

In the Name of God
The Beneficent, the Merciful



English Language Department
M.A. Thesis in Language Teaching

Content and Language Integrated Learning: a Meta-Analysis

By: Shiva Nakhaei

Supervisor:

Dr. Seyyed Ali Ostovar Namaghi

August 2017

شماره: ۲۴۱۹۷/۱۲۸
تاریخ: ۹۷/۷/۱

باسمه تعالی



مدیریت تحصیلات تکمیلی

فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم شیوا نخعی با شماره

دانشجویی ۹۴۱۸۰۰۴ رشته آموزش زبان انگلیسی تحت عنوان **Content and Language Learning: a Meta-analysis** که در تاریخ ۹۷/۰۶/۱۶ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می گردد:

قبول (با درجه: ۱۹) مردود
نوع تحقیق: نظری عملی

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استاد راهنمای اول	سید علی التکلی	رئیس	
۲- استاد راهنمای دوم			
۳- استاد مشاور			
۴- نماینده تحصیلات تکمیلی	زودتر زودتر	استاد	
۵- استاد ممتحن اول	ارباب	استاد	
۶- استاد ممتحن دوم	نظم مستند	استاد	



بمرد: در صورتی که کسی مردود نبود حداکثر یکبار دیگر (در مدت مجاز محصل) می تواند از پایان نامه خود دفاع نماید (دفاع مجدد نباید زودتر از ۴ ماه برگزار شود).

Dedication

This dissertation is dedicated to my supportive parents who have always encouraged me to succeed. I had promised to make my parents proud by the achievement of this monumental academic goal and I hope that I have fulfilled that promise. It is also dedicated to the memory of my beloved brother.

Acknowledgments

First and foremost, I would like to thank Allah, my God who has always supported me in ups and downs of my life. I would like to express my special gratitude to my supervisor Dr. Seyyed Ali Ostovar-Namaghi who gave me the golden opportunity to do this project. I came to know about so many things I am really thankful to them. Not only was he the best teacher ever for me but also a perfect inspiration of success. I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame. Finally, I would further like to acknowledge Dr. Iranmehr and Dr. Mozaffari of the English department at Shahrood University of Technology, and I am gratefully indebted to them for their very insightful comments and thoughtful feedback on this dissertation.

تعهد نامه

اینجانب شیوا نخئی دانشجوی دوره کارشناسی ارشد رشته آموزش زبان انگلیسی دانشکده مهندسی صنایع و مدیریت دانشگاه صنعتی شاهرود نویسنده پایان نامه ی

Content and Language Integrated Learning: A Meta-analysis

تحت راهنمایی جناب آقای دکتر سید علی استوار نامقی متعهد می شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود» و یا « **Shahrood University of Technology** » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ: ۱۳۹۷/۶/۱۲

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

Abstract

The research presents a meta-analysis on the effectiveness of content and language integrated learning (CLIL) on students' language proficiency. By establishing a set of inclusion/exclusion criteria, a total of 22 primary studies with 76 different effect sizes were extracted from different master's thesis, doctoral dissertation and articles in 2007–2017 and in English. Four moderator variables including language skills and components, subject matter, educational level and publication type were identified. It was found that the overall effect was $g = 0.81$ which represents a medium effect size with respect to Plonsky and Oswald' (2014) scale. As a result of moderator analysis, the highest effect sizes are in the master's theses. In addition, CLIL has the most effect on students' grammar and listening proficiency and in lower levels of education especially in elementary schools. It also has the highest effect when combining with hotel management as the content. Afterward, the study concludes with a discussion for implication and suggestion for future studies. The findings have clear implications for practitioners, researchers and curriculum developers.

Keywords

Content and language integrated learning (CLIL), English proficiency, meta-analysis, effect size, publication bias

Table of Contents

Title	Page
Dedication.....	I
Acknowledgments.....	II
Abstract.....	IV
Chapter One: Introduction.....	1
1.1. Overview.....	2
1.2. Statement of the Problems.....	3
1.3. Purpose of the Study.....	3
1.4. Limitations of the Study.....	3
1.5. Delimitations of the Study.....	4
Chapter Two: Literature Review.....	5
2.1. Overview.....	6
2.2. Theoretical Perspectives.....	6
2.2.1. Definition of CLIL.....	7
2.2.2. The 4Cs Framework.....	10
2.2.3. CLIL, Content-based Instruction and Immersion.....	11
2.2.4. Fundamental Theories behind CLIL.....	13
2.2.5. Problems with CLIL and its Implementation.....	16
2.2.6. Language Proficiency.....	17
2.3. Empirical Findings.....	18
2.3.1. CLIL and Different Language Skills.....	18
2.3.2. CLIL and Educational Levels.....	20
2.4. Summary of Empirical Findings and Statement of the Gap.....	21
Chapter Three: Research Method.....	24
3.1. Overview.....	25
3.2. The philosophy of Meta-Analysis.....	25
3.3. Sampling Procedure and Materials.....	26
3.4. Data Collection.....	28
3.5. Reliability of the Study.....	30
3.6. Data Analysis.....	30

3.6.1. Individual Effect Sizes and the Main Effect Size.....	30
3.6.2. Test of Homogeneity.....	32
3.6.3. Moderator Analysis (Sub-group Analysis).....	33
3.6.4. Publication Bias Evaluation.....	34
3.6.4.1. Funnel Scatter Plot.....	35
3.6.4.2. Trimm and Fill Method.....	35
3.6.4.3. Fail-Safe N Test.....	35
Chapter Four: Results.....	36
4.1. Individual and Main Effect Size Analyses.....	37
4.2. Moderator Analysis (Sub-group Analysis).....	39
4.2.1. Educational Level.....	41
4.2.2. Skill.....	41
4.2.3. Publication Type.....	42
4.2.4. Subject matter.....	43
4.3. Publication Bias Evaluation.....	44
4.3.1. Funnel Scatter Plot.....	45
4.3.2. Trimm and Fill Method.....	46
4.3.3. Classic Fail-safe N.....	47
Chapter Five: Discussion and Conclusion.....	47
5.1. Overview.....	48
5.2. Discussion and Conclusion.....	48
5.3. Implications for Practice.....	52
5.4. Suggestion for Future Research.....	52
References.....	54

List of Tables

Table 3.1. Features of the Studies Included in the Meta-Analysis.....	28
Table 4.1. Fixed and random effect model statistic.....	37
Table 4.2. Moderator analysis on the effectiveness of CLIL.....	39
Table 4.3. The result of Trim and Fill analysis.....	45
Table 4.4. Results of the classic fail-safe N.....	46

List of Figures

Figure 4.1. Forest plot of the effect sizes.....	38
Figure 4.2. Comparison of effect sizes between different educational levels.....	40
Figure 4.3. Comparison of effect sizes between different language skills and components..	41
Figure 4.4. Comparison of effect sizes between different publication types.....	41
Figure 4.5. Comparison of effect sizes between different subject matters.....	43
Figure 4.6. Funnel plot on observed studies.....	44
Figure 4.7. Funnel plot on observed and imputed studies.....	45

CHAPTER ONE: INTRODUCTION

1.1 Overview

Taking a brief look at the history of language teaching and learning, there was a big shift away from explicit language teaching to communicative language teaching. At the beginning of the introduction of teaching methods, especially emersion of the Audiolingual method, the main focus of teaching and learning in language classes was on form. In this era, linguistic competence was at the heart of curriculum process. Later in 1970s, language teachers and learners were suspicious about the role of grammar. During that time when the focus was on the introduction of descriptive grammar, it was found that there is a great number of native speakers, who speak fluently and accurately without knowing grammatical aspects of the language; hence, it was concluded that learning a language is totally different from learning its grammar and that the focus should be shifted from form to meaning. In addition, the advocators of communicative approach added that language functions and notions should be paid more attention than its grammar.

In 1980s, Brumfit reemphasized the primary function of grammar in language courses and mentioned that the syllabus should be multidimensional and needs a grammatical core and notions, functions and situations should revolve around this core (Finocchiaro & Brumfit, 1983). In fact, in an effective language classroom, both form and meaning of language should be focused on. If learners focus only on the form of the language, they will become linguistically competent. They can speak accurately but not fluently. On the other hand, if they focus only on meaning, they will become communicatively competent. They can speak fluently but they also make lots of grammatical mistakes. So, the best way is the integration of form and meaning. Language teaching is about enabling students to master forms and then to use those forms in communication. Only in this situation can learners become linguistically and communicatively competent. Taking this notion into account Del Hyme's (1972) famous quote "There are rules of use without which the rules of grammar will be useless" is better to be corrected in "There are rules of use without which the rules of grammar will be useless. And also, there are rules of grammar without which the rules

of use will be useless.” This fact has led to the introduction of CLIL (content and language integrated learning in which, the focus is on both form and meaning of the language.

1.2. Statement of the Problems

The increasing degree of requirement to an approach which combines the main aspects of language all together and provides a more natural efficient environment led to the introduction of CLIL by some scholars which aims to integrate form and meaning of language in teaching (e.g., Marsh, 2000; Mattheoudakis, Alexiou, & Laskaridou, 2014; Moghadam & Fatemipour, 2014). There is an increasing number of studies which have been conducted in relation to content and language integrated learning (CLIL) and its impact on different aspects of students’ proficiency. But, the problem is that the results are inconclusive. Different studies have come to different conclusions about the role of CLIL in teaching and learning a second language. These differences come from limitations that the studies may face with, such as sample size, time and other conditions; therefore, this negatively affects the generalizability of the findings.

1.3. Purpose of the Study

This meta-analysis quantitatively aims at combining the findings from the previous studies in the field of the effect of using CLIL on the proficiency of students at different grade levels in order to reach an agreement that is, the main effect size of this intervention on students’ proficiency. More specifically, this study aims at addressing the following questions:

1. What is the main effect of CLIL on students’ proficiency?
2. To what extent does the effect of CLIL on students’ proficiency vary according to different moderator variables?
- 3- Is there any publication bias in this meta-analysis? If so, what is the level of bias in this meta-analysis?

1.4. 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 are not included in the study because of this reason. 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 them. 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 the other ones have been paid little attention and 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. The final limitation is that the meta-analysis study combines the results of different studies that their experimental conditions are different from each other because they did not conduct the experiments using the same instructions and classroom procedures. This causes a high level of heterogeneity in the sample of this meta-analysis.

1.5. Delimitations of the Study

In order to narrow down the problem and define the boundaries of the research, experimental studies should have the following characteristics for being chosen. These characteristics are addressed under the title of inclusion criteria. That is, studies which did not meet these criteria were excluded. The study came up with a manageable but generalizable sample by setting the following inclusion criteria:

- Studies that investigated the effectiveness of CLIL published between the years 2007 and 2017. The latest date for the researches included in this study was set as September 2017;
- Quasi-experimental and experimental studies that used CLIL in the experiment groups are included in this meta-analysis study;
- Studies that do not report statistical information necessary to calculate effect sizes are not included in the analysis;
- Studies have to be written in English;
- Studies should give the sample size of the studied groups;
- Studies that provide sufficient information for calculating the effect size.

CHAPTER TWO:
LITERATURE REVIEW

2.1. Overview

This chapter presents the relevant literature of the study into two main sections of theoretical perspectives and empirical findings. In the first section, theories underlying Content and Language Integrated Learning (CLIL) and related frameworks will be discussed. The next section mainly deals with empirical findings of previously published works from the areas relevant to the present study.

2.2. Theoretical Perspectives

During ages, there has been a big effort to change the way of teaching language from traditional inefficient teacher-centered methods to more modern fruitful student-centered ones (Duckworth, 2009; Dupin-Bryant, 2004). One of the big problems with methods during method era was that each of them only focused on one aspect of language and learning and they ignored all other possible aspects. For instance, some of them only focus on the form of language and some other only focus on the meaning of language (Kumaravadivelu, 1994).

Considering these deficits, The increasing degree of requirement to an approach which combines the main aspects of language all together and provides a more natural efficient environment led to the introduction of a new innovative approach in terms of CLIL by some scholars which aims to integrate form and meaning of language in teaching (e.g., Marsh, 2000; Mattheoudakis, Alexiou, & Laskaridou, 2014; Moghadam & Fatemipour, 2014).

Reactions to the Communicative Approach and the Natural Approach led to the introduction and development of a new foreign language teaching approach called CLIL which is inspired by the success of Canadian immersion programs. It started to develop in Europe in the 1990s because of the need for higher level of foreign language proficiency by students (De Graaff *et al.*, 2007).

