Iranian Journal of Applied Linguistics (*IJAL*)

Vol. 22, No. 1, September 2019, 109-144

The Effect of CLIL on Language Skills and Components: A Meta-Analysis

Seyyed Ali Ostovar-Namaghi, **Shiva Nakhaee*** *Shahrood University of Technology, Iran*

Abstract

Content and Language Integrated Learning (CLIL) has recently been the focus of numerous studies in language education since it aims to overcome the pitfalls of formfocused and meaning-focused instruction by systematically integrating content and language. This meta-analysis aims to synthesize the findings of 22 primary studies that tested the effect of CLIL on language skills and components. Guiding the analysis are three questions: What is the overall combined effect of CLIL on language skills and components? How do moderators condition the effect of CLIL? To what extent the overall combined effect is conditioned by publication bias? The overall effect size was found to be g=0.81, which represents a medium effect size with respect to Plonsky and Oswald's (2014) scale. The results of moderator analysis show that CLIL has the highest effect on students' grammar and listening proficiency and in lower levels of education, especially in elementary schools. It also has the highest effect when combined with hotel management as the subject matter. Fail-safe N test of publication bias shows that the significant positive outcome of CLIL cannot be accounted for by publication bias. The findings have clear implications for practitioners, researchers and curriculum developers.

Keywords: CLIL; skills; components; meta-analysis; effect size; publication bias

Article Information:

Received: 28 February 2019 **Revised:** 11 July 2019 **Accepted:** 22 July 2019

Corresponding author: Shahrood University of Technology, Iran

Email address: shiva.nakhaee@gmail.com

1. Introduction

The Audio-lingual method, produced a host of learners who were grammatically competent but communicatively incompetent. Similarly, Communicative Language Teaching (CLT) produced learners who were communicatively competent but linguistically incompetent (Ma 2003). Content and Language Integrated Learning (CLIL) aims to solve this problem by systematically merging the strengths of formfocused and meaning-focused instruction in language teaching (Marsh, 2000; Moghadam & Fatemipour, 2014; Nikula, Dalton-Puffer & García, 2013). Many studies have been undertaken to test the effectiveness of this hypothetical solution. However, the overall strength of this intervention is not clear; hence, the field is in urgent need of a meta-analysis in order to synthesize the results of the previous studies by estimating the overall effect of CLIL on learner's language proficiency.

2. Review of the Literature

CLIL is a term created in 1994 by David Marsh. He defined CLIL as a situation in which, a school subject is taught through a foreign language and also that a foreign language is taught through a specific subject (Marsh, 2000; Moghadam & Fatemipour, 2014; Nikula, Dalton-Puffer & García, 2013). This approach is a form of bilingual education which aims to provide a bilingual experience for the pupil, even if only for a limited part of the school curriculum (Marsh, 2000; Mattheoudakis, Alexiou, & Laskaridou, 2014; Nikula, Dalton-Puffer & García, 2013).

In this approach, second language competence is an essential tool for content while the first language plays no or only a very subordinate role. In addition, students acquire language in real-life and natural situations rather than learning it through analysis; hence, this method prioritizes fluency and the ability to communicate over accuracy. CLIL is a long-term learning program in which students become proficient

in the second language after five to seven years in a good bilingual immersion program (Marsh, 2000; Mattheoudakis, Alexiou & Laskaridou, 2014).

There are some discrepancies and analogies between CLIL, content-based instruction (CBI) and language immersion (Snow, Met & Genesee, 1989; Tedick & Cammarata, 2012). All the three methods resemble each other since they all take integrated language learning into consideration. As such, they are vividly different from traditional language teaching in which, the focus is only on linguistic features and teaching language. Although CLIL, content-based instruction and language immersion are subparts of integrated language learning approach, they have some subtle differences (Tedick & Cammarata, 2012).

The diversity between CLIL, CBI and language immersion is better understood presenting the continuum of content and language integration (Met 1999; Tedick & Cammarata, 2012). It ranges from the most language-driven end which is frequently used in traditional language classrooms to the most data-driven end which is representative of immersion program. Since CBI is based on language courses, it tends to be nearer to the language-driven end. An ideal CLIL program must be in the middle of the continuum which shows the best integration of language and content in the course (Cenoz & Ruiz de Zarobe, 2015; Met, 1999). Additionally, CLIL is usually regarded as the European version of CBI in that, CBI is frequently used and more popular in the US and Canada (Ruiz de Zarobe 2008; Tedick & Cammarata, 2012).

2.1. The effectiveness of CLIL on learners' proficiency

Although previous empirical studies have suggested that CLIL has some significant effects on the development of language skills and components, the evidence they present is circumstantial and inconclusive. More specifically, different studies have come to different conclusions about the role of such a program in teaching and

learning a second language (e.g., Ackerl, 2007; Bret-Blasco, 2011; Cámara-Ortiz, 2014; Chostelidoua & Grivab, 2014; Dalton-Puffer, 2008; Jäppinen, 2005; JuanGarau, 2010; Kjellén-Simes, 2009; Lasagabaster, 2008; Moghadam & Fatemipour, 2014; Olaizola & Mayo, 2009; Ruiz de Zarobe, 2008; Serra, 2007; Xanthou, 2010).

To start with, previous studies do not seem to agree on the role of CLIL in developing EFL/ESL learners' writing proficiency. For instance, while Dalton-Puffer (2008) suggests that CLIL does not have a significant impact on learners' writing skills, some other studies concluded that CLIL has a positive impact on the development of adolescent learners' writing skills (e.g., Ackerl, 2007; KjellénSimes, 2009; Lasagabaster, 2008).

Considering reading skills, some researchers found that the approach has a significant positive impact on learners' reading skills (e.g., Cámara-Ortiz, 2014; Chostelidoua & Grivab, 2014). Additionally, Lasagabaster (2008) reported that CLIL is more conducive to the development of receptive skills than productive skills in European context. Also, some empirical findings suggest that CLIL learners significantly outperform the non-CLIL learners in listening and reading comprehension, fluency and vocabulary, but not a lot in pronunciation, accuracy and complexity of written and spoken language (Alonso et al., 2008; Dalton-Puffer, 2007; Lasagabaster, 2008; Naves, 2009; Ruiz de Zarobe, 2008). However, some other researchers concluded that CLIL has a positive effect on student's oral performance specially speaking skills (e.g., Bret-Blasco, 2011; Dalton-Puffer, 2008; Juan-Garau, 2010; Ruiz de Zarobe, 2008; Serra, 2007). Taking oral proficiency into account, almost all of the studies concluded that CLIL has a positive impact on students' oral performance and their accuracy and fluency of production (e.g., Bret-Blasco, 2011; Dalton-Puffer, 2008; Gallardo-del-Puerto & Lacabex, 2016; Juan-Garau, 2010; Ruiz de Zarobe, 2008; Serra, 2007).

