

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/207520>

Please be advised that this information was generated on 2020-01-01 and may be subject to change.

# Content and Language Integrated Learning in English as a Foreign Language

A European Perspective

José Goris

## **Colofon**

Content and Language Integrated Learning in English as a Foreign Language:  
A European Perspective  
José Goris

Published by: Behavioural Science Institute, Radboud University Nijmegen

ISBN/EAN: 978-94-6375-423-1

Cover design: Stijn Eikenaar, [persoonlijkproefschrift.nl](http://persoonlijkproefschrift.nl)

Lay-out: Stijn Eikenaar, [persoonlijkproefschrift.nl](http://persoonlijkproefschrift.nl)

Printing: Ridderprint BV | [www.ridderprint.nl](http://www.ridderprint.nl)

Copyright @ José Goris, 2019

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, without written permission of the author.

# Content and Language Integrated Learning in English as a Foreign Language

A European Perspective

Proefschrift

ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,  
volgens besluit van het college van decanen  
in het openbaar te verdedigen op dinsdag 3 september 2019  
om 14.30 uur precies

door

Josephina Adriana Goris  
geboren op 7 januari 1951  
te Bergen op Zoom

**Promotoren:** Prof. dr. L.T.W. Verhoeven  
Prof. dr. E.J.P.G. Denessen

**Manuscriptcommissie:**

Prof. dr. A. M.C. van Kemenade  
Prof. dr. L.J.M. van Balkom  
Prof. dr. H.C.J. de Graaff (Universiteit Utrecht)

**Paranimfen:**

Karin Kwisthout  
Hans Kwisthout

*“The future belongs to those who believe in the beauty of their dreams”*

*Eleanor Roosevelt*

*Voor Chris*



## Table of Contents

<b>Chapter 1</b>	General Introduction	9
<b>Chapter 2</b>	Effects of the Content and Language Integrated Learning approach to EFL teaching: A comparative study	23
<b>Chapter 3</b>	Effects of Content and Language Integrated Learning in Europe: A systematic review of longitudinal experimental studies	47
<b>Chapter 4</b>	The contribution of CLIL to learners' international orientation and EFL confidence	81
<b>Chapter 5</b>	Determinants of EFL learning success in Content and Language Integrated Learning	99
<b>Chapter 6</b>	Conclusion and Discussion	125





# 1

# CHAPTER

General introduction

### 1.1 Content and Language Integrated Learning

The acronym “CLIL”, short for Content and Language Integrated Learning, was introduced in Europe in the 1990s, when the approach gained momentum and was described as a dual focus methodology in which content and a foreign language are learnt together, in an integrated way (Marsh et al., 2001). CLIL was first introduced as a highly selective programme at secondary schools preparing for university studies. Over the past decades the innovative educational concept has increasingly been implemented in European mainstream schools of primary, secondary and vocational levels. Nowadays CLIL is associated with the teaching and learning of English as a Foreign Language (EFL), as through the years more and more CLIL programmes have adopted English as their target language. The preparedness for the CLIL methodology resulted from a dissatisfaction with the current L2 learning practise in secondary education around the 1980s and 1990s, which relied too much on knowledge of grammar for written skills while practising spoken skills by social interaction was a poor relation in the language classroom (Breen & Candlin, 2001). This is not to say that better language learning is the only CLIL target. Its aims reach much further and entail educating young people for life in an increasingly internationalised world, in which not only excellent L2 skills are needed for study or career, but also a feeling of confidence in discussions and an understanding of different cultures (Eurydice Report, 2006).

Throughout the European educational context considerable variation exists as to the way CLIL is put into practice. Differences relate to learner admittance criteria, the part of the curriculum involved in CLIL teaching, subjects, out-of-school L2 involvement and teacher education and skills. Research conducted in various European quarters has highlighted the better performance as to English as a foreign language of learners in - predominantly secondary school - CLIL classes when compared to mainstream classes (see e.g. Dalton- Puffer, 2008; Pérez-Cañado, 2012). Research into affective outcomes has also shown positive results in favour of CLIL learners. They tend to be more motivated to learn the foreign language than their non-CLIL peers and have more positive attitudes towards language learning (Doiz et al., 2014; Pérez-Cañado, 2012). However, robust research into the added value of CLIL as to EFL proficiency and learner development is still scarce. As to the wider, international scope of CLIL little research has been undertaken. The innovative educational approach aims to include a global dimension in the curriculum to prepare students for life in a global economy, but the way in which these dimensions are embedded in the subjects need to be explored (Mannion et al., 2011). The question whether CLIL prepares its learners adequately for life and study in an internationalised market place has hardly been investigated, which means that school leavers in European countries have only vague notions as to what to expect in international English-medium studies. In order to put CLIL developments in

a wider perspective, the present study was undertaken in various European quarters that introduced Content and Language Integrated Learning into mainstream education.