CLIL is somehow similar to the Communicative Approach as they both emphasize on the idea that proficiency and fluency in foreign language can be achieved

better by using it as a “functional medium of communication and information” rather than “making it the object of analysis in class” (De Graaff *et al*, 2007, p. 606). The most important difference between CLIL and “teaching the subject in the first language” is the fact that CLIL involves “additional language learning objectives” and “specific opportunities for communication and language use” (De Graaff *et al*, 2007, p. 606).

2.2.1. Definition of CLIL

CLIL is a term created in 1994 by David Marsh. He defined that CLIL is a situation in which, a school subject is taught through a foreign language and also that foreign language is taught within a specific subject. So, in this method the focus is on both language and content (Marsh, 2000; Moghadam & Fatemipour, 2014; Nikula, Dalton-Puffer, & García, 2013). CLIL is a long-term learning 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). Also, it aims at developing proficiency in both subject matter and language by teaching the subject matter not *in*, but *with* and *through* the foreign language (Harrop, 2012). CLIL models range from theme-based language modules to cross-curricular approaches in which, a subject is taught through the foreign language. Cross-curricular approach toward CLIL model has become the most frequent in Europe in the last few years (Harrop, 2012).

CLIL has multiple objectives including socio-cultural, socioeconomic, linguistic and educational purposes (Mattheoudakis, Alexiou & Laskaridou, 2014). Socio-cultural objective aims to make the learners more familiar with other cultures and increase their knowledge, tolerance and respect toward other cultures. However, socioeconomic objective aims at increasing learners’ recruitment in international societies. In addition, linguistic objective tries to develop students’ different language skills to have effective communications in different contexts. And finally, educational objective aims to help students develop their content and subject matter knowledge (Mattheoudakis, Alexiou & Laskaridou, 2014).

CLIL provides the opportunities for students to learn a subject matter through a foreign language and to learn a foreign language by studying a subject matter.

Therefore, 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). This approach can be used at different educational levels in preschool, primary school, secondary school and higher education. 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). In addition, students acquire language in real-life and natural situations rather than learning it. Thus, the process of second language development is similar to the process of native language acquisition in which, students mostly learn L2 implicitly, informally, incidentally and not through the explicit language instructions (Mattheoudakis, Alexiou & Laskaridou, 2014). So, fluency and ability to communicate is paid more attention than accuracy.

Additionally, since content and language are integrated with cognition and culture in this approach, it can also promote thinking skills, cultural awareness and intercultural communication skills (Mattheoudakis, Alexiou, & Laskaridou, 2014; Pistorio, 2010). According to Halbach (2009), integration of content and language in acquisition needs double cognitive effort than non-CLIL learners. It leads to a more positive effect on learner's learning and mental activities (de Diezmas, 2016; Halbach, 2009). Thus, as Dalton-Puffer (2008, p. 143) states "rather than being a hindrance, L2 processing actually has a strong potential for the learning of subject-specific concepts" (de Diezmas, 2016). Moreover, by integrating content and language together, this approach helps learners improve and apply their critical thinking to learn, integrate, apply and transfer the knowledge they acquire (Mattheoudakis, Alexiou & Laskaridou, 2014; Pistorio, 2010). As Muñoz (2002, p. 36) mentioned that "CLIL may strengthen learners' ability to process input, which prepares them for higher level thinking skills, and enhances cognitive development" (Muñoz, 2002, p. 36).

Also, considering cognitive skills, a range of functional strategic skills seem to develop in the CLIL courses. It means, there is “a move from an automatic to a deliberate level of analysis and action” (Greene, Pearson & Schoenfeld, 1999, p. 145).

Furthermore, by getting students participate in conversation activities and expecting them to use appropriate responses, this approach can improve learners’ interactive skills (Morgan, 2006). It also provides a suitable environment for cooperative learning by getting students work in pairs and groups (Jacobs & McCafferty 2006; Mattheoudakis, Alexiou & Laskaridou, 2014; Pistorio 2010).

On the other hand, Mehisto *et al.* (2008) believes that providing learners with learning skills that support integrated learning is essential in an environment where content and language are connected together. Therefore, the purpose of CLIL is not only content and language learning, but also it aims at developing learning skills (de Diezmas, 2016; Mehisto, Marsh, & Frigols, 2008).

All in all, improvement of metalinguistic skills, mental flexibility, increasing use of strategies and expansion of vocabulary range are other advantages of this approach (Baker, 2001; Bialystok, 2002; Cook, 1992; Dörnyei, 1995; Johnson & Swain, 1997).

According to Massler, Stotz, and Queisser (2014), there are three types of CLIL including type A, type B and type C lessons. Type A which also includes immersion is based on subject lessons. It is applied when the purposes of learning are defined based on the content of academic subject matter which is taught through a foreign language. In this type, assessment is based on the content of subject matter and not the foreign language. In another hand, type B is based on language lessons and content of academic subjects is used in language class. Here, the purposes of learning are defined based on the foreign language which is learned in a specific subject context. In this type, assessment is based on the foreign language and not the content of subject matter. Finally, type C lessons which is considered as a full integration of content and language. This is a rare phenomena and very difficult to implement (Cenoz & Ruiz de Zarobe, 2015; Massler, Stotz & Queisser, 2014).

This approach also provides an analysis of linguistic demands learners may face with. To this end, Coyle (2007) provided a model of linguistic progression in 3 phases:

- language of learning (needed to access basic concepts in a given context)
- language for learning (language needed to operate and interact with the content in a given context)
- language through learning (incidental language that results from active involvement with the task)

CLIL claims that it makes all language needed for successful knowledge acquisition transparent and accessible in a way that is not always found in content subjects (Coyle, 1999; Gajo, 2007; Harrop, 2012).

2.2.2. The 4Cs Framework

CLIL is based on a theoretical framework called 4Cs model which is a holistic approach and helps develop the rationale for introducing CLIL into the curriculum. There are four dimensions in this approach including content (such as subject matter, themes, cross-curricular approaches), communication (language learning and using), cognition (thinking and learning process) and culture (developing intercultural understanding and global citizenship, awareness of self and otherness) which are integrated together. So, in CLIL, the focus is on the interrelationship between these four dimensions (Coyle, 2005, Coyle, 2008; Coyle *et al.*, 2010).

According to Coyle *et al.* (2010) In order to implement CLIL effectively, five things should be fulfilled:

- progression in knowledge, skills and understanding of content
- engagement in associated cognitive processing
- interaction in the communicative context
- development of appropriate communication (language knowledge) skills
- acquisition of a deepening intercultural awareness through the positioning of self and ‘otherness’

2.2.3. CLIL, Content-based Instruction and Immersion

CLIL has some differences and similarities with language immersion and content-based instruction (Snow, Met, & Genesee, 1989; Spanos, 1990; Tarnopolsky, 2013). All the three methods are similar together on the basis of integrated language (ESP) learning in which, the focus has shifted from the target language to the integration of language and the non-linguistic disciplines. Hence, the focus is on both the language for professional communication and the subject matter of that communication (Snow, Met, & Genesee, 1989; Spanos, 1990; Tarnopolsky, 2013). They are different from Traditional ESP Teaching in which, the main focus is only on linguistic features and teaching language (Robinson, 1991; Tarnopolsky, 2013). On the other hand, although CLIL, content-based instruction and language immersion are subparts of integrated language (ESP) learning approach, they have some subtle differences.

Content- based instruction (CBI) provides a context in order to enable students to acquire a target language through using it in learning a contextualized subject matter such as history, physics, math and etc rather than learning it out of context (Brinton *et al.*, 1989; Tarnopolsky, 2013; Tedick, 2012). Thus, content learning leads to language learning while the language mastery facilitates students' content learning in their vocational or occupational areas (Stoller, 2008). Since the students' focus is mainly on the non-linguistic content presented in the target language, the language and its communicative skills are taught implicitly and learned subconsciously without focusing on its form which provides a favorable environment for language acquisition (Krashen, 1985). The important point is that CBI refers to language course. This language curriculum can be based on one or more non-linguistic subject matters (Tarnopolsky, 2013). It is so difficult to distinguish between CLIL and CBI (Brinton, Snow, & Wesche, 1989; Coyle, 2007; Marsh, 2000; Stoller, 2008; Tarnopolsky, 2013). They are usually considered as two different words of the same reality (Coyle *et al.*, 2010; Dalton-Puffer, 2008; Van de Craen, Ceuleers, & Mondt, 2007). CLIL is usually regarded as the European version of CBI in that, CBI is frequently used and more popular in the US and Canada (Cenoz & Ruiz de Zarobe, 2015; Ruiz de Zarobe, 2008; Tarnopolsky, 2013). Cenoz *et al.* (2014) concluded that, 'although the origins of CLIL in Europe might make it historically unique, this does not necessarily make it

pedagogically unique' (Cenoz *et al.*, 2014, p.244). In addition, Paran (2013) believes that the differences between CLIL and CBI are contextual rather than defining differences. Although these two approaches seem to share a lot in common, CLIL sounds broader in scope than CBI because CLIL is designed for both adult training and language teaching at secondary schools. However, CBI is mostly and specifically oriented at ESP and adult education (Tarnopolsky, 2013).

Language immersion is a second language teaching method through which, students educate at schools in their L2. Improving bilingualism is the main purpose of this program (Calvé, 1991; Johnson & Swain, 1997; Rehorick & Edwards, 1994; Walker & Tedick, 2000). The difference between content-based instruction and immersion is that the former includes language-based courses while the latter includes content-based courses taught in the target language (Tarnopolsky, 2013). Developing learner's communicative competence and cognitive advantages of bilingualism are other purposes of this program (Johnson & Swain, 1997). Immersion program is different from one country to another according to differences between historical antecedents, language policy or public opinion. In addition, such programs have different formats based on class time spent in L2 (complete or partial immersion), participation by native speaking (L1) students, learner age (from early to adult immersion), the subject matter and so on (Calvé, 1991; Johnson & Swain, 1997; Rehorick & Edwards, 1994; Walker & Tedick, 2000). Lasagabaster and Sierra (2009) made a comparison between CLIL and language immersion and believe that these two approaches have some differences in language of instruction, teachers, starting age, teaching materials, language objectives and immigrant students (Lasagabaster and Sierra, 2009). The language of instruction in CLIL is foreign language and students only contact with that in formal situations and contexts however, the language of instruction in immersion program is second language which is spoken locally in both formal and informal contexts outside of the school. In addition, teaching materials in language immersion are usually designed by native speakers who have a great command of the language of instruction. However, this is not usually in CLIL programs. In the case of starting age, a vast number of immersion programs are the early immersion type, while CLIL approach is more similar to late immersion in which, adult students can start learning the second languages in their maturity. Taking language objectives into account in both approaches, immersion

program has a long term objective in order to enable students master the second language similar to the native speaker acquisition process, whereas CLIL does not have such long term objectives. In addition, the results of some research have shown that immigrant students prefer to take part in immersion programs rather than CLIL ones (Lasagabaster & Sierra, 2009).

The difference between CLIL, CBI and language immersion can be better understood by presenting the continuum of content and language integration (Cenoz & Ruiz de Zarobe, 2015; Met, 1999; Tarnopolsky, 2013). The continuum ranges from the most language-driven end which is highly used in traditional language classrooms to the most data-driven end which shows immersion program. CBI is based on language courses. Thus, it tends to be nearer to the language-driven end. Showing the best integration of language and content in the course, a perfect CLIL program should be in the middle of the continuum (Cenoz & Ruiz de Zarobe, 2015; Met, 1999; Tarnopolsky, 2013).

2.2.4. Fundamental Theories behind CLIL

CLIL is based on some fundamental second language acquisition and learning theories including comprehensible input hypothesis (Krashen, 1985), output hypothesis (Swain, 1985), interaction theory (Long, 1996), theory of Basic interpersonal communication skills (BICS) and Cognitive academic language proficiency (CALP) (Cummins, 1992), cognitive constructivist theory (Piaget, 1963) and social constructivist theory (Vygotsky, 1978).

According to comprehensible input hypothesis, it is necessary for learners to be exposed to second language comprehensible input which is a little higher than their current level of proficiency (e.g., Doughty & Williams, 1998; Krashen, 1985; Lyster, 1998; Moghadam & Fatemipour, 2014). This is a favorable condition in which learners acquire the second language implicitly and are able to produce language subconsciously and spontaneously (Krashen, 1985). Later, he was criticized on the basis that the importance of focusing on the form is ignored in his theory (Doughty & Williams, 1998). Taking this for granted and focusing on both language and content, CLIL tries to provide both a meaningful context for students to receive comprehensible input and

form-focused explanation of the salient features of language presented in that input at the same time (Lyster, 1998).