Considering the vocabulary component of the language, some scholars found that CLIL has a positive impact on learner's vocabulary proficiency (DaltonPuffer, 2008; Juan-Garau, 2010; Moghadam & Fatemipour, 2014; Olaizola & Mayo, 2009; Olsson, 2015; Xanthou, 2010). Others however, emphasized the positive impact of this approach on both receptive and productive vocabulary (e.g., Dalton-Puffer, 2008; Xanthou, 2010). On the contrary, Austad (2013) found that the EFL students scored better on the vocabulary tests than the CLIL students and as such implicitly suggested that CLIL has in insignificant effect on developing learners' vocabulary.

2.2. CLIL and educational levels

CLIL has been implemented from kindergarten to university. Early on, however, the literature has mainly focused on secondary schools and less attention has been drawn to pre-primary and primary levels of education (Austad, 2013; Berendse, 2014; Dallinger, Jonkmann, Hollm & Fiege, 2015; Diéguez & Adrián, 2017; Lahuerta Martínez, 2017; Olsson, 2015; Sylvén & Ohlander, 2015; Moghadam & Fatemipour, 2014).

Crandall (1998) reported the early research on CLIL in primary schools. Afterward, many other programs implemented and examined this approach in primary levels (Korpela, 2013; Kubeš, 2012; Luprichova, 2013; Mäkinen, 2010; Mattheoudakis, Alexiou & Laskaridou, 2014; Menzlova, 2012). While most of them show the positive effect of CLIL in primary schools, only a few of them reported its null or negative effect (Kubes 2012; Mattheoudakis, Alexiou and Laskaridou 2014).

The effect of CLIL on language proficiency has also been tested in higher educational levels such as universities (e.g. Aguilar & Munoz, 2014; Chostelidoua

& Griva, 2014; Kothuri & Nageswari, 2017; Kováčiková, 2013). While most of these studies showed that CLIL has a significant effect on university students, some of them reported that this approach does not have any significant effect on students' language proficiency (Aguilar & Munoz, 2014; Gallardo del Puerto & Adrián, 2015; Kováčiková, 2013). For example, Aguilar and Munoz's (2014) study showed that the difference between the mean scores was significant in pre-and post-listening but it was not significant in pre-and post- grammar tests in university students. In addition, Kováčiková (2013) found that the experimental group reached higher mean scores in the reading and writing sections but the control group significantly outperformed the experimental group in the grammar and vocabulary section. Based on the results of a t-test, he concluded that none of the three scores (vocabulary and grammar, reading, writing) were significantly different. Similarly, Gallardo del Puerto and Adrián (2015) tested the effect of CLIL on university students' oral proficiency and found that EFL learners had significantly higher gains than CLIL learners.

2.3. Subject matter in CLIL

CLIL is a form of bilingual education which aims to provide a bilingual experience for the pupil, even if only for a limited part of the school curriculum (Marsh, 2000; Mattheoudakis, Alexiou, & Laskaridou, 2014). Regarding its role in the curriculum, it can refer to teaching one or more subjects through the second language (Cenoz & Ruiz de Zarobe, 2015). In CLIL, second language competence is an essential tool for learning the subject matter (content) while the first language plays no or only a very subordinate role (Mattheoudakis, Alexiou & Laskaridou, 2014).

CLIL programs mainly vary in terms of using different subject matters as the framework of language learning. The most frequent subject matters used in CLIL include accountancy, agriculture, biology, business, creative art, economics, engineering, English literature, geography, history, hotel management, math,

religion, science and social science. Reviewing different studies on the effectiveness of CLIL, it seems that science and math are the most popular contents in such programs respectively (Gallardo-del-Puerto & Lacabex 2013, 2016; Korpela, 2013; Kubes, 2012; ; Luprichova, 2013; Menzlova, 2012; Moghadam & Fatemipour, 2014; Olsson, 2015). At a second level, geography and history are the most practical subject matters in CLIL curriculum (Dallinger, Jonkmann, Hollm & Fiege, 2015; Gallardo-del-Puerto & Lacabex, 2016; Mattheoudakis, Alexiou, & Laskaridou, 2014; Korpela, 2013; Sylvén & Ohlander, 2015).

2.4. Purpose of the study

Although many studies have been conducted to test the effectiveness of CLIL/CBI on language skills and components at different levels of education, the evidence they present is circumstantial and rather inconclusive; hence, despite the abundance of empirical studies, which have tested the effect of this educational intervention, policy makers cannot make informed decisions as to whether to maintain this mode of practice or not; hence, this meta-analysis aims to systematically synthesize the findings of empirical studies regarding the effect of CLIL/CBI as the educational intervention on language skills and components. More specifically, this metaanalysis aims at: (1) estimating the weighted average effect, or what this study will refer to as combined effect size; and (2) explore the dispersion of effect sizes through sub-group or moderator analysis. These objectives can be achieved by More specifically, this study aims at answering the following questions:

1. What is the overall combined effect of CLIL on language skills and components?
2. How much does the effect of CLIL change according to three different subgroups of language skill, subject matter, and educational level?

3. To what extent the overall combined effect is conditioned by publication bias?

3. Method

Meta-analysis is a systematic method of gathering the results of several independent research studies which are carried out on the same subject but in different places and times. A meta-analysis uses a statistical approach to combine the results from multiple studies in an effort to increase power (over individual studies), to improve estimates of the size of the effect and/or to resolve uncertainty when reports disagree (Cohen, Manion & Morrison, 2007; Ergene, 1999; Hunter & Schmidt, 1990).

3.1. Sampling procedure (Study selection)

According to Little, Corcoran and Pilla (2008), participants of the study are all the participants of the related previous studies which test the effect of an educational intervention; hence, the focus in sampling procedure is mainly on the selection of relevant studies and materials rather than the selection of participants. To come up with a representative sample of the studies that tested the effect of CLIL/CBI on language proficiency, the study followed three steps:

1. ***Protocol registration/ Information sources*:** Relevant studies were identified through searching academic databases including Google Scholar, the Educational Resources Information Center (ERIC), Science Direct, Proquest and Elsevier. The search covered experimental studies including articles, master theses, and doctoral dissertations. The main reason for inclusion of master and Ph.D. dissertations is to reach a comprehensive sample of studies that address the domain of interest and also to reduce the probability of publication bias. In this phase of study, using the search keywords such as 'CLIL', 'content and

language integrated learning' and limiting the search to key words, titles and abstracts, all the studies including the terms were recorded. At this stage, a total number of 55 studies was identified to be included in the meta-analysis.

2. Investigations and selection of the studies were carried out in the form of a two-phase screening process. In the first phase, heading and abstract of studies were screened and in the second phase, the full text of papers were screened. In addition, a general search was done over the references of all studies included in order to detect further related published studies.

3. ***Eligibility criteria:*** In addition, the inclusion criteria were established as the final filter for the selection of the best studies to be included in this metaanalysis. Taking the inclusion criteria into account, we included quasiexperimental and experimental studies that: (a) investigated the effectiveness of CLIL in the recent decade; (b) used CLIL as the experimental condition; and (c) reported the sample size and the statistical information necessary to calculate effect sizes. Only 22 of these studies met the eligibility criteria to be included in this meta-analysis. The descriptive statistics related to the studies included in this meta-analysis are provided in Table 1.