Developments in higher education contributed to the importance of English as an academic language, a process accelerated by the signing in 1999 of the Bologna Declaration by Ministers of Education from 29 European countries. As a symptom of the ongoing internationalisation of curricula the number of English-taught programmes in higher education in 2007 at least tripled since 2001 and have seen a spectacular growth ever since. The numbers of identified English-taught programmes went up from 725 programmes in 2001 to 2,389 in 2007 and to 8,089 in 2014 (Wächter & Maiworm, 2014). Over the years, proficiency in the English language, the world's lingua franca, has turned out to be the tool for reaching the overall CLIL goal: promoting student mobility and educating confident L2 speakers and citizens for life in today's internationalised world. Moreover, and this is perhaps an unexpected but crucial benefit of CLIL, the use of a foreign language instead of the national language provides a challenge and a stimulus for teachers and learners alike, which has brought about a metamorphosis of didactics and educational practice. The introduction of CLIL is seen by many as a successful educational innovation and is presented as a promising concept for mainstream education in the 21<sup>st</sup> century (Marsh, 2013). However, up to the present day the extent to which CLIL contributes to a feeling of confidence in one's own EFL skills, of 'ownership' of the language, or to a feeling of 'international orientation', being at ease in an international audience, prepared for English-medium studies at a foreign university or for work abroad has not received sufficient attention. Many studies have discussed L2 gains within the context of the curriculum or school type, but without a further perspective on future career or studies in different parts of Europe and the rest of the world. Verspoor et al. (2015), for instance, tested receptive vocabulary and overall EFL skills at secondary schools in the Netherlands, with a focus on differences in performance between CLIL and mainstream learners; Dallinger et al. (2016) investigated skill development in subject content performance as well as EFL skills in German secondary schools, while Pérez-Vidal and Roquet (2015) investigated the effects of a newly introduced CLIL programme at secondary schools in Barcelona, to identify which areas of L2 competence benefitted the most from CLIL instruction.

It should be acknowledged that a fundamental innovation in mainstream education such as CLIL needs outcome-based information for further development, and that comparing various approaches across countries contributes to an insight into good practice which benefits both learners and teachers across countries. The aim of the present thesis is to investigate the effects of CLIL as practised in the early stages of its introduction, and is based on studies of a comparative nature. Data were collected between 2007 and 2009 in the first grades at secondary schools in four countries that

differ as to national language, presence of the English language on the curriculum and in society at large, CLIL implementation as to the number of content subjects involved, out-of-school activities promotive to language learning and teacher education. The research investigates EFL learning conditions and learner outcomes in the divergent CLIL contexts. It is clear that the CLIL approach has been received positively and with enthusiasm by its stakeholders - teachers, learners, parents and educational authorities. Even though much research has been done into CLIL benefits, its added value needs critical evaluation and its European scope needs a prominent place in CLIL literature. The aim of this thesis is to provide answers to the questions which learners profit the most from CLIL, what learner characteristics are promotive to successful L2 learning, if intensive CLIL programmes lead to better EFL results than moderate ones and if the presence of English in the media contributes to EFL acquisition. Across-language outcomes to side effects of CLIL add to the expertise necessary for further development of CLIL implementation and applying its didactics of effective teaching through a foreign language to a wider educational range, a development that is gaining ground in the European educational context (Marsh, 2013; Langé, 2014). With these concerns in mind the research was carried out in four European countries with divergent CLIL approaches at secondary schools preparing for university, involving comparable student populations as to educational level: the Netherlands, Germany, Italy and Hungary. Moreover, CLIL conditions and learner EFL outcomes in the wider European CLIL context of the past twenty years were investigated for inclusion in the present thesis.

A common factor in each of the various CLIL approaches is the teaching of content subjects, for instance history, geography, maths or biology through a foreign language, on the premises that the language used for interaction in the classroom and to describe subject matter will be processed more effectively and without having to invest additional lesson time (Coyle et al., 2010).

Another common factor is the fact that the target language is a foreign language, not the native language of the learners, who are basically beginners as to both the subject knowledge and the foreign language (Ball et al., 2015). Even though dual learning practice is not new - think of the Canadian example, where 'French immersion' has become the dominant L2 approach for Anglophone children, first investigated in the bilingual community in Quebec by Westhoff (1994), whose findings serve as a reference for European CLIL in its early stages - its large-scale implementation in mainstream schools all over Europe is. CLIL started as a highly selective programme in schools preparing the best and most motivated students for university. This condition applied to all schools involved in the present thesis, for which data were collected between 2007 and 2009. In later years the CLIL approach spread to various types of mainstream education, mostly secondary but also primary and vocational schools and admittance

became less elitist. Nevertheless, learner selection is still very much part of CLIL, either by means of school admittance criteria, parental preference or self-selection by pupils who feel attracted to the programme.