Output hypothesis is another theory which CLIL is based on (Mattheoudakis, Alexiou, & Laskaridou, 2014; Swain, 1985). In this theory, Swain (1985) emphasized on the fact that only exposing students to comprehensible input is not enough. It is essential but not sufficient. In fact, teachers should provide opportunities to push students to produce comprehensible outputs in spoken or written forms and focusing on their accuracy in production of language. CLIL, on the other hand is a suitable framework in which, not only students receive comprehensible content and language inputs, but also they are encouraged to produce both appropriate and accurate language based on their knowledge of the subject matter and language taught in the classroom (Mattheoudakis, Alexiou, & Laskaridou, 2014).

Interaction hypothesis can be another aspect of CLIL (Long, 1996; Mattheoudakis, Alexiou, & Laskaridou, 2014). In this theory like comprehensible input hypothesis, Long (1996) mentioned that comprehensible input is necessary for language acquisition. In addition, students communicate in a condition that one's comprehensible output could be the other one's comprehensible input. Furthermore, comprehensible input could be more effective for the learners when they try to negotiate the meaning. It means when the learner does not understand something, he tries to overcome this problem by using some communicative strategies such as asking for clarification, simplification, modification, paraphrasing, explanation and so on. CLIL, on the other hand provides such a condition in which interaction and negotiation of meaning happens among learners or between learners and teachers which enable students to understand both language-related and content-related information presented in the classroom (Mattheoudakis, Alexiou, & Laskaridou, 2014).

Another fundamental basis of CLIL is in terms of CALP and BICS conceptualization of language proficiency (Cummins, 1992; Mattheoudakis, Alexiou, & Laskaridou, 2014; Moghadam & Fatemipour, 2014). In this theory, Cummins (1992) introduced two major parts of language proficiency including Basic interpersonal communication skills (BICS) and Cognitive academic language proficiency (CALP).

BICS is the communicative capacity that all children acquire to be able to function in daily interpersonal exchanges that is language use. It is also called context-embedded communication which refers to the context-bound face-to-face communicative activities. On the other hand, CALP refers to formal academic learning which refers to deeper-level language proficiency that is necessary for dealing with more abstract, academic situations involving the development of literacy skills. These two components require different language and cognitive processes. Learning BICS proceeds learning CALP because the latter requires more cognitive endeavors than the former. CLIL on the other hand provides opportunities for gaining proficiency in both aspects by using academic textbooks and exercises and involving students in communication simultaneously based on these academic contents (Mattheoudakis, Alexiou, & Laskaridou, 2014; Moghadam & Fatemipour, 2014).

Cognitive constructivist theory is another supportive framework for the CLIL approach (Moghadam & Fatemipour, 2014; Piaget, 1963). This theory is based on the notion that learning is making a meaningful connection between what a learner is learning and what he has been learned until now (Piaget, 1963). CLIL is based on this notion and CLIL teachers try to design and teach the materials in the way that enable students to make a connection between the new knowledge and their prior knowledge to build their individual construction of both language and content knowledge. The stronger the connection, the easier and greater learning and internalization occur. On the contrary, the more the language and content are taught in isolation, the less effective process of learning will be (Moghadam & Fatemipour, 2014).

Social constructivism is another source of inspiration for CLIL (Mattheoudakis, Alexiou, & Laskaridou, 2014; Moghadam & Fatemipour, 2014; Vygotski, 1978). Vygotski (1978) believed that social interaction and purposeful meaningful functioning in social context through language is necessary for language acquisition and learning. The processes of mediation, conscious and meaningful imitation, private speech and internalization are central to learning. He proposed his mediation theory which is learning through the mediation of a more competent peer or adult that is called a mediator. The mediator should help and scaffold the learner to facilitate and accelerate his learning and also, he helps the learner go through his zone of proximal development

(ZDP) which is the distance between a learner's actual cognitive capacity and his level of potential development. He mentioned that a child can improve his actual cognition to his potential development through interaction with an adult who is cognitively more developed compared to him. As a result of such interactions and the consequent linguistic development their cognition will develop too. CLIL provides a framework in which scaffolding through interaction occurs. Teacher and learners interact in the classroom upon content and classroom activities. They try to help each other and push each other to upper levels of proficiency by group work and pair work or getting help from teacher (Moghadam & Fatemipour, 2014; Van de Craen, Ceuleers & Mondt, 2007).

The common idea behind all the above theories and CLIL approach is the requirement of meaningful contextualized linguistic input without any formal teaching (e.g., Marsh, 2000; Mattheoudakis, Alexiou, & Laskaridou, 2014; Moghadam & Fatemipour, 2014). Therefore, students should be exposed to contextualized comprehensible foreign language inputs in a content-oriented, discourse-based language teaching environment such as CLIL (Moghadam & Fatemipour, 2014). Generally, CLIL framework should include its multiple focus on language, learning and cognition, the construction of safe and enriching learning environments, the use of authentic materials and interactions, the promotion of active learning, the use of macro- and micro-scaffolding in students' learning so as to enhance their autonomy and the promotion of co-operation among students and teachers (Hammond, 2001).

2.2.5. Problems with CLIL and its Implementation

There are some criticisms over CLIL due to the difficulties in the implementation of CLIL programs (Harrop, 2012; Mattheoudakis, Alexiou, & Laskaridou, 2014). The first problem is the shortage of appropriate teaching materials, insufficient pre- or in-service teacher training programs and the lack of professionally trained CLIL instructors. The problem is that it is not clear between a language teacher and a subject teacher, which one is responsible for teaching in CLIL environment. According to Mehisto, Marsh and Frigols (2008) the question is whether it is the subject teacher who will be teaching their subject in another language or the language teacher who will be teaching a subject unrelated to his/her profession in the foreign language is a related issue of concern.

They offered a solution and stated collaboration and networking between language teachers, subject teachers, universities and organizations is necessary.

Parental reluctance is another problem. The second problem with CLIL is the attitudes of most parents that believe the intensive exposure to the foreign language in education may endanger students' mother tongue and may cause first language loss (Mattheoudakis, Alexiou, & Laskaridou, 2014). In another hand, parents worry about their children's subject matter achievement because they suppose that non-CLIL learners will outperform their children since they study all subjects in their 11 so they understand the subject better than their children (Mattheoudakis, Alexiou, & Laskaridou, 2014). However, reseach in the field of comparing CLIL and non-CLIL learner's content learning not only does not support their claim but also the findings of such studies have shown that it can facilitate content acquisition in comparison with non-CLIL programs (Mattheoudakis, Alexiou, & Laskaridou, 2014)..

2.2.6. Language Proficiency

Language proficiency can be defined as the ability of a speaker to use language skills successfully and accurately and to master its arts and functions of language required in special contexts (Hawkins, 2004). Lado (1961) defined the construct of language ability. He mentioned "language is a complex system of communication with various levels of complexity involving intricate selection and ordering of meanings, sounds, and longer units." (Lado, 1961, p. 2). On the contrary, Oller (1976) believed that language is an integrative and unitary phenomenon because it is a mixture of different skills and components. Later, focusing on sociolinguistics and its relation to different language components, Canale and Swain (1980) presented the concept of communicative competence as a necessary part of language proficiency. Bachman and Palmer (2010) presented a more comprehensive model for language proficiency by focusing on the interaction of language usage and its use, strategic competence and the metacognitive strategies. Betty and Boris (2004) defined speaking, writing, reading, and listening skills as the different levels of language proficiency. In addition, Lee and Schallert (1997)

mentioned that language proficiency is related to learners' ability of reading, speaking, listening and writing accurately and fluently in different situations and occasions. Furthermore, language proficiency refers to an individual's general knowledge of a language including vocabulary, grammar, and discourse conventions which may be called upon during any instance of oral or written language use (Peregoy & Boyle, 2008).

In this meta-analysis the four skills of speaking, listening, reading, writing and the four components of vocabulary, grammar and pronunciation are considered as different elements of language proficiency.

2.3. Empirical Findings

2.3.1. CLIL and Different Language Skills

Many researchers which have already conducted some research on CLIL only focused on one component or skill of the target language (e.g., Ackerl, 2007; Bret-Blasco, 2011; Cámara-Ortiz, 2014; Chostelidoua & Grivab, 2014; Dalton-Puffer, 2008; Juan-Garau, 2010; Kjellén-Simes, 2009; Lasagabaster, 2008; Ruiz de Zarobe, 2008; Serra, 2007). Lasagabaster (2008) 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. One reason may be because of the student's exposure to the spoken and written forms of target language.

Many studies have been done to investigate the effect of CLIL on reading skills proficiency (e.g., Cámara-Ortiz, 2014; Chostelidoua & Grivab, 2014). For example, Chostelidoua and Grivab (2014) founded that students in CLIL group performed better than students in EFL group in both reading skills and content knowledge showed a considerably higher positive attitude towards the CLIL course than their peers in the control group. Cámara-Ortiz (2014) also reported that this approach may have a positive effect on the learners' reading skills, as the results obtained by the CLIL learners in both

reading comprehension tests, were better than the ones obtained by the students who were not involved in this method.

Some research in CLIL studies focused on investigating the effect of this method on developing writing skills proficiency and they came to different contradictory conclusions (e.g., Ackerl, 2007; Dalton-Puffer, 2008; Kjellén-Simes, 2009; Lasagabaster, 2008). For instance, Dalton-Puffer (2008) suggests that CLIL doesn't seem to have a significant impact on learners' writing skills in particular. However, other investigators such as Kjellén-Simes (2009) in Sweden, Ackerl (2007) in Vienna and Lasagabaster (2008) concluded that this approach has a positive impact on the development of adolescent learners' writing skills.

Some other studies in the field of CLIL took the effect of this method on learner's oral skills proficiency into the consideration (e.g., Bret-Blasco, 2011; Dalton-Puffer, 2008; Juan-Garau, 2010; Ruiz de Zarobe, 2008; Serra, 2007). For example, Bret-Blasco (2011) reported that the results reported in this study seem to indicate that CLIL exposure may have a positive effect on students' oral performance, as the language they produce was generally more complex, more accurate and more fluent than that of EFL learners. Some researchers found that CLIL learners speak more fluently than learners who exclusively benefit from EFL education. Their use of second language is creative, they are eager to take risks and experience the language in meaningful and challenging ways (Dalton-Puffer, 2008; Juan-Garau, 2010; Ruiz de Zarobe, 2008; Serra, 2007).

A number of studies have been conducted to find out the impact of using this method for developing learner's vocabulary (e.g., Dalton-Puffer, 2008; Juan-Garau, 2010; Moghadam & Fatemipour, 2014; Olaizola & Mayo, 2009; Xanthou, 2010). Olaizola and Mayo (2009) reported that CLIL learners have better mastery of some morphological elements of the language. Juan-Garau (2010) also found that CLIL participants are seen to produce a higher number of words per minute and their pauses are shorter. Furthermore, Moghadam and Fatemipour (2014) reported that students of CLIL schools have the ability to develop and retain vocabulary better than ordinary school students because of this method and textbooks which they are taught for science and mathematics subjects. Some scholars found that CLIL has a positive impact on both

receptive and productive vocabulary (e.g., Dalton-Puffer, 2008; Xanthou, 2010). Mattheoudakis, Alexiou and Laskaridou (2014) indicated that research in the impact of CLIL on grammar, and especially syntax, on the other hand, has yielded mixed results and further research is needed in this area.

A number of studies tried to take the content aspect of CLIL into consideration and they did some experimental research to see how much using that approach has a positive impact on learning a subject matter (e.g., Jäppinen, 2005; Serra, 2007; Xanthou, 2011). For example, Jäppinen (2005) conducted a study in Finland and concluded that CLIL has a positive effect on learning math and science. Xanthou (2011) examined the effects of that on the learning of science by primary Cypriot learners, and found that science learning was positively affected by this approach. Serra (2007) also found that CLIL has some positive effect of learning math in Swiss.

2.3.2. CLIL and Educational Levels

CLIL can be implemented in different types of school. Considering learner's age, it can be implemented from kindergarten to secondary level of education. In early ages of CLIL that it was recently introduced, most of the literature related to this approach focused on secondary schools and less attention was drawn to pre- primary and primary levels of education (Austad, 2013; Berendse, 2014; Dallinger *et al.*, 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). Almost all results of the studies show the positive effect of CLIL in primary schools and only few numbers of them reported its null or negative effect (Kubes, 2012; Mattheoudakis, Alexiou & Laskaridou, 2014).

In some studies, it has also been incorporated in higher educational levels such as universities (e.g. Aguilar & Munoz, 2014; Chostelidoua & Griva, 2014; Kothuri & Nageswari, 2017; Kováčiková, 2013). The results of most studies show that CLIL has a

great effect on university students and only a few of them reported that this approach does not have any special effect on students' language proficiency (Aguilar & Munoz, 2014; Gallardo del Puerto & Martínez Adrián, 2015; Kováčiková, 2013). For example, Aguilar and Munoz (2014) conducted a test and resulted that the difference between the mean scores in the pre-and post-listening test was significant but it was not for the 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 scores from the grammar and vocabulary section were significantly higher in the control group. Using t-test, he finally concluded that none of the three scores (vocabulary and grammar, reading, writing) were significantly different. Gallardo del Puerto and Martínez Adrián (2015) conducted the same study about the effect of CLIL on oral proficiency and concluded that CLIL students did not perceive that their English language skills had improved after the oral presentation training in comparison with the significant gains reported by EFL students.