Table1. *Features of the studies included in the meta-analysis*

Characteristic								**Total**
Publication year		***2007***	***2008***	***2009***	***2010***	***2011***	***2012***	
	N	-	-	1	2	-	1	
	%	0	0	4.5	9	0	4.5	
		2013	***2014***	***2015***	***2016***	***2017***	***2018***	
	N	6	4	4	1	3	-	N= 22
	%	27.5	18.2	18.2	4.5	13.6	0	% =100

Type of study		***Master thesis***	***Doctoral thesis***	***Article***	
					N=22
	N	5	3	14	%=100
		22.8	13.6		
%				63.6	

3.2. Data collection and coding

Selected studies were coded based on a data coding procedure in a Microsoft Excel file. The data coding form was created prior to the statistical analyses and the coding process was done according to this coding form. Data forms capture identifying information on studies, descriptions of interventions, sample characteristics, research methods, outcome measures, and the raw data and statistical information needed to calculate effect sizes (Littel, Corcoran & Pillai, 2008).

The coding form used in this study is made up of the data items including research code, name of the study, author, year of publication, country, type of the study, i.e., article, master thesis, Ph.D. dissertation, sample size, study design, subject matter or content, language skill or component, age, gender, language level, educational level, data collection tool, contact hours, number of participants in experimental and control group, experimental and control groups pre/ post test mean scores and standard deviations.

To determine the reliability of the coding system and to avoid study selection that is biased by a coder exercising her personal judgment, it is important to use a systematic and standardized approach to the evaluation of studies. Ideally, coding studies should be made by two coders performing the coding process independently in order to gain the inter-rater reliability. In this study, the coding process has been done by two independent coders. Both of the coders were M.A. TEFL candidates at English language department of Shahrood University of

Technology. They coded the whole sample of the studies. Besides, Cohen's kappa reliability coefficient between the coders was calculated using SPSS software. Cohen's kappa reliability index was determined to be 0.91 which shows agreement higher than 91% between the two raters. This figure shows almost a perfect consistency between the coders. Finally, the disagreement between the two coding forms was checked and corrected.

3.3. Data analysis

Individual effect sizes and the combined effect size

Having collected the data from studies and coded them based on pertinent features, the effect size of each individual study was calculated. Effect sizes were calculated by Hedge's g in this study. Different scales have been presented for interpreting the calculated effect size. Cohen, Manion and Morrison (2007) interpret the effect size as follows:

- $0 \leq$ Effect size value ≤ 0.20 insignificant,
- $0.21 \leq$ Effect size value ≤ 0.50 small,
- $0.51 \leq$ Effect size value ≤ 0.8 medium,
- $0.81 \leq$ Effect size values, large effect size;

The scale presented by Plonsky and Oswald (2014) yields a different interpretation:

- $0.0 \leq$ Effect size value < 0.40 insignificant,
- $0.4 \leq$ Effect size value < 0.70 small,
- $0.7 \leq$ Effect size value < 0.1 medium,

$0.1 \leq$ Effect size values, large effect size;

If studies reported the effect of CLIL on different groups, more than one effect size was calculated for these studies. Finally, the weighted average effect size or the

combined effect size was calculated to see whether CLIL has a significant positive effect or not

Choice of model

The meta-analyst should choose between the fixed effects model (SEM) and the random effects model (REM) (Hedges & Olkin, 1985). The former assumes that the difference in effect sizes is only due to sampling error and that the true effect size is the same in all studies, and the combined effect is our estimate of this common effect size. In other words, it is assumed that there is no heterogeneity or that it is negligible. Conversely, the latter assumes the existence of heterogeneity. It is obvious that the set of populations being studied are naturally heterogamous.

Taking this inherent heterogeneity into account, this meta-analysis used the REM.

Publication bias evaluation

According to publication bias notion which is also called lost data, research on a specific subject is partially published. This is because studies that do not have statistically significant relationships or those that have low significant relationships are not considered worth enough to be published (Borenstein et al., 2009); hence, defining and evaluating publication bias is a vital and necessary step which indicates the presence of bias in the sample of effects. In this meta-analysis, the funnel plot is used to visually represent the existence of publication bias in our study. Then, classical fail-safe N test is conducted to estimate the number of lost studies with non-significant results and average zero effect size needed to nullify the calculated combined effect size in this study.

Sub-group analysis

Although meta-analysis is now increasingly used as a tool to find out whether the combined effect size of an educational intervention is statistically significant or not,

it can equally be used to explore the dispersion of effect sizes related to subgroups through what is commonly known as moderator or subgroup analysis, which is planned in line with the objective and procedure of the study (Littel, Corcoran & Pillai, 2008); hence, in addition to estimating the combined size effect, this study investigates how variables such as language skills and components, subject matter and educational level moderate the effect of CLIL.

4. Results

Figure 1, the main outcome of this meta-analysis, is a forest plot, which graphically shows the treatment effect of each individual study coupled with an estimate of the overall or combined effect size and associated confidence interval. The point estimate of each study is represented by a box, the size of which represents the study's weight in the generation of the meta-analysis. The whiskers through the boxes show the length of the confidence interval (CI). The length of the line shows the precision of the study. The longer the line, the less precise the results of the study are. The vertical line in the middle of the graph shows the line of no effect.

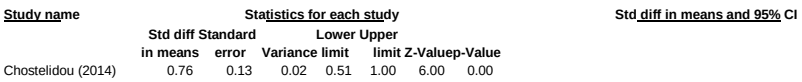
As the forest plot vividly depicts, a great majority of confidence intervals are entirely on the right side of the line. These studies show that CLIL has a significantly positive effect. Some other confidence intervals are situated entirely to the left side of the line of no effect. These studies show that CLIL has a negative effect. Finally, there are very few confidence intervals that cross the line of no effect and as such they show that the effect of CLIL is not statistically significant.

The bottom row of the forest plot summarizes and combines the effect of individual studies into the weighted average effect or the combined effect and turns the plot into a meta-analysis. The middle of the diamond in the bottom row shows the overall or the combined effect of the meta-analysis. Since the diamond is far to the right of the line of no effect, the combined effect is statistically significant.

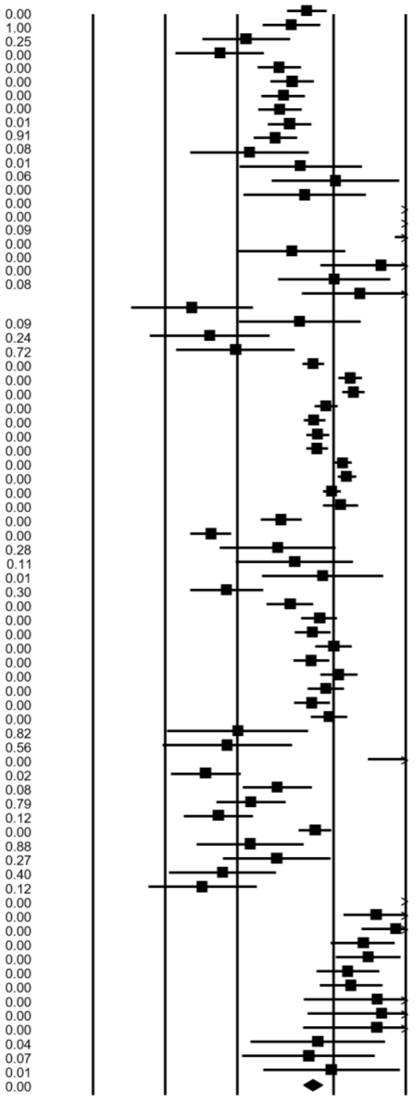
Table 2 summarizes the results of the forest plot numerically. Since p=0.00 ≤0.05, the combined size effect of 0.81 shows that CLIL has a significant positive effect on overall language proficiency.