The teaching and learning of subject content through a foreign language is a core characteristic of CLIL teaching, but whether it is the content or the language that gets the most attention, the number of subjects, teacher linguistic skills and didactical practice, target language preparation of the learners, the amount of regular EFL lessons, extra-curricular activities, admittance criteria: these are all factors that vary per country, and depend on the educational context (Coyle et al, 2010). Each context has its own background which provides possibilities and limitations as to dual teaching (Eurydice Report, 2006). In countries that are very much English-orientated, learners will soon be familiar with the language side of CLIL, while the opposite is true for countries where the English language is largely absent: they will have to focus more on language and meaning. Having a closer look at CLIL conditions in the countries in which my research was conducted will illustrate this.

In **the Netherlands** the CLIL approach is accredited by the European Platform for Education - merged with Nuffic<sup>1</sup> in 2015 - which coordinates and monitors all CLIL schools that form the National Network by means of applying standards for bilingual education in English. Once in every five years a one-day accreditation visit takes place, in which schools are assessed on indicators pertaining to CLIL and educational standards (De Graaff & van Wilgenburg, 2015). The English language is very popular in the Netherlands and much present in everyday life (Berns et al., 2010). Even without formal training children become soon familiar with it, also because of its relatedness and similarity to Dutch. This offers extra possibilities for CLIL teaching: in line with the standards of the Platform, a large part of the curriculum is involved: in the lower grade levels a minimum of 50%, in upper grades 25%. Pupils in CLIL classes have no extra preparatory EFL training - they followed the same mainstream curriculum as their non-CLIL peers. All non-linguistic subjects may form part of the CLIL curriculum. On the whole the content teachers have Dutch as their first language and are qualified to teach their subject, but not the language. They need to give proof of sufficient EFL proficiency to teach their subject in English - level B2 on the Common European Framework of Reference (CEFR)<sup>2</sup> - even though schools try to involve native speakers of English in CLIL

- 
- 1 Nuffic is the Dutch organisation for internationalisation in higher education. The name of the new organisation after the merger is EP-Nuffic (<https://www.nuffic.nl/>).
  - 2 The Common European Framework of Reference for Languages (CEF or CEFR) was put together by the Council of Europe as a way of standardising the levels of language exams in different regions. It is very widely used internationally and all important exams are mapped to the CEFR.

teaching. Ideally some form of cooperation between language and content teachers is part of the CLIL implementation.

As in other parts of Europe, there is a growing belief in the Netherlands that CLIL education may also be beneficial for a larger population. Whereas CLIL first started as an elitist programme for selected pre-university students, the bilingual lessons are increasingly being offered to other secondary school types as well, including general education and lower vocational training (Mearns & de Graaff, 2018). Moreover, the issues of personal development, citizenship and intercultural competences will be receiving extra attention in the near future (Nuffic, 2018).

In the neighbouring country, **Germany**, the background of English-medium CLIL practice is rather different. CLIL started with French as the target language (Eurydice Report, 2006). Even though closely related to the national language German, the English language is less present in society at large and moreover, not as popular with young people as in the Netherlands (Berns et al., 2010). The English language is on the primary school curriculum and prospective learners of CLIL classes receive extra EFL training, but nevertheless the German CLIL approach is often characterised as cautious. There are generally only two English-medium subjects on the curriculum, history, geography or politics and the natural science subject biology (Prüfer, 2013). German CLIL teachers usually have a dual qualification, for their subject as well as the language. The lessons are supposed to be 'bilingual', meaning that the German language is used in addition to the target language. This practice is meant to provide a supportive role in the acquisition of content, however, German is used increasingly less (Rumlich, 2018). Developments in German CLIL reflect growth to include a broader audience. Breidbach and Viebrock (2012) mention the fact that the general perception of CLIL as a 'success story' has led school authorities to implement CLIL even in school types formerly assumed to be inappropriate for CLIL, such as primary schools, and that initiatives have been taken by local governments to make CLIL less elitist - to include children from less wealthy and migrant backgrounds.

The languages of countries in southern Europe are less related to English, moreover its presence used to be minimal. In **Italy** the place of the English language in mainstream education and society at large was almost non-existent around the turn of the century (European Commission, 2006). Italy was - and still is - a nation with low-EFL proficiency, geographically and linguistically remote from Great-Britain and the English language. The introduction of CLIL took place in a fragmentary way and without any precise direction, usually involving a limited number of subjects (Serragiotto, 2007). The Italian secondary schools with CLIL programmes involved in the present study are situated in the North, in the regions where CLIL was introduced first (Langé, 2014). They used the modular approach, in the form of a limited number of modules for certain

subjects during part of the school year. For the problem of teaching English-medium lessons by linguistically untrained subject teachers the schools followed the overall Italian CLIL practice of team-teaching, meaning EFL teacher presence in the subject lessons, providing language help and feedback.