2.4. Summary of Empirical Findings and Statement of the Gap

CLIL seem to have a magnificent effect on both learning a subject matter and learning the target language in general and its different skills and components in particular. It is obvious that the results are not conclusive. Different studies have come to different conclusions about the role of CLIL in teaching and learning a second language and its impact (e.g., Ackerl, 2007; Bret-Blasco, 2011; Cámara-Ortiz, 2014; Chostelidoua & Grivab, 2014; Dalton-Puffer, 2008; Jäppinen, 2005; Juan-Garau, 2010; Kjellén-Simes, 2009; Lasagabaster, 2008; Moghadam & Fatemipour, 2014; Ruiz de Zarobe, 2008; Olaizola & Mayo, 2009; Serra, 2007; Xanthou, 2010). Dalton-Puffer (2008) suggests that CLIL doesn't seem to have a significant impact on learners' writing skills. However, some other investigators concluded that CLIL has a positive impact on the development of adolescent learners' writing skills (e.g., Ackerl, 2007; Kjellén-Simes, 2009; Lasagabaster, 2008).

Considering reading skills, some researchers found that CLIL has a significant positive impact on learners' reading skills (e.g., Cámara-Ortiz, 2014; Chostelidoua & Grivab, 2014). On the other hand, Lasagabaster (2008) 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. 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).

The results also proved that CLIL students mostly outperform the non-CLIL learners in listening and reading comprehension, fluency and vocabulary, but not alot in pronunciation, accuracy and complexity of written and spoken language (Dalton-Puffer, 2007; Lasagabaster, 2008; Alonso *et al.*, 2008; Naves, 2009; Ruiz de Zarobe, 2008).

Taking vocabulary into consideration, some scholars found that CLIL has a positive impact on learner's vocabulary (e.g., Juan-Garau, 2010; Moghadam & Fatemipour, 2014; Olaizola & Mayo, 2009). Other emphasized on the positive impact of this method on both receptive and productive vocabulary (e.g., Dalton-Puffer, 2008; Xanthou, 2010). Mattheoudakis, Alexiou and Laskaridou (2014) indicated that different research on the impact of CLIL on grammar has come to different conclusions and further research is needed in this area.

From the content knowledge point of view different studies have been conducted and all of them have concluded that CLIL has a positive impact on student's content knowledge such as math and science (e.g., Jäppinen, 2005; Serra, 2007; Xanthou, 2011).

Considering educational levels, most of the literature related to this approach focused on secondary schools and less attention was drawn to pre- primary and primary levels of education (Austad, 2013; Berendse, 2014; Dallinger *et al.*, 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). Almost all results of the studies show the positive effect of CLIL in primary schools and only few numbers of them reported its null or negative effect (Kubes, 2012; Mattheoudakis,

Alexiou & Laskaridou, 2014). In some studies, it has also been incorporated in higher educational levels such as universities (e.g. Aguilar & Munoz, 2014; Chostelidoua & Griva, 2014; Kothuri & Nageswari, 2017; Kováčiková, 2013). The results of most studies show that CLIL has a great effect on university students and only a few of them reported that this approach does not have any special effect on students' language proficiency (Aguilar & Munoz, 2014; Gallardo del Puerto & Martínez Adrián, 2015; Kováčiková, 2013).

As the review vividly shows, the results of the previous empirical findings are inconclusive and at times contradictory; hence, the field is in urgent need of a rigorous meta-analysis which not only synthesizes the previous empirical findings but also sheds some light on practice by presenting the reader interested in the effectiveness of CLIL with an overall effect size.

CHAPTER THREE: RESEARCH METHOD

3.1. Overview

This chapter presents the methodology of the study by introducing the steps which are taken in order to do a meta-analysis to investigate and determine the effectiveness of CLIL approach on students' overall achievement and proficiency. To make the reader cognizant with the theory and practice of meta-analysis, what follows presents the philosophy of meta-analysis, sampling procedure and materials, data collection, reliability of the study, data analysis and publication bias evaluation.

3.2. The Philosophy of Meta-Analysis

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 time and using statistical analysis in order to synthesize those results. To put it in a nutshell, meta-analysis is the analysis of other analysis (Littel, Corcoran, & Pillai, 2008; Petticrew & Roberts, 2006). 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). According to Cooper and Rosenthal (1980), meta-analysis is used in order to avoid Type II errors by systematically synthetic of research results. It also helps gather hypotheses for further research.

This method is a great replacement for intuition which is frequently reported narratively. Narrative reviews are selective and they are not adequately comprehensive. They also usually ignore confliction among different research findings. In addition, they hardly ever take this notion into account that sampling error can play an important role in creating variations in results of studies. On the other hand, they neglect the important role of intervening variables in research. Finally, they are not replicable because the procedures for integrating the research findings are not reported explicitly (Cook *et al.* 1992).

Taking these deficits into consideration, the quantitative method of meta-analysis has been brought into the center of attention since the past few decades by Glass *et al.*

(1981) and others (e.g. Hedges & Olkin, 1985; Rosenthal, 1991) in order to substitute narrative intuition.

Glass *et al.*

. (1981) and Hunter *et al.* (1982) suggest several stages in meta-analysis procedure:

Step 1 formulation of the problem, defining an appropriate question and identifying the variables relevant to the research question including independent and dependent variables.

Step 2 searching the related literature, identifying all the studies which focus on the related variables and final selection of related studies based on inclusion criteria

Step 3 Coding each study based on characteristics that may predict outcome results and effect sizes (e.g. age of participants, gender, ethnicity, duration of the intervention, date, publication status, design characteristics, quality of design, status of researcher).

Step 4 Estimating the effect sizes for each pair of dependant and independant variables, so that results can be measured on a common scale, and then weighting the effect size of each study according to the sample size.

Step 5 Calculating the mean effect size and the standard deviation of effect sizes across the studies

Step 6 Determining the effects of sampling errors, measurement errors and range of restriction.

Step 7 If a large proportion of the variance is attributable to the issues in Step 6, then the average effect size can be considered an accurate estimate of relationships between variables.

Step 8 If a large proportion of the variance is not attributable to the issues in Step 6, then review those characteristics of interest which correlate with the study effects.

3.3. Sampling Procedure and Materials

According to what presented in Little, Corcoran and Pilla (2008), Participants of the study will be all the participants of the related previous studies that the experiment has been done on. So, the focus in sampling procedure of a meta-analysis is on the selection of relevant studies and materials that have the same problematic and the necessary statistical data analysis and of articles published in scientific journals rather than selection of participants.

In this study, in order to determine and locate the studies that will be included in meta-analysis, an in-depth search was made on Google Scholar search engine. In addition, Science Direct, Proquest, Elsevier and Eric online databases are also scanned and different experimental studies including articles published in refereed or non-refereed journals, master thesis, doctoral dissertations, seminars and books are gathered from this various online open-access databases to conduct the meta-analysis study. The main reason for inclusion of master and PhD dissertations is to arrive as a comprehensive sample of studies that address the domain of interest and also to nullify the probability of publication bias.

In this phase of study, using the term 'CLIL', 'content and language integrated learning', 'the effect of CLIL on', 'the impact of CLIL on' and limiting the search to key words, titles and abstracts, all the studies including the terms are recorded. In addition, a general search was done over the references of all studies included in order to detect further related published studies. The total number of 55 studies is determined as candidate studies to be included in the meta-analysis. After the analysis, thesis having the appropriate problematic and the criteria are included in this study. Investigations and selection of the studies were carried out in the form of a two-phases screening process. In the first phase, heading and abstract of studies are screened and in the second phase the full text of papers are screened.

Studies should have the following characteristics for being chosen. These characteristics are called under the title of inclusion criteria:

- Studies that investigated the effectiveness of CLIL published between the years 2007 and 2017. The latest date for the researches included in this study was set as September 2017;
- Quasi-experimental and experimental studies that used CLIL in the experiment groups are included in this meta-analysis study;
- Studies that do not report statistical information necessary to calculate effect sizes are not included in the analysis;
- Studies have to be written in English;
- studies should give the sample size of the studied groups;

- Studies that provide sufficient information for calculating the effect size (Higgins & Green, 2011; Littel, Corcoran, & Pillai, 2008);

Table3.1. Features of the Studies Included in the Meta-Analysis

Characteristic							Total	
Publication year of research	2007	2008	2009	2010	2011	2012	N= 22 %=100	
	N	-	-	1	2	-		1
	%	0	0	4.5	9	0		4.5
	2013	2014	2015	2016	2017			
	N	6	4	4	1	3		
%	27.5	18.2	18.2	4.5	13.6			
Type of reseach	Master thesis	Doctoral thesis	Article				N=22	
	N	5	3	14				%=100
	%	22.8	13.6	63.6				

Afterwards, the research papers are subjected to detailed examination to determine their eligibility in the meta-analysis and categorized according to the data they include. Then the research abstracts were reviewed. Upon reviewing the abstracts, total of 20 studies were excluded according to the criteria stated above. In the second phase, the remaining research studies were analyzed in detail and 22 of these articles were found to be appropriate while the other 2 were deemed inappropriate. Finally, 22 studies met the criterion were included in the further analyses. Descriptive statistics on those 22 publications are provided in Table 3.1.

3.4. Data Collection

The data coding forms is a bridge to connect the primary research studies and the research synthesis and serve as a historical record of reviewers' decisions (Higgins & Green, 2011). It is a mean to the end of handling a huge amount of data easily and precisely. In addition, the coding form is broad enough to see all of the studies and specific

enough to avoid any missing data of any research. Selected studies are coded based on a data coding procedure in a Microsoft Excel file. In this study, the data coding form is created prior to the statistical analyses and the coding process was made 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).

Coding form used in this research is made up of the components including :

- Research Code
- Name of the study
- Author
- Year of publication
- Country
- Type of the study (article, master thesis, PHD dissertation)
- Sample size
- Study design
- Subject matter(content)
- Language skill and component
- Age
- Gender
- Language Level
- Educational level
- Data collection tool
- Contact hours
- Number of participants in experimental group
- Number of participants in control group
- Experimental group pretest mean score and standard deviation
- Control group pretest mean score and standard deviation
- Experimental group post test mean score and standard deviation
- Control group post test mean score and standard deviation

3.5. Reliability of the Study

To determine the reliability of the coding system and to avoid study selection that is biased by the 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 are MA TEFL candidates at the department of applied linguistics, Shahrood University of Technology. They were supposed to code the whole sample of the studies. To do so, Cohen's kappa reliability coefficient between the coders was calculated using SPSS software. Cohen's kappa reliability is a measure of inter-rater agreement only between two raters for categorical (qualitative) items. Cohen's kappa reliability index was determined to be 0.91 which shows agreement higher than 91% between the two raters. The result shows almost a perfect consistency between the coders. Finally, the disagreement between the coding sheets is checked again by two coders and is corrected by them after reaching a common agreement.

3.6. Data Analysis

3.6.1. Individual Effect Sizes and the Main Effect Size

After collecting the data from studies and coding them according to their important information, the effect size of each individual study must be calculated. It is clear that different studies use different interpretation of the results about the difference between experimental and control groups' means. Some studies have reported statistics in terms of mean and standard deviation while other studies may have reported them in terms of P value, T value or F value. Calculating the effect size of each study is a mean to the end of integration of all the results in different studies to an identical scale. Therefore, the process of comparison and combining the results of the studies with each other would be easier, more practical and more precise. Effect size (standardized mean difference), is a simple way of computing the difference between the means of the two groups. It tells you how much the experimental group of a study differs from its control group (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Cohen's d, Hedges's g and Glass's delta are main measures for calculation the effect size of experimental studies. Hedges' g and Cohen's d are very similar except when sample sizes are below 20, then Hedges' g calculates more precisely and is better than Cohen's d. Hedges' g is therefore sometimes called the corrected effect size. When sample size is more than 20, the results for both methods are almost the same. In fact, The Cohen's d which is based on sample averages, especially for small samples, results in a biased estimate of the population effect size. Furthermore, the main difference between Hedge's g and Cohen's d is that Hedge's g uses pooled weighted standard deviations (instead of pooled standard deviations). It is weighted based on the each study's sample size and standard error. That is, studies conducted with greater sample sizes were given more weight. Glass's delta is used if standard deviations are significantly different between the groups (Experimental and control) in which, Glass's delta uses only the control group's standard deviation (SD_C).