Table 2. Random effect model statistic

		Effect size and 95% confidence interval					**Test of null (2-tail)**		**Heterogenity**			
Model	Number studies	Point Estimate	Standard Error	Variance	Lower Limit	Upper Limit	Z Value	P value	Q Value	Df	P Value	I Squared
Random	76	0.81	0.06	0.00	0.69	0.93	13.65	0.00	871.98	75	0.00	91.39



Zarepour(2014) 0.56 0.18 0.03 0.20 0.92 3.07 Mattheoudakis(2014-a) 0.00
0.28 0.08 -0.55 0.55 0.00 Mattheoudakis(2014-b) -0.33 0.28 0.08 -0.88 0.23
-1.16 Olsson(2015-a) 0.41 0.14 0.02 0.14 0.68 2.96
Olsson(2015-b) 0.58 0.14 0.02 0.30 0.85 4.13
Olsson(2015-c) 0.46 0.14 0.02 0.19 0.73 3.34
Olsson(2015-d) 0.42 0.14 0.02 0.15 0.69 3.01
Olsson(2015-e) 0.54 0.14 0.02 0.27 0.82 3.90
Olsson(2015-f) 0.36 0.14 0.02 0.09
0.63 2.62 Gallardo(2013-a) 0.04 0.38
0.14 -0.70 0.78 0.11
Gallardo(2013-b) 0.68 0.39 0.15 -0.08 1.44 1.75
Gallardo(2013-c) 1.12 0.41 0.17 0.32
1.91 2.75 Gallardo(2013-d) 0.73 0.39
0.15 -0.03 1.50 1.88 Berendse(2013-
a) 6.25 1.01 1.03 4.27 8.24
6.17
Berendse(2013-b) 5.05 0.70 0.49 3.68 6.43 7.20
Berendse(2013-c) 3.06 0.62 0.38 1.85
4.27 4.96 Gallardo(2016-a) 0.57 0.34
0.12 -0.10 1.24 1.68 Gallardo(2016-
b) 1.68 0.39 0.15 0.92 2.44
4.33
Gallardo(2016-c) 1.10 0.36 0.13 0.40 1.80 3.06
Gallardo(2016-d) 1.42 0.37 0.14 0.69 2.15 3.81
Gallardo(2016-e) -0.67 0.39 0.15 -1.43
0.09 -1.74 Gallardo(2016-f) 0.67 0.39
0.15 -0.09 1.43
1.72
Gallardo(2016-g) -0.45 0.38 0.15 -1.20 0.30 -1.18
Gallardo(2016-h) -0.14 0.38 0.14 -0.88 0.60 -0.36
Dallinger(2015-a) 0.83 0.07 0.00 0.70 0.97 12.17
Dallinger(2015-b) 1.30 0.08 0.01 1.15 1.45 16.89
Dallinger(2015-c) 1.34 0.07 0.01 1.20 1.48 18.41
Dallinger(2015-d) 1.00 0.07 0.01 0.85 1.14 13.45
Dallinger(2015-e) 0.85 0.07 0.00 0.71 0.98 12.34
Dallinger(2015-f) 0.89 0.07 0.01 0.75 1.03 12.16
Dallinger(2015-g) 0.88 0.07 0.00 0.75
1.02 12.84 Lorenzo(2009-a) 1.20 0.06
0.00 1.08 1.31 20.57
Lorenzo(2009-b) 1.25 0.06 0.00 1.14 1.37 20.82
Lorenzo(2009-c) 1.07 0.06 0.00 0.95 1.18 18.45 Lorenzo(2009-d) 1.18
0.11 0.01 0.96 1.40 10.42
Aguilar(2014-a) 0.44 0.13 0.02 0.18 0.70 3.31
Aguilar(2014-b) -0.44 0.13 0.02 -0.70
-0.18 -3.31 Korpela(2013-a) 0.40 0.37
0.14 -0.33 1.12 1.07
Korpela(2013-b) 0.60 0.37 0.14 -0.13 1.33 1.61
Korpela(2013-c) 0.96 0.39 0.15 0.20 1.71 2.48
Austad(2013) -0.24 0.23 0.05 -0.70
0.22 -1.03 Sylvén(2015) 0.55 0.15
0.02 0.26 0.84 3.67 Menzlova(2013-
a) 0.91 0.12 0.01 0.68 1.14
7.87
Menzlova(2013-b) 0.83 0.11 0.01 0.60 1.05 7.20
Menzlova(2013-c) 1.09 0.12 0.01 0.86 1.32 9.28
Menzlova(2013-d) 0.81 0.11 0.01 0.59 1.04 7.09
Menzlova(2013-e) 1.15 0.12 0.01 0.92 1.39 9.72
Menzlova(2013-f) 1.00 0.12 0.01 0.77 1.22 8.53
Menzlova(2013-g) 0.82 0.11 0.01 0.59 1.04 7.11
Menzlova(2013-h) 1.03 0.12 0.01 0.80 1.26 8.81
Kubes(2012-a) -0.10 0.45 0.20 -0.98 0.78 -0.23
Kubes(2012-b) -0.24 0.41 0.17 -1.04 0.57 -0.58
Luprichova(2013) 2.27 0.39 0.15 1.52
3.03 5.87 Kovacikova(2013-a) -0.50 0.22
0.05 -0.94 -0.07 -2.28
Kovacikova(2013-b) 0.39 0.22 0.05
-0.04 0.82 1.76
Kovacikova(2013-c) 0.06 0.22 0.05 -0.37 0.49 0.27 Kovacikova(2013-d) -0.34 0.22 0.05 -0.78 0.09 -1.56 Lahuerta Martínez(2017) 0.86 0.10 0.01
0.65 1.06 8.21 Gallardo(2015-a) 0.05 0.34 0.12 -0.62 0.71 0.14
Gallardo(2015-b) 0.38 0.34 0.12 -0.29 1.05 1.11
Gallardo(2015-c) -0.29 0.34 0.12 -0.96 0.38 -0.85
Gallardo(2015-d) -0.54 0.35 0.12 -1.22 0.13 -1.57
Kothuri(2017) 5.29 0.39 0.15 4.53 6.05 13.66
Mäkinen(2010-a) 1.62 0.21 0.04 1.21 2.03 7.73
Mäkinen(2010-b) 1.86 0.22 0.05 1.44 2.29 8.55
Mäkinen(2010-c) 1.46 0.20 0.04 1.06 1.86 7.11
Mäkinen(2010-d) 1.52 0.21 0.04 1.11 1.92 7.35
Mäkinen(2010-e) 1.27 0.20 0.04 0.88 1.66 6.35 Mäkinen(2010-f) 1.31
0.20 0.04 0.91 1.70 6.52
Diéguez(2017-a) 1.64 0.47 0.22 0.72 2.56 3.49
Diéguez(2017-b) 1.69 0.47 0.22 0.76 2.62 3.57
Diéguez(2017-c) 1.63 0.47 0.22 0.71 2.55 3.47
Diéguez(2017-d) 0.89 0.43 0.18 0.05 1.73 2.09
Diéguez(2017-e) 0.78 0.42 0.18 -0.05 1.61 1.84
Diéguez(2017-f) 1.06 0.43 0.19 0.21 1.91 2.44
0.81 0.06 0.00 0.70 0.94 13.72