Recent developments in CLIL practice relate to the introduction of English-taught programmes into a broader range of school types. To this end the Italian Ministry of Education launched a project in 2010 to make CLIL mandatory in some form in the last year of secondary high schools (*Licei* and *Istituti Tecnici*) in order to provide not only the selected CLIL groups but all pupils with extra EFL practice in preparation of life in the 21<sup>st</sup> century. (Langé, 2014). Its sudden implementation was hampered by a shortage of teachers, as all responsibility for CLIL implementation was placed in the hands of non-language content teachers. In order to be admitted to long-term CLIL training courses, focussing on both the target language and CLIL didactics, these teachers had to possess a language competence of at least B1 level on the CEFR (see footnote 2 on p. 14). The training project aims to provide an opportunity for content teachers to reach the level of language competence needed for CLIL, ideally C2 or C1 (Di Martino & Di Sabato, 2012).

Also in Eastern Europe the role of the English language used to be limited in the previous century. As described by Dörnyei et al. (2006), in **Hungary** the only foreign language on the school curriculum was Russian, until schools obtained more freedom in 1985 to introduce other European languages, mainly German and English. After Hungary joined the European Union in 2004 these languages also became the medium of instruction for secondary school CLIL subjects that amounted up to 50% of the curriculum. As pupils sometimes left primary school without any prior knowledge of the target language, most CLIL schools offered an intensive preparatory L2 learning year, in which the target language had to be mastered thoroughly for the CLIL demands of the following years. The sudden development as to the enhanced need for L2 and CLIL education brought about a shortage of teachers, remedied by an active recruitment of skilled native speaker teachers.

In recent years knowledge of foreign languages has become increasingly important in Hungary. Secondary schools still have the possibility to offer the preparatory 'language booster' year, while apart from English and German - the most frequent languages - also French, Italian and Spanish can be found in CLIL secondary education. CLIL is also increasingly being introduced in primary education, in programmes that are much in demand with parents of young children (Kovács & Trentinné Benkő, 2014).

Apart from sufficient EFL skills, CLIL teacher needs effective CLIL language pedagogy to teach his or her subject in a language that is not the usual medium of instruction (De Graaff et al., 2007). Teachers in CLIL need to have linguistic skills,



possessing insight into how language functions, in addition to being able to use the language as a tool in the classroom (Bentley, 2010). The CLIL classroom provides a real-life situation: pupils need the language in their efforts to understand the content of subject matter, the teacher needs it to explain lesson procedures and to teach his subject (Dalton-Puffer, 2007). It provides opportunities for the teacher to involve the students in the learning process of both language and content: the interaction between participants is of wider scope than in the formal language classroom. The language can be put to real and spontaneous use, with less inhibition on the part of the learners as - unlike in the language lessons - the language functions as a means of communication rather than being the aim of teaching (Ball et al., 2015; Cimermanova, 2017).

### **1.2 Previous research on effects of CLIL**

As from its early stages the CLIL approach has been the focus of research, notably into linguistic outcomes. Dalton-Puffer (2008, 2009, 2011) reported widely on L2 gains in divergent European contexts, mostly German-speaking countries in which positive effects were noted for receptive language skills, vocabulary and morphology as well as creativity, risk-taking, fluency and speaking confidence. Favourable effects mentioned in research frequently relate to EFL proficiency and indicate that CLIL learners achieve better scores than their mainstream counterparts; however, many studies are of a descriptive nature and have only one measure. In the Spanish context Lasagabaster (2008) compared vocabulary, grammar, fluency, listening, pronunciation of secondary CLIL and non-CLIL learners in the Spanish Basque country, a bilingual community in which both Basque and Spanish are official languages. The results showed better competence for the CLIL learners as to English, the third language included in the curriculum while hardly present in society at large. In 2017 Artieda et al. investigated the achievement of two groups of Catalan-Spanish intermediate learners of English in secondary school regarding receptive and productive L2 skills as well as grammatical knowledge. The group with extra L2 exposure in CLIL lessons was significantly better than the non-CLIL group in reading comprehension and in several dimensions of writing. Large-scale studies into linguistic skills of CLIL learners were also undertaken. In Spain a large group of learners in their second year at secondary school took part in a study conducted by Nieto Moreno de Diezmas (2016) to provide evidence on the effectiveness of CLIL as to the acquisition of English language competences (reading, writing, listening and spoken production and interaction) compared to traditional learning of English. To do so, results of CLIL and non-CLIL learners were examined and contrasted. Significant findings in favour of CLIL learners were found regarding spoken production and interaction, but measuring progress over a longer period of time was not part of the research. Another large-scale project was performed in southern Spain by Lorenzo

et al. (2010), comparing European CLIL with other, non-European bilingual education initiatives and focussing amongst other things on CLIL learner linguistic competence. A downside to this research was that there were no pre-test scores of the learners involved, no clear starting point so that progress was not measured and compared with that of control groups. It remained unclear which outcomes could truly be contributed to CLIL.