Taking the merit's of Hedge's g over other formulas, the effect sizes are calculated by this formula presented below:

$$\text{Hedges' } g = \frac{M_1 - M_2}{SD_{pooled}^*}$$

Where "M₁ – M₂" is difference between experimental and control groups means and SD_{pooled}^* is pooled and weighted standard deviation.

Hedges' g (like Cohen's d) is biased upwards for small samples (under 50). The effect sizes have also been weighted and corrected for small sample sizes using the following formula:

$$g = \frac{M_1 - M_2}{SD_{pooled}^*} \times \left(\frac{N-3}{N-2.25} \right) \times \sqrt{\frac{N-2}{N}}$$

Cohen's d and Hedges' g are interpreted in a similar way. Cohen suggested using the following rule for interpreting results (Cohen, Manion & Morrison, 2007):

- $0 \leq$ Effect size value ≤ 0.20 insignificant,

- $0.21 \leq \text{Effect size value} \leq 0.50$ small,
- $0.51 \leq \text{Effect size value} \leq 0.8$ medium,
- $0.81 \leq \text{Effect size values}$, large effect size

However, Plonsky and Oswald (2014) suggested the following scale for interpretation of effect sizes in L2 research (Plonsky & Oswald, 2014):

- $0.0 \leq \text{Effect size value} < 0.40$ insignificant,
- $0.4 \leq \text{Effect size value} < 0.70$ small,
- $0.7 \leq \text{Effect size value} < 0.1$ medium,
- $0.1 \leq \text{Effect size values}$, large effect size

Each individual study contributed one effect size. If studies reported effect of CLIL on achievement of different groups, more than one effect sizes was calculated for these studies. Finally, the main effect size (average effect size) is going to be calculated in order to combine the effect sizes of the studies, to see whether CLIL has positive effect on students' overall proficiency in overall and to what extend it affects their language proficiency.

3.6.2. Test of Homogeneity

Before calculating the effect sizes of studies, first the analysis model which will be used for calculation of effect size should be determined. Generally, there are two models of analysis in meta-analysis studies including, the fixed effects model (SEM) and a random effects model (REM) (Hedges & Olkin, 1985).

Fixed effects models take only within-study variability into consideration and believe that differences are only due to within-study variation. They also hold the assumption that studies in a meta-analysis research are homogeneous which means they use identical methods, participants, and measurements and thus, they should produce identical results. They also presuppose that the groups of effect sizes are part of the same population of effect sizes and thus are not influenced by any other variable. So, the results from a homogenous set of studies can be combined and the mean effect size will be calculated for a single population

Random effect models take both between-study and within-study variability into consideration and hold the assumption that studies in a meta-analysis research are heterogeneous and are not functionally identical which means they are a random sample from the universe of all possible studies. Therefore, results of heterogeneous studies can not be combined and the calculated effect size should be generalized to bigger populations. Random effects analysis can be done on all of the studies but it may be better to identify an important subgroup difference and then do a fixed effects analysis of each and report all of the results.

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. In other words, it is assumed that there is no heterogeneity or that it is negligible. Conversely, the latter assumes the existence of heterogeneity. That is, it takes both between-study and within-study variability into consideration. The combined effect size calculated in a meta-analysis is not related to a probability or random sample. Rather, it is observed in a pooled sample taken from different populations; hence the combined effect is an inference about the effect in the population of these populations; 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.

3.6.3. Moderator Analysis (Sub-group Analysis)

After calculating the effect size of each study and the main effect size, sub-group analysis has been done in order to study the effect of CLIL on students' proficiency in detail according to different sub-groups and test the moderator variables. Moderator is always considered as the source of the variations between the mean effect sizes of the variables of the study. Moderators are variables that are supposed to have an effect on the results of meta-analysis and are used to determine the level of this effect. This analysis is done to find statistical differences between subgroups and between the average effect sizes of the variables. It also allows for some comparisons between the groups. Sub-group analysis is planned according to the objective and procedures of the study (Littel, Corcoran & Pillai, 2008).

In this meta-analysis study four variables have been chosen as the moderators which are supposed to play a role in the main effect size measurement:

- Publication type
- Language skills and components (listening, speaking, reading, writing, vocabulary, grammar, pronunciation)
- Subject matter (content)
- Educational level

3.6.4. Publication Bias Evaluation

According to publication bias notion which is also called lost data, research on a specific subject is not completely 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; Kulinskaya *et al.*, 2008). In other words, articles reporting positive or statistically significant results are more likely to be published than those reporting negative or insignificant results. This directly has an effect on the meta-analysis research based on those articles and in consequence, such publication bias can have a negative effect on the overall results of meta-analysis investigations. The publication bias over a certain level may effect on the calculated main effect size and show it higher than its real value (Rothstein, Sutton & Borenstein, 2005).

So, defining and evaluating publication bias is a vital and necessary step which should be done to guarantee the validity of the meta-analysis study.

For publication bias evaluation of this study, the following questions have been asked:

- Is there any existence of publication bias in this meta-analysis study?
- Is it possible that publication bias affects on the value of the main effect size?
- How much of the effect size is due to the publication bias?

There are different methods for evaluation of publication bias in a meta-analysis study. In this study, Funnel Scatter Plot and Rosenthal and Orwin's Fail-Safe N test are used as the tools for calculating publication bias.

3.6.4.1. Funnel Scatter Plot

Funnel Scatter Plot is a graph which helps us visually in recognizing whether there is a publication bias in the study or not. Since the Funnel Plot provides visual information, the result of evaluation by this method is not considered to be accurately objective and numerical. X axis shows effect size values and Y axis shows standard error values in Funnel Scatter Plot. Funnel Scatter Plot is mostly interpreted according to the funnel's symmetry or asymmetry. If the plot forms almost asymmetric funnel, it shows no potential publication bias. On the other hand, if effect sizes of the studies show a relatively symmetrical distribution around the main effect size, the studies seem to have publication bias. In this research, this method is used to see whether there is a publication bias in research over the effect of CLIL or not.

3.6.4.2. Trimm and Fill Method

The trim and fill method is a nonparametric (rank-based) technique suggested by Duval and Tweedie (Duval, 2005). This method is applied in order to estimate the number of studies missing from a meta-analysis due to the suppression of the most extreme results on one side of the funnel scatter plot. Therefore, this method argues the observed data so that the funnel plot is more symmetric. The method should not be regarded as a way of yielding a more "valid" estimate of the overall effect or outcome, but as a way of examining the sensitivity of the results to one particular selection mechanism (i.e., one particular form of publication bias).

3.6.4.3. Fail-Safe N Test

Classical Fail-safe N is another method which is used to evaluate whether there is a publication bias in studies or not. Classical fail-safe N test locates and estimates the number of studies with non-significant results and average zero effect size (lost studies) that are required to nullify the main effect size which means, to reduce the main effect size value to 0.01 and to bring the significant level down to $p = .05$. The more the number of required studies, the less probability of publication bias exists.

CHAPTER FOUR: RESULTS

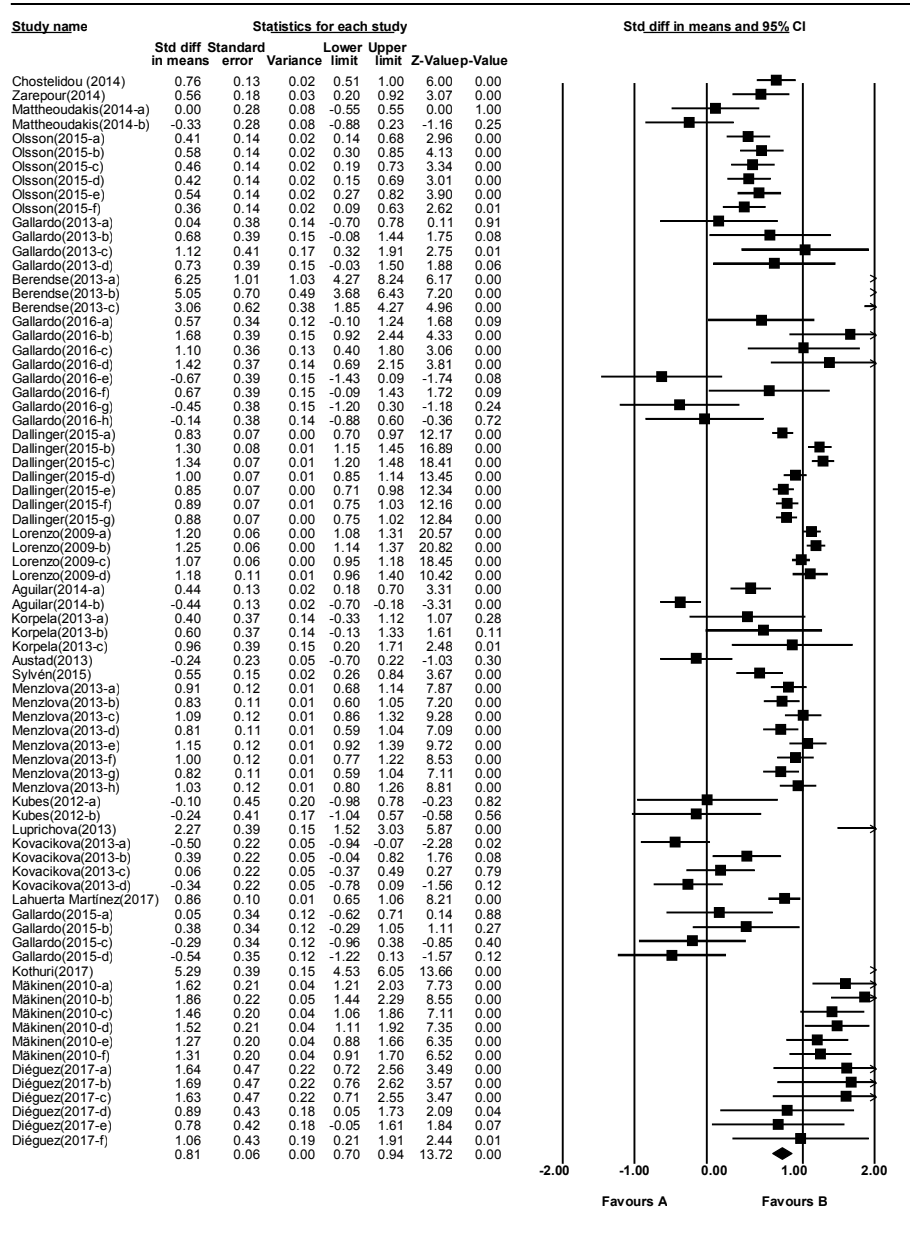
4.1. Individual and Main Effect Size Analyses

Table 4.1 shows effect size, degrees of freedom, 95% confidence intervals and some other statistics for random effect models. The overall effect sizes for random model is 0.81 and is considered to have medium magnitudes (when converted to Hedge's g) according to Plonsky and Oswald's scale and large magnitudes according to Cohen's scale for interpretation of effect sizes. The overall positive effect size indicates a positive effect of CLIL method on student's proficiency. Furthermore, the significant Q statistic result ($Q(76) = 871.98, p < 0.05$), shows a heterogeneity of effect sizes, suggesting the need to use random effect model for subsequent analysis. In addition, that I squared value in Table 4.1 is bigger than 75 % shows that the distribution of the effect values of studies on CLIL are highly heterogeneous.

Table 4.1. Fixed and random effect model statistic

		Effect size and 95% confidence interval					Test of null (2-tail)			Heterogeneity		
Model	Number studies	Point Estimate	Standard Error	Variance	Lower Limit	Upper Limit	Z Value	P value	Q Value	Df	P Value	I Squared
Fixed	76	0.91	0.01	0.00	0.88	0.94	58.44	0.00	871.98	75	0.00	91.39
Random	76	0.81	0.06	0.00	0.69	0.93	13.65	0.00				

The distribution of the effect sizes of 76 studies is shown in Figure 4.1 the forest plot. Among the effect sizes represented in Figure 4.1, Gallardo del Puerto (2016-e) had the lowest effect size (-0.67) and Berendse (2013-a) had the highest effect size (6.25). Meanwhile, among all effect sizes, twelve of them are negative.



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

Figure 4.1. Forest plot of the effect sizes

4.2. Moderator Analysis (Sub-group Analysis)

As mentioned before, moderator is always considered as the source of the variations between the mean effect sizes of the variables of the study and is supposed to have an effect on the results of meta-analysis and is used to determine the level of this effect. This analysis is done to find statistical differences between subgroups and between the average effect sizes of the variables. Sub-group analysis has been done in order to study the effect of CLIL on students' proficiency in detail according to different sub-groups and test the moderator variables. In order to reveal the reasons of heterogeneous distribution, educational level, skill, publication type and subject were used as moderator variables. The results of the examination of the effects of CLIL on students' proficiency according to moderator variables are presented in Table 4.2.