-2.00 -1.00 0.00 1.00 2.00

Favours A Favours B

Figure1. Forest Plot of the Effect Sizes

Mean Effect Size= 0.81 std= 0.06 Variance= 0.00 I^2 = 91.39
p value= 0.00 Q value= 871.78 Z value= 13.72

The first research question addresses the effectiveness of CLIL or what is known as CBI in North America and Canada, by combining the effect sizes from a total of 22 primary studies. As shown in Table 2, the overall combined effect of 0.81 represents a large effect size on Cohen's (1987) scale but a medium effect size with respect to Plonsky and Oswald's (2014) scale. All in all, the combined effect size of 0.81 shows that CLIL learners scored 0.81 standard deviations above the non-CLIL participants on the outcome measures.

Moderator analysis

In addition to estimating the overall effect of CLIL on language skills and components, this meta-analysis was concerned with locating any systematic differences in the effectiveness of CLIL at different levels of education, across different skills and components, and across different subject matters. Exploring the dispersion of effect sizes through moderator analysis is critically important in our case since the I^2 = 91.39 is much bigger than 75%. Taking I^2 which suggests a high level of heterogeneity into account, the *p*-value and the combined effect size should be treated cautiously and the meta-analyst should focus on the dispersion of true effect sizes through sub-group or moderator analysis. Table3 shows the results of sub-group or moderator analysis and the dispersion of effect sizes.

Educational level

Table 3 clearly shows that education level distinctly moderates the overall combined effect size. While based on the overall combined effect size, one may conclude that CLIL affects students of different educational levels indiscriminately, the effect sizes of 1.02, 0.81, and 0.40, for primary, secondary and university levels respectively, vividly show that CLIL differentially affects learners at different level of education. While it is most effective in primary school, it is less effective among university students.

Skills and components

Based on the overall combined effect size, which shows that CLIL has a significant positive effect on language skills and components, policy makers may come up with the large-scale implementation of CLIL as an effective intervention. Subgroup analysis and the dispersion of effect sizes show that skills and components moderate the overall effect significantly. The effect size of 1.32 shows that CLIL is mostly effective in developing learners' grammar. Conversely, with an effect size of 0.00, CLIL is found to be the least effective in teaching pronunciation. Based on Plonsky and Oswald's (2014) scale, CLIL has a medium effect in teaching writing, vocabulary, reading and listening.

Subject matter

While the overall effect size may lead the reader to conclude that CLIL yields the same results irrespective of what subject matter is chosen as the content to be integrated with language, sub-group analysis shows that different subjects differentially affect language proficiency. Table 3 shows the dispersion of effect sizes based on subject matter as a moderator. As shown in Table 3, some of the subject matter effect sizes, including those related to agriculture, literature, and religion are on the negative side of the line of no effect. This shows that these subjects yield a significant negative effect. However, based on Plonsky and Oswald's (2014) scale, the effect of these subject matters on developing learners' overall proficiency, indicated by the combined effect of CLIL on skills and components, is insignificant since despite being negative, they approximate zero, which signifies the insignificant effect of these subject matters. On the other hand, effect sizes corresponding to creative arts, hotel management, social sciences, natural sciences, mathematics, history, geography, accountancy are entirely to the positive side of zero. This shows that these have a significant positive effect on developing language skills and components. The dispersion of effect sizes corresponding to subject matter shows that using hotel management with an effect size of 5.25, and creative arts with an effect size of 1.20 are most effective in developing learners' proficiency, or the combined effect of CLIL on skills and components.

Table 3. *Moderator analysis on the effectiveness of CLIL*

Sub-groups	K	variance	Z value	P value	M (g)	SE	95% confidence interval Lower	Upper
Educational Level								
Primary	26	0.00	16.00	0.00	1.02	0.06	0.89	1.14
Secondary	38	0.00	10.77	0.00	0.81	0.07	0.66	0.95
University	12	0.07	1.44	0.14	0.40	0.28	-0.14	0.95
Skill								
General	10	0.01	6.62	0.00	0.76	0.11	0.54	0.99
Grammar	17	0.07	4.86	0.00	1.32	0.27	0.78	1.85
Listening	8	0.01	7.93	0.00	0.91	0.11	0.69	1.14
Pronunciation	3	0.12	-0.00	0.99	0.00	0.35	-0.69	0.69
Reading	6	0.02	4.58	0.00	0.72	0.15	0.41	1.03
Speaking	8	0.07	1.27	0.20	0.35	0.28	-0.19	0.90
Vocabulary	20	0.01	6.57	0.00	0.90	0.13	0.63	1.17
Writing	4	0.02	5.61	0.00	0.81	0.14	0.53	1.10
Subject matter								
Accountancy	1	0.01	6.00	0.00	0.75	0.12	0.50	1.00
Agriculture	4	0.04	-0.49	0.62	-0.09	0.20	-0.49	0.29
Biology	13	0.02	3.21	0.00	0.54	0.17	0.21	0.88
Business	6	0.00	8.14	0.00	0.45	0.05	0.34	0.57
Creative art	6	0.03	6.81	0.00	1.20	0.17	0.85	1.55
Economics	6	0.00	8.14	0.00	0.45	0.05	0.34	0.57
Engineering	2	0.18	0.00	1.00	0.00	0.43	-0.84	0.84
English literature	4	0.03	-0.49	0.62	-0.09	0.19	-0.48	0.28
Geography	22	0.00	9.08	0.00	0.74	0.08	0.58	0.90
History	24	0.00	11.08	0.00	0.80	0.07	0.66	0.94
Hotel Management	1	0.14	13.66	0.00	5.25	0.38	4.50	6.00
Math	14	0.00	12.95	0.00	0.85	0.06	0.72	0.98
Religion	1	0.05	-1.03	0.30	-0.23	0.23	-0.69	0.21
Science	31	0.00	10.21	0.00	0.71	0.07	0.57	0.84
Social science	7	0.10	3.03	0.00	0.98	0.32	0.34	1.61

Publication bias evaluation

In order to answer question three, the funnel plot was used to visually represent the existence of publication bias in our study. If the plot forms an almost symmetric funnel, it shows no potential publication bias. On the other hand, if effect sizes of the studies show a relatively asymmetrical distribution around the main effect size, the studies seem to have publication bias.