Even though many studies focus on the benefits of CLIL the research is often not directed towards longitudinal effects. Solid empirical studies, making use of a pre-test / post-test design and control groups have been sparse in the first decades of European CLIL. In the meta-analysis conducted in 2012, Pérez-Cañado presented a review of the way in which CLIL, the new educational approach had manifested itself on our continent. The survey pertained to countries in northern, central, eastern and southern Europe and discussed a broad field of cognitive, educational and affective variables affected by CLIL education, such as motivation, teaching practice, content learning and also linguistic merits as to EFL competence. The results unquestionably indicated that CLIL affects L2 language learning outcomes, though longitudinal studies were found to be scarce. In order to gain insight into the effectiveness of the CLIL approach as to linguistic outcomes, numerous authors called for increased prominence of longitudinal outcomes (e.g. Lasagabaster & Ruiz de Zarobe, 2010; Bruton, 2011; Pérez Cañado, 2016; Piesche, Jonkmann, Fiege & Kebler, 2016).

Also non-cognitive factors that were found to influence L2 learning such as motivation, attitudes towards the L2 speaking community, language aptitude and anxiety (Gardner, 2010) have been the focus of research in the CLIL field. Similar to research on CLIL effects on L2 learning only few studies have investigated long-term outcomes so that the relation between CLIL and affective variables promotive to L2 learning such as language aptitude, motivation, language confidence, identification with the English-speaking world remains unclear - in spite of the fact that positive affects towards language learning are often part of the selection criteria for student admittance to CLIL classes. In Spain the interaction between motivation and language achievement was investigated (Doiz et al., 2014; Navarro Pablo & García Jiménez, 2018) studies which demonstrated the motivating influence of learning in a CLIL class, albeit at only one point in time. Moreover, negative effects ascribed to difficulties involved in learning through a foreign language were also encountered, effects that were not monitored over a longer period of time so that no conclusions could be drawn as to the contribution of CLIL. In Finland Seikkula-Leino (2007) found CLIL to involve plenty of language beyond the pupils' current competence, which made it a demanding experience with the inherent possibility that pupils felt incompetent and doubtful as to their L2 skills, more so than in mainstream classes. An interesting question - that

was not part of her study - would be to see whether these feelings were overcome at later stages or remedied by different teachers. Lasagabaster and Doiz (2015) did arrive at longitudinal conclusions resulting from research analysing the impact of CLIL on different interrelated affective components such as motivation and confidence. The authors found that CLIL does not have long-term positive effects on students' motivation towards English language learning. Their initial motivation to learn the subject matter was maintained. The study was conducted in the Basque country, where English is a medium of instruction together with Spanish and Basque, a trilingual context that is not a standard CLIL background (Lasagabaster & Sierra, 2010). Different results might be obtained in other European countries, which focus entirely on bilingual CLIL practice.

### **1.3 The present thesis**

In view of the above, the aim of the present study is to contribute solid longitudinal outcomes to the CLIL research literature, as to both linguistic and affective variables, obtained not just in one country but across a variety of diverging educational settings across Europe. This was the reason underpinning our choice for the Netherlands, Germany, Italy and Hungary, countries with fundamental differences as to native language, involvement in the English language and CLIL implementation as discussed in the first paragraph.

The focus of the present research relates to the educational goals of CLIL, described in a report edited by David Marsh (2002) as promoting student mobility, and educating confident L2 speakers and citizens for life in today's internationalised world. The present study investigates whether CLIL has succeeded in contributing to these goals across divergent contexts in its first twenty years of existence by offering better EFL learning opportunities than mainstream teaching, enabling students to develop more confidence in their L2 skills when compared to their mainstream peers and preparing them better for the international world by involving them in the target language speaking community. Research showing that learners in CLIL classes are above-average performers as to language skills abounds, but as to the contribution of CLIL to personal development into citizens with positive attitudes towards other nations and speaking the lingua franca with confidence in a globalised world not much is known. Nor has much previous research been done as to the influence of individual learner variables on language learning success in CLIL classes. In short, the thesis investigated three research questions:

- To what extent does CLIL lead to better EFL skills than mainstream teaching?
- To what extent does CLIL contribute to learners' confidence in their EFL skills and international orientation?