Table 4.2. Moderator analysis on the effectiveness of CLIL

Moderator variables	Number of calculated effect sizes	variance	Z value	P value	Effect size	95% confidence interval		Standard Error
						Lower	Upper	
Educational Level								
Primary	26	0.00	16.00	0.00	1.02	0.89	1.14	0.06
Secondary	38	0.00	10.77	0.00	0.81	0.66	0.95	0.07
University	12	0.07	1.44	0.14	0.40	-0.14	0.95	0.28
Skill								
General	10	0.01	6.62	0.00	0.76	0.54	0.99	0.11
Grammar	17	0.07	4.86	0.00	1.32	0.78	1.85	0.27
Listening	8	0.01	7.93	0.00	0.91	0.69	1.14	0.11
Pronunciation	3	0.12	-0.00	0.99	0.00	-0.69	0.69	0.35
Reading	6	0.02	4.58	0.00	0.72	0.41	1.03	0.15
Speaking	8	0.07	1.27	0.20	0.35	-0.19	0.90	0.28
Vocabulary	20	0.01	6.57	0.00	0.90	0.63	1.17	0.13
Writing	4	0.02	5.61	0.00	0.81	0.53	1.10	0.14
Publication type								
Article	50	0.00	11.87	0.00	0.74	0.61	0.86	0.06
Doctoral thesis	7	0.08	0.66	0.50	0.18	-0.36	0.74	0.28
Master thesis	19	0.03	7.62	0.00	1.46	1.08	1.84	0.19
Subject matter								
Accountancy	1	0.01	6.00	0.00	0.75	0.50	1.00	0.12
Agriculture	4	0.04	-0.49	0.62	-0.09	-0.49	0.29	0.20
Biology	13	0.02	3.21	0.00	0.54	0.21	0.88	0.17
Business	6	0.00	8.14	0.00	0.45	0.34	0.57	0.05
Creative art	6	0.03	6.81	0.00	1.20	0.85	1.55	0.17
Economics	6	0.00	8.14	0.00	0.45	0.34	0.57	0.05
Engineering	2	0.18	0.00	1.00	0.00	-0.84	0.84	0.43
English literature	4	0.03	-0.49	0.62	-0.09	-0.48	0.28	0.19
Geography	22	0.00	9.08	0.00	0.74	0.58	0.90	0.08
History	24	0.00	11.08	0.00	0.80	0.66	0.94	0.07
Hotel	1	0.14	13.66	0.00	5.25	4.50	6.00	0.38
Management								
Math	14	0.00	12.95	0.00	0.85	0.72	0.98	0.06
Religion	1	0.05	-1.03	0.30	-0.23	-0.69	0.21	0.23
Science	31	0.00	10.21	0.00	0.71	0.57	0.84	0.07
Social science	7	0.10	3.03	0.00	0.98	0.34	1.61	0.32

4.2.1. Educational Level

In order to examine whether the effect levels of CLIL on students' proficiency differentiate according to educational level, the studies included in this research were classified under three groups as primary, secondary and university level. According to results of the analysis in Table 4.2, average effect size of the studies calculated according to primary, secondary and university levels were found to be 1.02, 0.81 and 0.40 respectively. Where primary level had the maximum (1.02) and university level had the minimum (0.40) group averages.

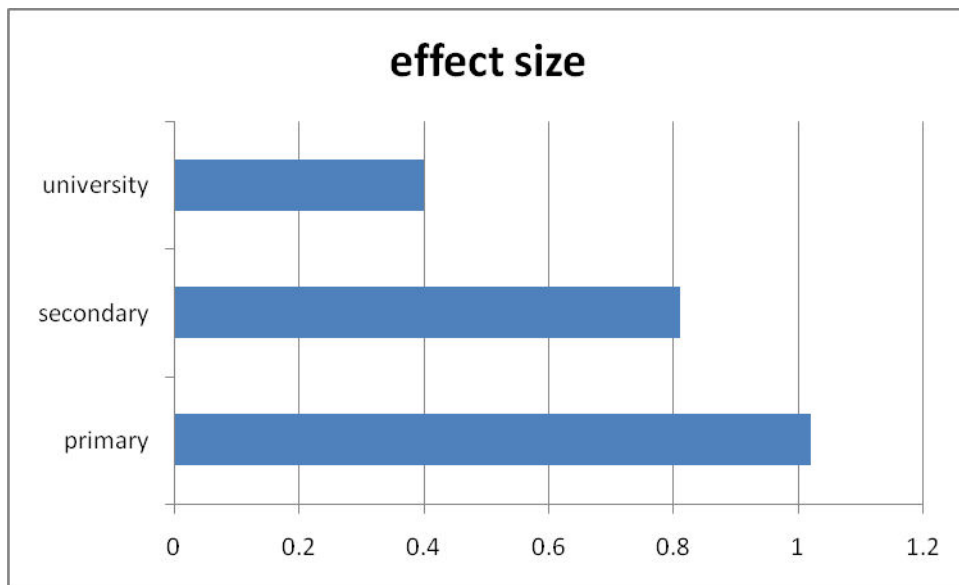


Figure 4.2. Comparison of effect sizes between different educational levels

4.2.2. Skill

In order to determine whether the effect of CLIL on students' proficiency differentiate according to skills, the studies were classified under eight groups according to the skills and components under instruction as general proficiency, grammar, listening, pronunciation, reading, speaking, vocabulary and writing. According to results of the analysis in Table 4.2, average effect size of these studies was calculated as 0.76, 1.32, 0.91, 0.00, 0.72, 0.35, 0.90 and 0.81 respectively. Where grammar had the maximum (1.32) and pronunciation had the minimum (0.00) group averages.

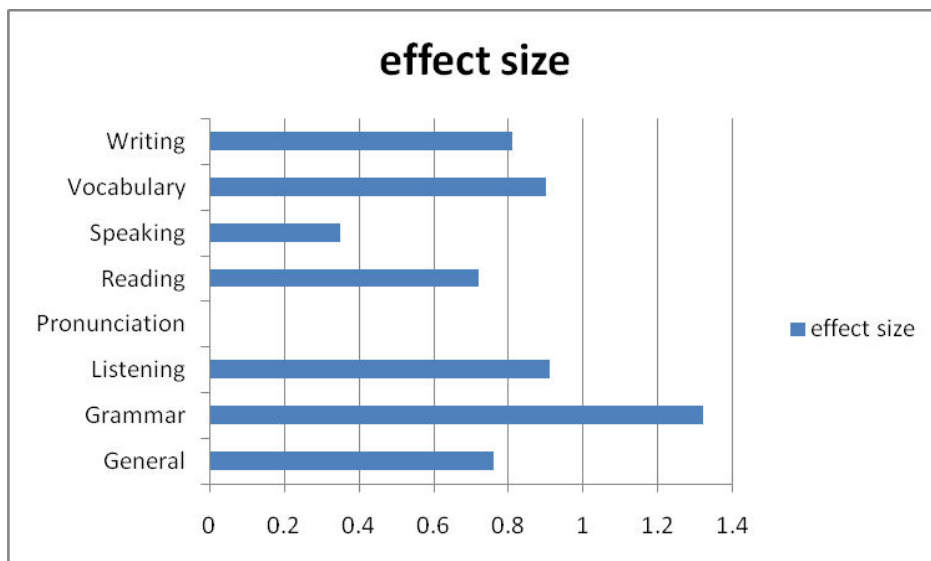


Figure 4.3. Comparison of effect sizes between different language skills and components

4.2.3. Publication Type

In order to determine whether the effect of CLIL on students' performance, the studies were classified under three groups according to their type as articles, master's thesis and doctoral thesis. As a result of the analysis in Table 4.2, average effect size of practices in articles, master's thesis and doctoral thesis were calculated as 0.74, 1.46 and 0.18 respectively. Where master thesis result had the maximum (1.46) and doctoral thesis result had the minimum (0.18) group averages.

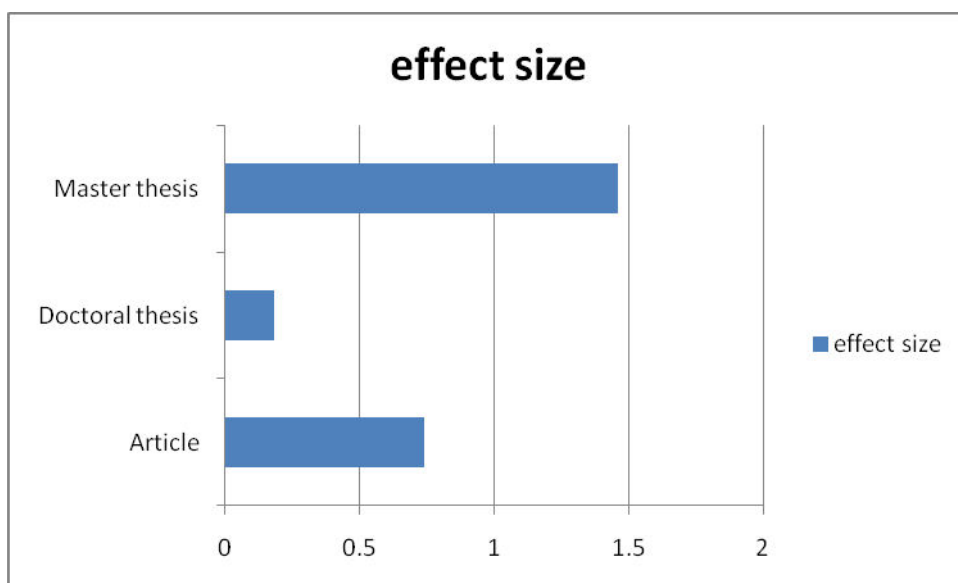


Figure 4.4. Comparison of effect sizes between different publication types

4.2.4. Subject matter

Different studies used different content in implementation of CLIL in education. In order to find the main effect of CLIL on students' English proficiency, the studies were categorized according to different academic subjects including accountancy, agriculture, biology, business, creative art, economics, engineering, English, geography, history, hotel management, math, religion, science and social science. When studying the result of the subject matter as a moderator variable, it should be taken into account that some studies didn't mention what subject matter they used and some studies involved more than one subject matter in their CLIL programs. That is why the total number of effect sizes studies is more than 76. As a result of the analysis according to subject, CLIL seems to have negative effect size on students' language proficiency based on three subject matters including agriculture (-0.09), English literature (-0.09) and religion (-0.23) in which, CLIL has the most negative effect on students language proficiency when the content is religion among all the academic subjects. On the other hand, this approach has positive effect size on students' language proficiency based on accountancy, biology, business, creative art, economics, geography, history, hotel management, math, science and social science in which the highest positive effect size value (5.25) was calculated for hotel management and the lowest positive effect size value (0.45) was calculated for business. As it is shown, CLIL seems to have no effect on students' language proficiency when the subject matter was engineering.

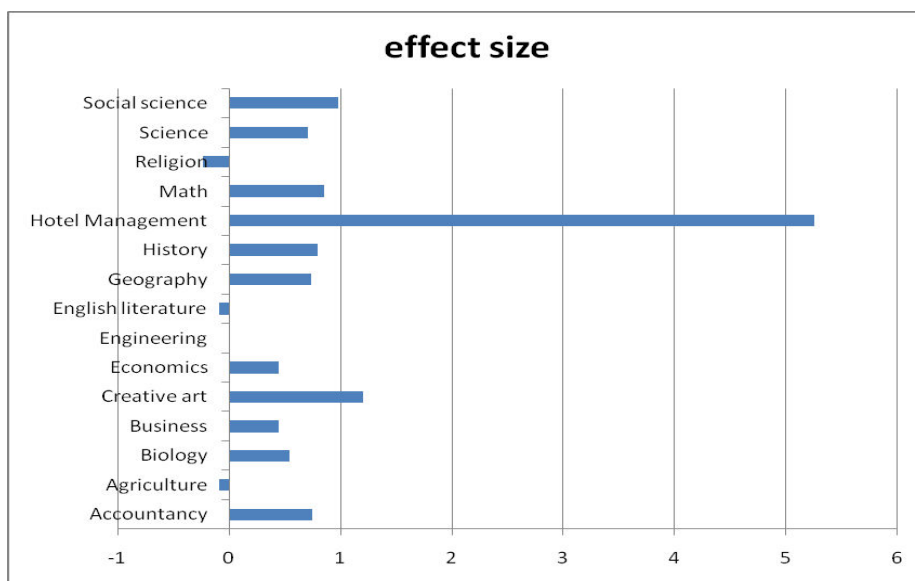


Figure 4.5. Comparison of effect sizes between different subject matters

4.3. Publication Bias Evaluation

The main question in this section is whether there is any existence of publication bias in this meta-analysis. If so, what is the level of publication study in this meta-analysis?