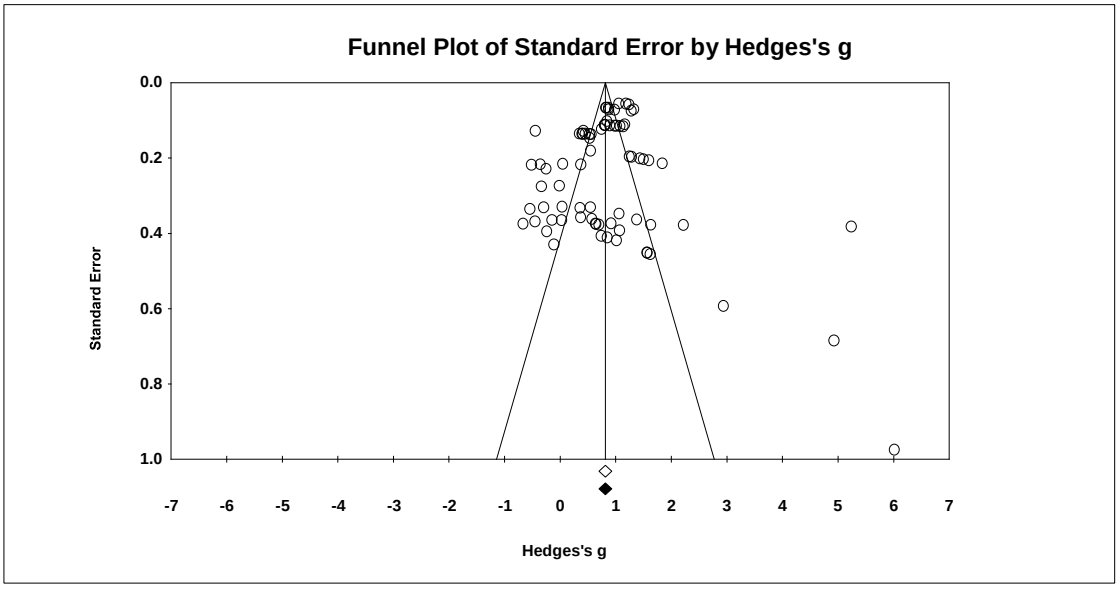


Figure 2. Funnel plot on observed studies

As it is shown in Figure 2, the funnel plot forms almost asymmetric distribution, therefore, it shows some potential publication bias. However, the result of evaluation by this method is subjective; hence, this study used a second method, i.e., Fail-safe N test to avoid any subjective interpretation. As mentioned before, classical fail-safe N test was conducted to estimate the number of lost studies with non-significant results and average zero effect size needed to nullify the calculated combined effect size in this study. As shown in Table 4, an additional 9605 lost studies with an average zero effect size would be needed in order to nullify the effect size. Overall, these results indicated that publication bias could not explain the significant positive outcomes detected across all studies.

Table 4. *Results of the classic fail-safe N*

Classic Fail Safe N	
Z value for observed studies	44.78
P value for observed studies	0.00
Alpha	0.05
Tails	2.00
Z for alpha	1.95
Number of observed studies	76

Number of missing studies that would bring p value to > alpha	9605

5. Discussion and Conclusion

The result section addressed the three research questions posed in this metaanalysis. The first question was answered with the overall combined effect of 0.81, which represents a large effect size on Cohen's (1987) scale but a medium effect size with respect to Plonsky and Oswald' (2014) scale. However, Hedges (2008) believes that combined effect sizes such as the one estimated in this meta-analysis are best interpreted when compared with other overall combined effect sizes.

Several meta-analyses have explored the effect of bilingual programs on learners' academic achievement in the United States. To start with, Willing (1985) synthesized the results of 23 primary studies and found an overall combined effect size of 0.33 and based on this point estimate concluded that participation in bilingual programs is favored in preparing learners for tests of reading, language skills, mathematics and achievement when the tests were in English. Similarly, Rolstad, Mahoney and Glass (2008) synthesized the results of 17 studies and found an overall combined effect size of 0.23 and based on this size effect concluded that bilingual education is superior to all-English programs. Finally, Krashen and McField (2005) synthesized the previous empirical findings and found the overall combined effect size of 0.26. Although in line with Plonsky and Oswald' (2014) scale the overall combined effect of 0.81, is medium in size, once compared with the combined effect sizes of 0.33, 0.23, and 0.26, as reported by previous metaanalyses in the same domain, the combined overall effect of CLIL on language skills and components is vividly large.

Although the magnitude of the overall combined effect is large, it should be interpreted cautiously since the $I^2 = 91.39$ is suggestive of a high level of heterogeneity. Borenstein et al. (2009) suggest that I^2 be used as a criterion to decide whether moderator analysis is needed or not. As he suggests, when I^2 is high, then a moderator analysis should be undertaken to explore the dispersion of effect sizes. In this study, I^2 is 91.39%, hence, sub-group analysis was undertaken to see how the specified moderators moderate the effectiveness of CLIL.

Table 3 answers the second question by showing how educational level, skills and components, and subject matter moderate the effectiveness of CLIL. The disparity

of the effect sizes which reflects how moderators affect the overall effectiveness of CLIL is a better basis for deciding on the large-scale implementation or replacement of CLIL as an effective educational intervention since, in drastic contrast with the overall combined effect which shows the effectiveness of CLIL, subgroup analysis shows that the effect of CLIL can have varied effects. At times, it has a significant positive effect as in the case of primary education, grammar, and when the subject is creative arts or hotel management, since they show large effect sizes based on Plonsky and Oswald's (2014). In other cases, it has a significant negative effect as in the case of agriculture, English literature and religion. In rare cases, e.g., engineering with an effect size of 0.00, it happens to have an insignificant effect.

Along these lines, CLIL has a maximum effect at primary level (1.02), a moderate effect at secondary level (0.81) and a minimum effect (0.40) at university level. It is crystal clear that as the students' educational level and age increases, the impact of CLIL on their language proficiency decreases. The students seem to be more successful when they are in CLIL environment from early ages and primary levels of school. This leads to the conclusion that younger students in lower educational levels are under the positive effect of CLIL on their second language more than those in higher levels. Therefore, implementation of this method in lower bilingual educations is prior to its implementations at higher levels of education; hence, the biggest part of financial budget, energy and time should be spent on implementing CLIL at lower educational levels.

Moreover, hotel management was the subject matter that had the largest effect size among other subjects (5.25). However, since only one study took hotel management as subject matter, this figure should be interpreted cautiously. Accountancy and religion as the content of CLIL program, yielded a negligible effect size . However, creative arts, social science, math, history, geography, science and biology yielded the largest effect size among all the studies.

