyportals.com/intelligence/asia-rising/>
- Ouazizi, K. (2016). The effects of CLIL education on the subject matter (mathematics) and the target language (English). *Latin American Journal of Content and Language Integrated Learning*, 9(1), 110–137. doi:10.5294/laclil.2016.9.1.5
- Piesche, N., Jonkmann, K., Fiege, C., & Keßler, J. (2016). CLIL for all? A randomised controlled field experiment with sixth-grade students on the effects of content and language integrated science learning. *Learning and Instruction*, 44, 108–116. doi:10.1016/j.learninstruc.2016.04.001
- Rallo-Fabra, L., & Juan-Garau, M. (2011). Assessing FL pronunciation in a semi-immersion setting: The effects of CLIL instruction on Spanish-Catalan learners' perceived comprehensibility and accentedness. *Poznan Studies in Contemporary Linguistics*, 47(1), 96–108. doi:10.2478/psicl-2011-0008
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. doi:10.1207/s15516709cog1202_4
- Xanthou, M. (2011). The impact of CLIL on L2 vocabulary development and content knowledge. *English Teaching: Practice and Critique*, 10(4), 116–126.

- Yang, W. (2015). Content and Language Integrated Learning next in Asia: Evidence of learners' achievement in CLIL education from a Taiwan tertiary degree programme. *International Journal of Bilingual Education and Bilingualism*, 18(4), 361–382. doi:10.1080/13670050.2014.904840
- Zhao, J., & Dixon, L. Q. (2017). *English-medium instruction in Chinese universities: Perspectives, discourse and evaluation*. New York, NY: Routledge.