- What is the influence of individual learner variables on language learning success in CLIL classes?

#### **1.4 Outline of the present thesis**

The research questions will be answered in four chapters. The following two chapters provide an answer to the question whether CLIL leads to better EFL proficiency than mainstream education and form a synthesis of our own and others' longitudinal research into effects of CLIL on EFL skills in the course of the past twenty years. Chapter 2 presents the findings of our longitudinal study into the effects of CLIL in its early forms - before 2010 - on EFL learning in the first two years at grammar school - the type of secondary school preparing for university. The outcomes are of a comparative nature, i.e. three of our four research countries with divergent CLIL types are compared: the Netherlands, Germany and Italy. In the previous paragraph we introduced four countries, however, Hungary was excluded from this part of the research as no matching control groups turned out to be present at the start of the first testing round. An overwhelming majority of new students had opted for the German or English-medium CLIL streams on offer and only very few for the mainstream classes. Moreover, mainstream first year classes appeared to accommodate a considerable number of drop-outs from CLIL streams, which made them unfit as valid control groups. In the remaining three countries four comparable groups took part: two groups in classes with a CLIL programme and two groups in mainstream classes as control groups. The scores for various written linguistic skills in the English language - vocabulary, idioms, grammar and text comprehension - of CLIL and mainstream pupils were compared by means of a pre-test / post-test design. With regards to the research outcomes, which showed differences in EFL gain scores and CLIL implementation across countries, it seemed a relevant concern to extend our field of study to comparable research in the rest of Europe conducted in both earlier and later years, for outcomes to be compared and contrasted. Chapter 3 presents a systematic review of longitudinal studies into the effects of CLIL on specified learner EFL skills, carried out by CLIL researchers in Europe in the past twenty years in primary and secondary mainstream education. Studies on the benefits of CLIL abound in European educational research, and seem to be unanimous in their positive findings as to its linguistic side: learners in CLIL classes perform better as to the target language than mainstream learners. However, the great majority of studies measure EFL performance at only one point in time, either at the start of CLIL or at some later moment, but without a follow-up so that it is impossible to monitor learner progress and compare it to the progress of mainstream learners. As it is our aim to gain an insight in CLIL learner development and come to solid conclusions as to

the added value of CLIL, we selected only studies with a pre-test / post-test design and the presence of a mainstream control group.

Chapter 4 provides an answer to the research question whether the early CLIL programmes implemented before 2010 had the potential to educate young people better for life in an internationalised world in which English is the lingua franca, and whether this holds good across contexts, which, as we have seen, diverge greatly as to involvement in EFL learning. It investigates if pupils in CLIL classes in their own specific national setting will come closer to the profile of the confident and internationally oriented learner than mainstream peers in the course of the first two years at secondary school. Again, Hungary had to be excluded because of the absence of mainstream control groups. The chapter investigates the question if having more confidence in one's EFL skills and feeling more involved in the international world is a specific feature of learners that have chosen for CLIL across European cultures, and if their mainstream peers possess these characteristics to a lesser extent, so that they may be attributed to selection criteria. The contribution of two years of CLIL would lie in the enhancement of these affects to a greater extent than mainstream education.

Chapter 5 concentrates on individual differences in language learning and is a comparison of the effects of divergent CLIL approaches in four countries, before 2010: the Netherlands, Germany, Italy and Hungary. In view of the fact that control groups were absent in the Hungarian schools the study was carried out without a comparison with mainstream learners. It focusses exclusively on the question if the better EFL performance of pupils selected for CLIL in the four diverse educational settings can be attributed to certain learner variables interacting with the CLIL intervention - a relevant issue in recent criticism on CLIL selectiveness. The effects of three learner constructs on EFL proficiency progress are investigated - again by the pre-test / post-test design - 'EFL aptitude', aptitude to learn the English language, 'EFL confidence', the confidence with which pupils use the L2 in the classroom and 'international orientation', considered to encompass a general interest in other languages and cultures, a sense of identification with target language speakers and a desire to use that language for international communication, study and work . In addition, the effects of the presence of the English language in everyday life were taken into account, often seen as promotive to language learning but not the same in all countries, so that some might profit more than others.

Finally, the last chapter, Chapter 6, summarizes the contents of the present thesis and discusses the results. Based on our discussion of previous research we will look at future directions as to the role of CLIL in education and discuss what aspects need further attention.