Taking skills and components as a sub group, it was found that CLIL had the largest positive effect on grammar (1.32), a large positive effect on vocabulary (0.90) and a no effect on students' pronunciation component (0.00); therefore, while language education program can take CLIL as an effective intervention in developing learners' grammar and vocabulary, they should not generalize this effectiveness to teaching pronunciation and as such find alternative strategies for

teaching language pronunciation. In addition, among the four language skills, CLIL had the most effect on students' listening skill and the least effect on their speaking skill. This conclusion is in line with Lasagabaster (2008) who reported that the receptive skills such as reading and listening are under the positive influence of CLIL more than the productive skills such as writing and speaking in various European countries.

Although meta-analysis is increasingly used as a tool for testing the effectiveness of educational interventions, the discrepancy between the effect sizes presented in Table 3 and the overall combined effect suggest that meta-analysts are much better off if they use it to explore the dispersion of effect sizes and make more informed decisions on the basis of this dispersion; hence, it is suggested that: (1) meta-analyst undertake moderator analysis if I^2 shows a high level of heterogeneity as in the current meta-analysis; (2) policy makers base their decisions on the combined effect size of an educational intervention if the studies covered in meta-analysis are homogeneous and decide on its implementation after a careful consideration of moderating variables.

Moreover, since a forest plot, the main outcome of any meta-analysis, graphically presents very useful information including estimates of the effect size of each study, the corresponding confidence interval, the precision of each study and the overall combined effect, it is essential that meta-analyst not ignore it in reporting the findings of their meta-analysis. Despite its vividly significant role in clarifying and summarizing the findings of any meta-analysis, only a few metaanalysts make use of this powerful tool.

Finally, taking the results of this meta-analysis into account, this study has clear implications for:

- The curriculum developers of bilingual education institutes, since they enable them to decide in line with the overall combined effect size of CLIL and its differential effects in the light of moderator analysis rather than decide based on contradictory empirical findings presented by individual studies.
- Practitioners, since not only do they give them insight into the differential effect of CLIL but also enable them to make more precise lesson plans before implementing CLIL in their classrooms.

- Policy makers since they help them make informed decisions based the combined effect of a large number of empirical studies rather than decide based on individual studies which present circumstantial and inconclusive evidence.

6. Limitations of the Study

Although the study is rigorous in design, like any studies, whether quantitative or qualitative, this study has its own limitations. Since some studies were unavailable to access because of limitations of some online databases, the selected studies were only gathered from free open-access databases and some rich studies may be missing. Many studies which were worth to be included in this research had to be excluded because they didn't present all the statistical information required for this meta-analysis research. The number of studies done on the effect of CLIL on students' language pronunciation and writing skills is too few. Thus, more experimental studies are needed to be done in this field in order to come to a conclusive conclusion about these areas. In the case of content, some subject matters such as biology, geography, history, math and science have been frequently used in different CLIL programs while many other content areas have rarely been used; hence, there is not enough information about the effect of CLIL on students' language learning when language is integrated with these subject matters as the content.

7. Suggestion for Future Research

This meta-analysis aimed at presenting the overall combined effect size of CLIL on language skills and components coupled with its differential effect at different levels of education, language skills and components and across different subject matters. The completeness and precision of this meta-analysis consists in the precision of the primary studies. As shown in the forest plot, the size of the boxes situated in line with effect sizes reflect the weight of each study in estimating the overall combined effect, and whiskers which go through the boxes depicts the lengths of the confidence interval. The longer the lines, the less precise the findings of the study are. The length of the whiskers shows that a large number of primary studies included in this meta-analysis are not precise enough; therefore, it is suggested that researchers who are interested in testing the effect on CLIL or any other educational interventions assure the methodological rigor of their study and be more meticulous

in reporting descriptive statistics needed to aggregate and compare findings from different studies. A large number of studies are excluded because they do not report statistics such as standard deviations, sample sizes, means, reliability indexes, effect sizes and confidence intervals. Moreover, there is a paucity of studies undertaken to test the effectiveness of CLIL on students' language pronunciation and writing skills. Thus, more experimental studies are needed to be undertaken in these areas. Finally, to reduce the high level of heterogeneity which characterized this meta-analysis, future studies should further specify the dependent variable or the outcome. This caveat will enable them to synthesize the findings of studies which present more homogeneous effect sizes.

8. References

Ackerl, C. (2007). Lexico-Grammar in the essays of CLIL and non-CLIL students: Error analysis of written production. *Vienna English Working Papers, 16*(3), 6–11.

*Aguilar, M., & Muñoz, C. (2014). The effect of proficiency on CLIL benefits in engineering students in Spain. *International Journal of Applied Linguistics, 24*(1), 1-18.

Alonso, E., Grisalena, J. & Campo, A. (2008). Plurilingual education in secondary schools: Analysis of results. *International CLIL Research Journal, 1*(1), 36-49.

*Austad, N. K. (2013). *Vocabulary Testing in CLIL: The Effect of incidental vocabulary learning in CLIL on the vocabulary of learners* (Master's thesis). Universitas Osloensis.

*Berendse, E. P. H. (2014). *Acquiring L2 English prepositions in an L1 Dutch environment: The effect of immersion through CLIL teaching* (Master's thesis). Utrecht University.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* London, UK: Wiley.

Bret-Blasco, A. (2011). *Implementing CLIL in a primary school in Spain: The effects of CLIL on L2 English learners' oral production skills.* Mémoire, Universitat Autònoma de Barcelona, Barcelona.

Cámara Ortiz, C. (2014). *The impact of adopting a CLIL approach on EFL learners reading skills in Catalonia school.* Barcelona: Universitat Central de Catalunya.

Cenoz, J., & Ruiz de Zarobe, Y. (2015). Learning through a second or additional language: Content-based instruction and CLIL in the twenty-first century. *Language, Culture and Curriculum, 28*(1), 1–7.

*Chostelidoua, D., & Grivab, E. (2014). Measuring the effect of implementing CLIL in higher education: An experimental research project. *Social and Behavioral Sciences, 116*, 2169–2174.

Cohen, J. (1987). *Statistical Power Analysis for the Behavioral Sciences*. Hillside, NJ: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education (6th Edition).* New York: Routledge.

Crandall, J. (1998). The expanding role of the elementary ESL teacher: Doing more than teaching language. *ESL Magazine, 1*(4), 10-14.

*Dallinger, S., Jonkmann, K., Hollm, J., & Fiege, C. (2015). The effect of content and language integrated learning on students' English and history competences–Killing two birds with one stone?. *Learning and instruction*, *41*, 23-31.

Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms*. Amsterdam: John Benjamins.

Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): Current research from Europe. In W. Delanoy & L. Volkmann (Eds.) *Future Perspectives for English Language Teaching (*pp. 139-157). Heidelberg: Carl Winter.

*Diéguez, K. I., & Adrián, M. (2017). The influence of CLIL on receptive vocabulary: A preliminary study. *Journal of English Studies*, *15*, 107-134.

Ergene, T. (1999). *Effectiveness of test anxiety reduction programs: A metaanalysis review* (PhD Thesis). Ohio: Ohio University.