For evaluating the publication bias in this study the following questions were answered:

- Is there any existence of publication bias in this meta-analysis study?
- Is it possible that publication bias affects on the value of the main effect size?
- How much of the effect size is due to the publication bias?

4.3.1. Funnel Scatter Plot

Funnel Scatter Plot is the first tool used to helps us visually in recognizing whether there is a publication bias in the study or not. As it is presented in figure 4.6, X axis shows effect size values and Y axis shows standard error values in Funnel Scatter Plot. In the plot, the studies with small samples are mostly distributed in the bottom of the graph. On the contrary, the studies with large samples are distributed on the top of the graph around the main effect size value. The range of distribution is wide in the bottom of the graph because studies with smaller sample sizes have more sampling error variation in effect sizes. However, the range of distribution is narrow on the top of the graph because studies with bigger sample sizes have less variation in effect sizes. This pattern forms a funnel-shaped plot. This plot is interpreted according to the funnel's symmetry or asymmetry. If the plot forms 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. In this research, this method is used to see whether there is a publication bias in research over the effect of CLIL or not. As it is shown in Figure 4.6, the funnel plot forms almost asymmetric distribution, therefore, it presents a little evidence of potential publication bias.

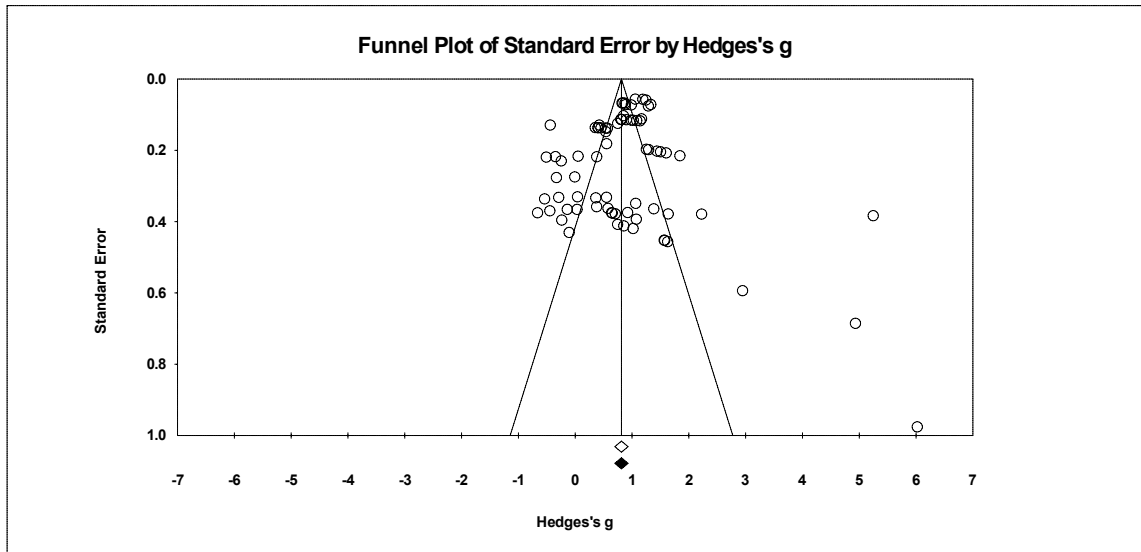


Figure 4.6. Funnel plot on observed studies

4.3.2. Trimm and Fill Method

On the other hand, in the graph of Duval and Tweedie created by the Trimm and Fill method (figure 4.7), if 11 studies are placed at the right side of the graph, it will be a fully symmetric structure and totally unbiased. In this case, effect size value of the studies increases from 0.81 to 1.00 regarding random effect model of analysis. According to Cohen's (1987) effect size classification, both values show a large effect and there is not a big difference between observed effect size in publication-biased case and adjusted effect size in a totally unbiased case. Therefore, in case of publication bias, the studies which were not obtained will not change the calculated effect size for success. In addition, adding only 11 dummy studies for a study that combined with the aid of the meta-analysis of 76 effect sizes presents a low publication bias.

Table 4.3. The result of Trim and Fill analysis

	Studies Trimmed	Fixed Effects			Random Effects			Q value
		Point estimate	Lower limit	Upper limit	Point estimate	Lower limit	Upper limit	
Observed values		0.91	0.88	0.94	0.81	0.69	0.93	871.98
Adjusted values	11	0.97	0.94	1.00	1.00	0.87	1.12	1220.77

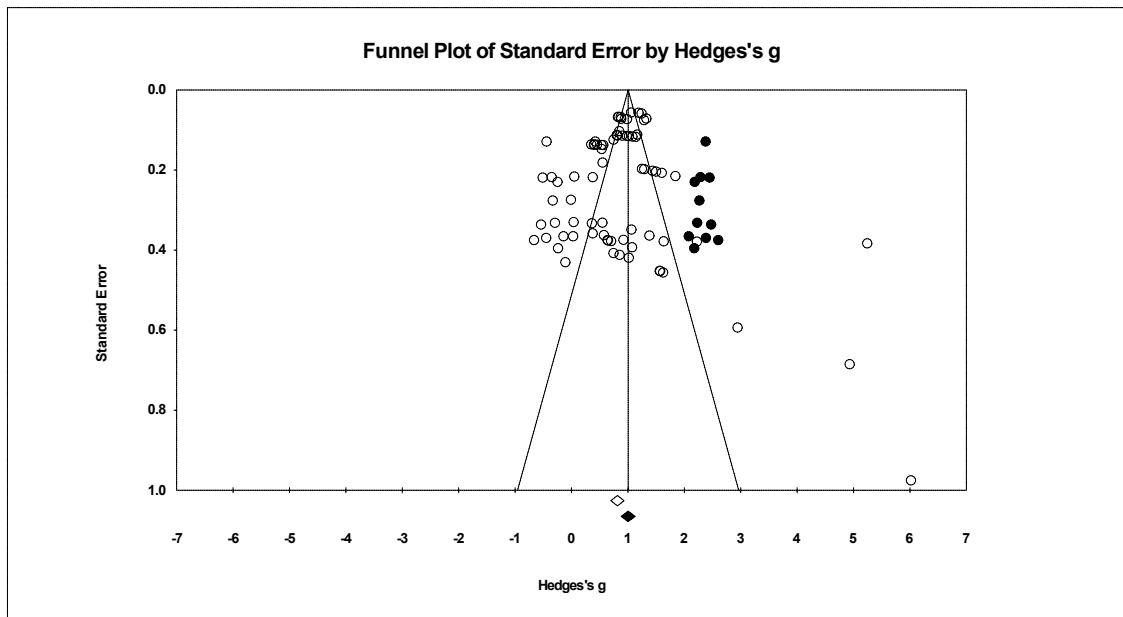


Figure 4.7. Funnel plot on observed and imputed studies

4.3.3. Classic Fail-safe N

Another examination of the possibility of publication bias was performed in terms of the classic fail-safe N analysis to locate and estimate the number of studies with non-significant results and average zero effect size (lost studies) that are required to nullify the main effect size which means, to reduce the main effect size value to 0.01 and to bring the significant level down to $p = 0.05$. The more the number of required studies, the less probability of publication bias exists. As presented by the data in Table 4.4 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 4.4. 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

**CHAPTER FIVE:
DISCUSSION AND CONCLUSION**

5.1. Overview

The objectives of the present study are to investigate the effectiveness CLIL on students' overall language proficiency, to imply a subgroup analysis to find out its effectiveness on different groups of language skills and components and to calculate the effect of this approach on students' proficiency in combination with other moderator variables such as educational levels and subject matter. To this end, 22 experimental studies on the effect of CLIL on student's language proficiency were included in this meta-analysis which resulted in 76 effect sizes and embodies 5591 participants among all the studies. This chapter presents a summary of the findings, a discussion of the findings in relation to the previous studies in the field, the pedagogical implications of the study, and the recommendations for further research.

5.2. Discussion and Conclusion

The results section addressed the three research questions posed in this meta-analysis. 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. Among all the calculated effect sizes, 64 of them are positive and 12 of them are negative which shows most studies agree on the positive effect size of CLIL on students' proficiency.

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 meta-analyses in the same domain, it is vividly large.

Despite the large magnitude of the overall combined effect found in this meta-analysis, it should be interpreted cautiously since the $I^2 = 91.39$ is suggestive of a high level of heterogeneity. I^2 shows the proportion of observed variance that reflects differences in effect sizes. Borenstein *et al.* (2009) suggest that I^2 be used as a criterion to decide whether the 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.

Table 4.2 answers the second question by showing how education level, skills and subskills, and subject matter moderate the effectiveness of CLIL. According to the subgroup analysis which was done based on four moderator variables, it was found that CLIL method has been used through a variety of school subjects, on different language skills and in many different school levels. Moreover, CLIL studies have been published for different purposes including PhD dissertations, master theses and research papers. Taking publication type moderator variable into account, the results revealed that CLIL had the highest to the lowest group average in studies published as master thesis, articles and doctoral thesis respectively which shows studies published as doctoral thesis have the most critical point of view on the effect of CLIL on students' proficiency. On the other hand, the result of educational level moderator analysis have shown that CLIL has the maximum effect on language proficiency in primary level, a moderate effect in secondary level and the minimum effect in 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, the implementation of this method in lower bilingual educations is prior to the higher levels of education and the biggest part of financial budget, energy and time

should be spent on the lower educational levels when using the CLIL approach. Additionally, taking school subject as the moderator variable, science, history and geography were the most popular subjects in CLIL programs. Moreover, hotel management was the subject that had the most effect size among other subjects. However, since only one study worked on hotel management, accountancy and religion as the content of CLIL program, the results of their effect size are not significant enough. Therefore, creative art, social science, math, history, geography, science and biology were the subjects that got the highest effect size among all the studies. Also, English literature and agriculture were the subjects that CLIL had a negative effect on students' language proficiency in the programs using them as content. Taking these results into consideration, some school subjects seem to be more effective on students' language learning in CLIL environment than the others. So, subjects such as creative art, social science, math, history, geography, science and biology are more appropriate to be used as the content of CLIL educations because they gained the highest effect size on students' language proficiency in this case.

In order to determine whether the effect of CLIL on students' proficiency differentiates according to skills, the studies were classified under eight groups according to the skills and subskills under instruction as general proficiency, listening, speaking, reading, writing, grammar, pronunciation and vocabulary. As the results show, CLIL had the highest effect on grammar as a component. It also had a large positive effect on vocabulary and a no effect on students' pronunciation component. However, since only three studies investigated the effect of CLIL on pronunciation, the final result is not significant and reliable enough. In addition, among the four language skills, CLIL had the highest 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; therefore, through this approach, students can obtain grammatical and vocabulary proficiency but not pronunciation proficiency. This shows that the CLIL environment is not sufficient to improve students' pronunciation skills. For this reason, other strategies for teaching language pronunciation should be added to the program to compensate this lack.

The disparity of effect sizes which reflect 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 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, creative arts, and hotel management, which 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.

Three basic tests of evaluation of publication bias including Funnel Scatter Plot, Trimm and Fill Method and Fail-Safe N test reported a low level of publication bias among the studies in this meta-analysis. The results show that there are few unpublished lost studies which may change the calculated main effect size in this meta-analysis study. Hence, the final result of these three methods guarantees the validity of the meta-analysis study.

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 4.2 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.

5.3. Implications for Practice

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

- The curriculum developers of bilingual education institutes, so they design their educational program according to the strengths and weaknesses of CLIL over different skills and components and different educational levels and different ages. It also helps them to choose the most appropriate subject matter in CLIL programs.
- Practitioners, so the results give them an overall insight and reality into this approach and enable them to make better and more precise lesson plans before implementing the CLIL environment in their classrooms.
- Researchers, who can reach a conclusive result about the impact of CLIL on second language learner's proficiency by this study. So, they do not waste their time and energy over reviewing the huge number of previous studies to find out the overall effect of this approach.
- 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.

5.4. Suggestion for Future Research

This meta-analysis aimed at presenting the overall combined effect size of CLIL 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 (Appendix A), 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 are the findings of the study. Having a look at the length of the whiskers show 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. Furthermore, 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 the other ones have been paid little attention and 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. Therefore, more programs are needed to implement them and report their effect on learning in CLIL environment. Finally, to reduce the high level of heterogeneity which characterized this meta-analysis, future studies should further specify the dependent variable or the outcome to enable them to synthesize the findings of studies which present more homogeneous effect sizes.