# 2

## CHAPTER

Effects of the Content and Language  
Integrated Learning approach to EFL  
teaching: A comparative study



## Abstract<sup>1</sup>

This study investigates the effects of English-medium CLIL on EFL proficiency in three European countries. Seven mainstream grammar schools spread across the Netherlands, Germany and Italy participated with a total of 263 pupils aged 12 to 16. Several language skills, viz. receptive vocabulary, idioms, grammar and text comprehension were measured by means of written tests in a pre/post- test design and the use of control groups. The first test was held when the experimental classes were all at the start of the CLIL intervention in secondary education; the second test two years later. Each of the three countries participated with two CLIL classes and two mainstream classes as control groups. The results show that CLIL classes had better EFL skills from the start: they outperformed the mainstream groups at both test rounds. However, gain scores varied per country and the apparent head start of CLIL pupils makes it difficult to interpret EFL results.

*Keywords:* bilingual learning, EFL innovations, CLIL in Europe, CLIL effects, English-medium learning.

### Reference:

Goris, J., Denessen, E., & Verhoeven, L. (2013). Effects of the Content and Language Integrated Learning approach to EFL teaching. *Written Language & Literacy* 16(2), 186-207.

---

1 Data for this chapter were collected between 2007 and 2009.

## 2.1 Introduction

The teaching of English as a foreign language (EFL) has become increasingly important in most types of education. The Eurydice network on education systems and policies in Europe states that English is by far the most taught foreign language in nearly all countries at all educational levels; the 2012 report mentions an increase in the percentage of pupils learning English at all educational levels since 2004/2005, particularly at primary level. In 2009/10, on average, 73 % of pupils enrolled in primary education in the EU were learning English. In lower secondary and general upper secondary education the percentage was higher than 90 %, while in upper secondary pre-vocational and vocational schools it reached 74.9 %.

Moreover, developments in higher education during the last few decades have contributed to the importance of English as an academic language (Ruiz-Garrido et al., 2010), a process accelerated by the signing in 1999 of the Bologna Declaration by Ministers of Education from 29 European countries. As a symptom of the ongoing internationalisation of curricula the number of English-taught programmes in higher education in 2007 at least tripled since 2002 (Wächter & Maiworm, 2008). In line with this increased importance of the English language for international study careers secondary schools, notably those preparing for studies in higher education, have adopted enhanced and innovative approaches to the teaching of EFL. One innovation in particular - Content and Language Integrated Learning, known under the acronym CLIL - has become popular and will be discussed in this paper. Though not primarily intended as an instrument for language learning - there is also the content side, of equal weight - CLIL already proved its merits in the field of EFL acquisition when it was first introduced at grammar schools in the Netherlands (Huibregtse, 2001).

In theory a CLIL target language could be any foreign language and does not necessarily have to be English; however, in the European context it almost always is (Dalton-Puffer, 2011) which gave rise to the focus of this study. Our aim is to investigate the EFL benefits of several CLIL approaches in the first two years, when pupils with only little EFL knowledge from different countries and with different native language backgrounds are at the start of English-medium learning and to compare EFL pupil performance. We restricted ourselves to pre-university grammar school practice, as at the time CLIL was most often and in some countries exclusively practised in this school type. We selected three countries that answered our criteria: the Netherlands, Germany and Italy, three countries with diverse CLIL approaches and pupils more or less of the same age, though the Italian pupils were two years older than those in the Netherlands and Germany when their CLIL lessons started.

## 2.2 CLIL as an innovative way of teaching EFL

Even though CLIL is more than a way of foreign language learning - its object is dual focussed: acquiring content through a foreign language, aiming at integration of form and function, with its own specific didactics (De Bot, 2002) - research projects in CLIL contexts used to exhibit a strong focus on issues of foreign language acquisition or linguistic competence. This may be explained by the fact that the first large-scale CLIL research projects were conducted by trained linguists (Breibach & Viebrock, 2012) who were interested in the English-medium content lessons as a source of L2 exposure and an opportunity to put language into practice to an extent that could not be equalled in the mainstream EFL lesson, a new and welcome phenomenon.

By the early 1950s, foreign language teaching had become one of the educational priorities of European governments. Its primary goal - practising translation exercises and teaching formal language rules to learners aspiring to a job in international business - shifted in the early seventies. In 1971 the Council of Europe commissioned a project to produce a system of units in foreign language instruction for adult learners, 'those who would wish to be able to communicate non-professionally with foreign language speakers in everyday situations on topics of general interest' (Van Ek, 1976, pp. 1-2). The specification of what this group of adult learners would minimally have to be capable of doing in the foreign language led to the development of an objective for foreign language learning along the same lines in compulsory secondary education. In the following decades the ability to use real, appropriate language to communicate with others came to be seen as the primary goal of most foreign language learning methods (Ellis, 1985; Krashen, 1982, 2002). The functional-notional approach, introduced in the 1970s, provided a basis for practising communication within the classroom (Shehadeh, 2005; Wilkins, 1976), even though situations relied largely on role play. The CLIL approach carries communication a step further: content or non-language subjects are taught with a foreign language as the medium of instruction. The CLIL classroom actually provides a real-life situation, a naturalistic context for those involved (Dalton-Puffer, 2007) in which the language is put to real use.