*Gallardo del Puerto, F., & Gomez Lacabex, E. (2013). The impact of additional CLIL exposure on oral English production. *Journal of English Studies*, *11,* 113-131.

*Gallardo del Puerto, F., & Adrián, M. (2015). The use of oral presentations in higher education: CLIL vs. English as a foreign language. *Revista de Educación, 38*, 73-106.

*Gallardo del Puerto, F., & Lacabex, E. G. (2016). Oral production outcomes in CLIL: An attempt to manage amount of exposure. *European Journal of Applied Linguistics, 5*(1), 31-54.

Georgiou, S. I. (2012). Reviewing the puzzle of CLIL. *ELT Journal*, 66(4), 495–504

Hedges , L. , & Olkin , I. (1985). *Statistical models for meta-analysis* . New York : Academic Press.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. London: Sage Publications.

Jäppinen, A. K. (2005). Thinking and content learning of mathematics and science as cognitional development in content and language integrated learning (CLIL): Teaching through a foreign language in Finland. *Language and Education*, *19*(2), 147-168.

Juan-Garau, M. (2010). Oral fluency development in secondary education CLIL learners. *Vienna English Working Papers*, 19(3), 42-48.

Kjellén-Simes, M. (2009). Content-based language learning in English: A model for high proficiency in English in Swedish schools. In S. Granath, B. Bihl & E. Wenno (Eds.). *Pathways to Language and Literature* (pp. 103114). Karlstad: Karlstad Universitet.

*Korpela, L. (2013). *Learning English grammar in content and language integrated learning: Comparing the grammatical proficiency of CLIL students and students receiving mainstream EFL instruction* (Pro Gradu thesis). University of Helsinki.

*Kothuri, S., & Nageswari, R. (2017). CLIL to develop hotel management (HTM) learners' vocabulary; Ongoing research. *Man in India, 97*(2), 745751.

*Kovacikova, E. (2013). *Modernization of teaching English as a foreign language by means of CLIL methodology in higher vocational education* (Doctoral dissertation, PhD thesis). Nitra: Univerzita Konštantína Filozofa.

Krashen, S., & McField, G. (2005). What works? Reviewing the latest evidence on bilingual education. *Language Learner, 1*(2), 7–10.

*Kubes, M. (2012). *Aplikácia prístupu CLIL na hodinách matematiky v 4. ročníku ZŠ* (Doctoral dissertation, Doctoral thesis). Bratislava: Univerzita Komenského.

*Lahuerta Martínez, A. C. (2017). Analysis of the effect of CLIL programs on the written competence of secondary education students. *Revista de Filología*, *35*, 169-184.

Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, 1, 3041.

Littel, H. J., Corcoran, J. & Pillai, V. (2008). *Systematic reviews and metaanalysis*. NY: Oxford University Press.

*Lorenzo, F., Casal, S., & Moore, P. (2009). 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.

*Luprichova, J. (2013). *Modernization of teaching English as a foreign language by means of CLIL methodology* (Doctoral dissertation, Doctoral Thesis). Nitra: Constantine the Philosopher University.

Ma, Y. (2003). Communicative language teaching in China. *Journal of Suzhou Education Institute, 7*(1), 101-102.

*Mäkinen, M. (2010). *Yesterday I met the biggest ice cream and drinked the biggest coke of my life": are there differences in the competence of English grammar between CLIL students and non-CLIL students?* (Pro Gradu thesis), University of Jyväskylä.

Marsh, D. (2000). An introduction to CLIL for parents and young people, In D. Marsh & G. Langé (Eds.). *Using languages to learn and learning to use languages* (pp. 1-14). Jyväskylá, Finland.

*Mattheoudakis, M., Alexiou, T., & Laskaridou, C. (2014). To CLIL or not to CLIL? The case of the 3rd experimental primary school in Evosmos. In N. Lavidas, T. Alexiou, & A. Sougari (Eds.), *Selected papers from the 20th International Symposium of Theoretical and Applied Linguistics* (pp. 215234). Versita Publications.

*Menzlova, B. (2012). Obsahovo a jazykovo integrované vyučovanie (CLIL) na 1. stupni základnej školy. In S. Pokrivčáková *et al*. (Eds.),

Obsahovo a jazykovo integrované vyučovanie (CLIL) v ISCED 1 (pp. 1360). Bratislava: ŠPÚ

Met, M. (1999). *Content-based instruction: Defi ning terms, making decisions* (NFLC Reports). Washington, DC: The National Foreign Language Center.

*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.

Navés, T. (2009). Effective CLIL programs, In Ruiz de Zarobe, Y. & Jimenez Catalan, R.M. (Eds.), *CLIL: evidence from research in Europe*. Bristol: Multilingual Matters.

Olaizola, I. V., & Mayo, M. P. G. (2009). Tense and agreement morphology in the interlanguage of Basque/Spanish bilinguals: CLIL verus non-CLIL. In R. Zarobe Y. & Catalan R. M. J. (Eds.), Content *and language integrated learning: Evidence from research in Europe* (pp.157-175). Buffalo: Multilingual Matter.

*Olsson, E. (2015). Progress in English academic vocabulary use in writing among CLIL and non-CLIL students in Sweden. *Moderna språk*, *109*(2), 51-74.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878-912.

Rolstad, K, Mahoney, K., & Glass, J. V. (2008). The big picture in bilingual education: A Meta-analysis corrected for Gersten's coding error. Journal of Educational Research and Policy Studies, 8(2), 1-15.

Ruiz de Zarobe, Y. (2008). CLIL and foreign language learning: A longitudinal study in the Basque country. *International CLIL Research Journal*, 1(1), 60-73.

Serra, C. (2007). Assessing CLIL at primary school: A longitudinal study. *International Journal of Bilingual Education and Bilingualism*, 10(5), 582602.

Snow, M. A., Met, M., & Genesee, F. (1989). A conceptual framework for the integration of language and content in second/foreign language instruction. *TESOL Quarterly*, *23*(2), 201-217.

*Sylvén, L. K., & Ohlander, S. (2015). The CLISS project: Receptive vocabulary in CLIL versus non-CLIL groups. *Moderna språk*, *108*(2), 80114.

Tedick, D., & Cammarata, L. (2012). Content and language integration in K-12 contexts: Student outcomes, teacher practices, and stakeholder perspectives. *Foreign Language Annals, 45*(1), 28–S53.

Willing, A. C. (1985). Meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research,* 53(3), 269-317.

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*.

Notes on Contributors:

Seyyed Ali Ostovar-Namaghi is currently a full-time associate professor of TEFL at the department of applied linguistics, Shahrood University of Technology (SUT), Iran. He teaches both graduate and undergraduate courses including language teaching methodology, research methodology, materials development, and EAP. His chief research interest is language teacher education, grounded theory, and theories of practice. He has published in a number of leading peerreviewed journals. He is also a member the editorial board of some journals in applied linguistics and language teaching.

Shiva Nakhaee received her MATEFL from Shahrood University of Technology. She has been teaching language skills at private language schools of Shahrood in the past few years.