References

References marked with an asterisk indicate studies included in the meta-analysis.

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.

Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

Baker, C. (2001). *Foundations of Bilingual Education and Bilingualism*. Clevedon: Multilingual Matters.

*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.

Betty, L. and Boris, S. (2004). *Developing Professional-Level Language Proficiency*. Cambridge: Cambridge University Press.

Bialystok, E. (2002). Acquisition of literacy in bilingual children: A framework for research. *Language Learning*, 52(1), 159-199.

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.

Brinton, D., Snow, M., & Wesche, M. (1989). *Content-Based Second Language Instruction*. New York: Newbury House.

Calve, P. (1991). Vingt-cinq ans d'immersion au Canada 1965–1990. *Etude de Linguistique Appliquée*, 82, 7–19.

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

Canale, M. and Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1, 1-46.

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.

Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied linguistics*, 35(3), 243-262.

*Chostelidou, 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.

Cook, V. (1992). Evidence of multicompetence. *Language Learning*, 42(4), 557-591.

Cook, D. J., Guyatt, G. H., Salena, B. J., & Laine, L. A. (1992). Endoscopic therapy for acute nonvariceal upper gastrointestinal hemorrhage: A meta-analysis. *Gastroenterology*, 102(1), 139-148.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological bulletin*, 87(3), 442.

Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *The International Journal of Bilingual Education and Bilingualism*, 10, 543–562.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. New York: Cambridge University Press.

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

Cummins, J. (1992). Language proficiency, bilingualism, and academic achievement. In P. A. Richard-Amato & M. A. Snow (Eds.), *The multicultural classroom: Readings for content-area teachers* (pp. 16-26). New York: Longman.

*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. Volkman (Eds.) *Future Perspectives for English Language Teaching* (pp. 139-157). Heidelberg: Carl Winter.

De Diezmas, E. N. M. (2016). The impact of CLIL on the acquisition of L2 competences and skills in primary education. *International Journal of English Studies*, 16(2), 81-101.

De Graaff, R., Jan Koopman, G., Anikina, Y., & Westhoff, G. (2007). An observation tool for effective L2 pedagogy in content and language integrated learning (CLIL). *International Journal of Bilingual Education and Bilingualism*, 10(5), 603-624.

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

Dörnyei, Z. (1995). On the teachability of communication strategies. *TESOL Quarterly* 29(1), 55-85.

Doughty, C. & Williams, J. (1998). *Focus on Form: Second language Acquisition*. Cambridge: Cambridge University Press.

Duckworth E (2009). Helping students get to where ideas can find them. *The New Educator*, 5, 185-188.

Dupin-Bryant, P. A. (2004). Variables related to interactive television teaching style: In search of learner-centered teaching styles. *International Journal of Instructional Technology and Distance Learning*, 1(4), 3-14.

Duval, S. J. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.) *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 127–144). Chichester, England: Wiley.

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

Finocchiaro, M., & Brumfit, C. (1983). *The functional-notional approach: From theory to practice*. Oxford University Press, 200 Madison Ave., New York, NY 10016.

Gajo, L. (2007). Linguistic knowledge and subject knowledge: How does bilingualism contribute to subject development? *The International Journal of Bilingual Education and Bilingualism*, 10, 563-581.

*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.

Genesee, F., & Lindholm-Leary, K. (2013). Two case studies of content-based language education. *Journal of Immersion and Content-Based Language Education*, 1(1), 3-33.

Glass, G. V., McGaw, B. & Smith, M. L. (1982). *Meta-analysis in social research*, Beverly Hills, CA: Sage.

Greene, J.G., Pearson, P. & Schoenfeld, A. (1999). Achievement and theories of knowing and learning. In R. McCormick & C. Paechter (Eds.) *Learning and Knowledge*. London (pp. 136-153). Paul Chapman/Open University.

Halbach, A. (2009). The Primary School Teacher and the Challenges of Bilingual Education In *CLIL Across Educational Levels: Experiences from Primary, Secondary and Tertiary Contexts*, edited by E. Dafouz (pp. 19–26). Madrid: Santillana.

Hammond, J. (2001). *Scaffolding, teaching and learning in language literacy education*. Australia: PETA.

Harrop, E. (2012). Content and Language Integrated Learning (CLIL): Limitations and Possibilities. *Encuentro*, 21, 57–70.

Hawkins, M.R. (2004). Researching English Language and Literacy Development in Schools. *Educational Researcher* 33(3), 14-25.

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

Higgins, J. P., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions*. UK: John Wiley & Sons.

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

Hunter, J. E., Schmidt, F. L. & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Hymes, D. (1972). On communicative competence. In Pride, J. D. & Holmes, J. (eds.), *Sociolinguistics* (pp. 269- 2939). Harmondsworth: Penguin.

Jacobs, G. M. & McCafferty, S. G. (2006). Connections between Cooperative Learning and second language learning and teaching. In S. G. McCafferty, G. M. Jacobs & C. DaSilva Iddings (Eds.) *Cooperative Learning and Second Language Teaching* (pp. 18-19). New York: Cambridge University Press, 18-29.

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.

Johnson, R. K., & Swain, M. (Eds.). (1997). *Immersion education: International perspectives*. Cambridge University Press.

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. 103-114). 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), 745-751.

*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.D. (1985). *The Input Hypothesis*. London: Longman.

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.

Kumaravadivelu, B. (1994). The Postmethod condition: Emerging strategies for second/foreign language teaching. *TESOL Quarterly*, 28(1), 27-48.

Lado, R. (1961). *Language Testing*. New York: McGraw-Hill.

*Lahuerta Martínez, A. C. (2017). Analysis of the effect of CLIL programmes 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, 30-41.

Lasagabaster, D., & Sierra, J. M. (2009). Immersion and CLIL in English: more differences than similarities. *ELT Journal*, 64(4), 367-375.

Lee, J. & Schallert, D. L. (1997). The Relative Contribution of L2 Language Proficiency and L1 Reading Ability to L2 Reading Performance: A Test of the Threshold Hypothesis in an EFL Context. *TESOL Quarterly*, 31(4), 713-739.

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

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Richards and T. Bhatia (Eds.) *Handbook of Second Language Acquisition*. San Diego: Academic 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.

Lyster, R. (1998). Form in immersion classroom discourse: In or out of focus? *Canadian Journal of Applied Linguistics*, 1, 53-82.

*Mäkinen, M. (2010). *Yesterday I met the biggest ice cream and dranked 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). Jyväskylä, Finland.

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

*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. 215-234). Versita Publications.

Mehisto, P., Marsh, D. & Frigols, M.J. (2008). *Uncovering CLIL. Content and language integrated learning in bilingual and multilingual education*. Oxford: MacMillan.

*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. 13-60). Bratislava: ŠPÚ

Met, M. (1999). *Content-based instruction: Defining 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.

Muñoz, C. (2002). Relevance and potential of CLIL. *CLIL/EMILE The European Dimension: Action, Trends and Foresight Potential*. Brussels: European Commission, 33-34.

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

Nikula, T., Dalton-Puffer, C., & García, A. L. (2013). CLIL classroom discourse: Research from Europe. *Journal of Immersion and Content-Based Language Education*, 1(1), 70-100.

Olaizola, I. V., & Mayo, M. P. G. (2009). Tense and agreement morphology in the interlanguage of Basque/Spanish bilinguals: CLIL versus 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.

Oller, J.W. (1976). Language Testing. In R. Wardhaugh & H.D. Brown: *A Survey of Applied Linguistics* (pp. 275-300). Michigan University Press.

*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.

Paran, A. (2013). Content and language integrated learning: Panacea or policy borrowing myth?. *Applied Linguistics Review*, 4(2), 317-342.

Peregoy, S. F., & Boyle, O. F. (2008). *Reading, Writing, and Learning in ESL: A Resource Book for K-12 Teachers* (5th ed.). New York, NY: Addison-Wesley.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences*. MA-USA: Blackwell Publishers Ltd.

Piaget, J. (1963). *The origins of intelligence in children*. New York: Norton and Company.

Pistorio, M. I. (2010). A blend of CLIL and cooperative learning creates a socially constructed learning environment. *Latin American Journal of Content & Language Integrated Learning*, 3(1), 1-10.

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

Pokrivčáková, S., Menzlová, B., & Farkašová, E. (2010). Creating Conditions for Effective Application of CLIL Methodology in Slovakia. In Pokrivčáková, S. *et al.* (Eds.), *Modernization of Teaching Foreign Languages: CLIL, Inclusive and Intercultural Education* (pp. 7-20). Brno: Masarykova Univerzita.

Rehorick, S., & Edwards, V. (Eds.). (1991). *French immersion: Process, product and perspectives*. Welland, ON: Canadian Modern Language Review.

Robinson, P. C. (1991). *ESP today: A practitioner's guide*. Hamel, Hempstead: Prentice Hall.

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.

Rothstein, H., Sutton, A. J., and Borenstein, M. (2005). *Publication Bias and Meta-Analysis: Prevention, Assessments and Adjustments*. Chichester: John Wiley and Sons.

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), 582-602.

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.

Spanos, G. (1990). On the integration of language and content instruction. *Annual Review of Applied Linguistics*, 10, 227-240.

Stoller, F. L. (2008). Content-based instruction. In *Encyclopedia of language and education* (pp. 1163-1174). Springer.

Swain, M. (1985). Communicative competence. Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.). *Input in Second Language Acquisition* (pp. 235-53). Rowley, MA: Newbury House.

*Sylvén, L. K., & Ohlander, S. (2015). The CLISS project: Receptive vocabulary in CLIL versus non-CLIL groups. *Moderna språk*, 108(2), 80-114.

Tarnopolsky, O. (2013). Content-based instruction, CLIL, and immersion in teaching ESP at tertiary schools in non-English-speaking countries. *Journal of ELT and Applied Linguistics (JELTAL)*, 1(1), 1-11.

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), 28–53.

Van de Craen, P., Ceuleers, E., & Mondt, K. (2007). Cognitive development and bilingualism in primary schools: teaching maths in a CLIL environment. *Diverse Contexts—Converging Goals. CLIL in Europe. Frankfurt am Main: Peter Lang*, 185-200.

Vygotsky, L.S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.

Walker, C. L., & Tedick, D. J. (2000). The complexity of immersion education: Teachers address the issues. *The Modern Language Journal*, 84(1), 5-27.

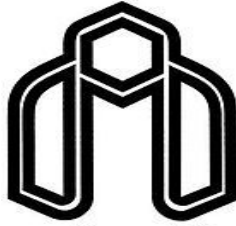
Willing, A. C. (1985). Meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, 53(3), 269-317.

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.

چکیده

پژوهش حاضر روش فراتحلیل در محاسبه تاثیر تلفیق زبان و محتوا بر مهارت زبانی را ارائه نموده است. با بکارگیری مجموعه ای از معیار شمول و غیر شمول، ۲۲ مطالعه اصلی شامل ۷۶ میزان تاثیر از منابع مختلف اعم از پایان نامه های کارشناسی ارشد، رساله های دکتری و مقالات مختلف در سال های ۲۰۰۷ تا ۲۰۱۷ و به زبان انگلیسی جمع آوری شده است. چهار متغیر واسطه شامل مهارت ها و اجزای زبانی، موضوع درسی، مقطع تحصیلی و نوع نشر شناسایی شدند. میزان تاثیر کلی ۰.۸۱ محاسبه شد که ضریب تاثیر متوسطی را بر اساس مقیاس پلانسکی و اسوالد نشان می دهد. نتایج تحلیل واسطه نشان میدهد که بالاترین میزان تاثیر در پایان نامه های کارشناسی ارشد مشاهده شده است. به علاوه، تلفیق زبان و محتوا بیشترین تاثیر را بر گرامر و مهارت شنیداری زبان آموزان و در مقاطع پایینی و ابتدایی تحصیلی داشته است. تلفیق زبان و محتوا بیشترین تاثیر خود بر زبان آموزان را هنگامی داشته است که موضوع درسی مدیریت هتل داری بوده است. در ادامه، این پژوهش با بحث درباره کاربرد یافته ها و توصیه هایی برای مطالعات آینده به نتیجه گیری می پردازد. یافته های این پژوهش کاربرد های صریحی برای مربیان، محققان و برنامه ریزان آموزشی دارد.

کلمات کلیدی: تلفیق زبان و محتوا، فراتحلیل، مهارت زبانی، ضریب تاثیر



دانشگاه صنعتی شاهرود

گروه زبان انگلیسی
پایان نامه کارشناسی ارشد آموزش زبان

تلفیق زبان و محتوا: فرا تحلیل

نگارنده: شیوا نخئی

استاد راهنما
دکتر سید علی استوار نامقی

شهریور ۱۳۹۷