Other factors seen as CLIL success factors are the longer period of exposure, better learning conditions owing to authentic lesson materials, the presence of native speakers, extra EFL lessons and richer linguistic content of the CLIL classes (Wolff, 1997). Marsh (2002) also mentions egalitarianism (p.10) as a beneficial factor in view of the fact that the CLIL approach is seen to open doors on languages for a broader range of learners, notably in the field of early language learning and vocational education, in which both below average and above average ability learners benefit from exposure. However, other research (Weenink, 2005) points out that the admittance policy of secondary schools, in particular those preparing students for university, often includes

strict selection: their CLIL classes only admit motivated and academically bright pupils. Bruton (2011) holds this to be a likely contribution to the better EFL results of CLIL learners, as their talents compare favourably with the more average mainstream performance.

### 2.3 English-medium CLIL in Western Europe

The introduction and increasing use of English-medium CLIL programmes into mainstream secondary schools across Western Europe is a fairly recent development: it started off on a limited scale in the late 1980s or early 1990s (Eurydice, 2006). In most countries pupil admittance to the CLIL programme is selective and based on a combination of sufficient motivation and EFL skills. However, as to contents and implementation CLIL programmes show a great variety across nations. Their diversity lies mainly in CLIL history, curriculum content, teacher training and national or regional coordination.

Each of the three countries in the present study has its own developmental path towards English-medium CLIL. In the Netherlands the roots are in international education; in 1989 one of the international secondary schools opened a bilingual department for Dutch students, who could follow the regular curriculum partly in Dutch and partly in English. This initiative led to the successful introduction of *tweetalig onderwijs* (literally: bilingual education) into mainstream secondary schools, mainly grammar school classes. Since its introduction the number of CLIL schools, almost exclusively with English as the language of instruction, has risen sharply: in the 2009/2010 school year there were 112 schools with CLIL grammar school classes. Recently also schools for general secondary and vocational training have started preparations for CLIL, predominantly English-taught (European Platform, 2013).

In Germany the switch to English-medium CLIL was made gradually; the phenomenon of CLIL – *bilingualer Fachunterricht* (literally: bilingual subject teaching) started off in mainstream secondary education as early as the sixties with French as the target language, to facilitate student exchange projects with France. Other frequent CLIL languages were those of adjacent countries, practised in areas close to the borders and Italian in areas where many Italian workers were present (Eurydice Report, 2006). However, since the early 1990s the CLIL target language has increasingly become English, which is nowadays used the most. Breidbach and Viebrock (2012) mention that the general perception of CLIL as a ‘success story’ has led school authorities to implement CLIL even in school types formerly assumed to be inappropriate for CLIL, such as primary schools, and that CLIL has clearly reached the political agendas of local governments as can be seen in programmes such as the North Rhine-Westphalian ‘CLIL for all’.

In Italy, whose plurilingual environment has of old favoured bilingual teaching, the concept of dual teaching traditionally refers to different subjects being taught in one or two different languages: the official state language of instruction and a second regional language. The term most frequently used to explain the English acronym CLIL, which is now commonplace in Italy is *insegnamento veicolare* (literally: vehicular teaching). It refers to a learning environment in which language and content are combined, by alternating use of the school's official language of instruction and the foreign language concerned, during the last decade increasingly English. At the start of the present study English-medium CLIL in the participant Italian schools was still in its early stages: both schools had initially started off with a limited number of English-medium subjects in several modules spread across the school year. In the meantime there have been drastic changes in Italian English-medium CLIL, which was made compulsory in secondary schools, but as these developments do not affect the outcome of the present study they will not be discussed here.

Table 1: *English-medium CLIL in the Netherlands, Germany and Italy: general characteristics.*

CLIL	The Netherlands	Germany	Italy
Year of introduction	1989	1990	1998
English-medium instruction	50% of the curriculum	15% of the curriculum	Modular approach
Subjects in CLIL	All non-language subjects	History, Geography, Politics	All subjects except modern foreign languages
Pupil selection criteria	Good overall performance and motivation	Good overall performance and motivation	Entrance test
Pupil age at the start of CLIL at grammar school	12	12	14
Initial pupil EFL proficiency <sup>2</sup>	English as a subject in grades 5 and 6 of primary school	English as a subject in primary grades 5 and 6 (Eastern states) or in secondary grades 1 and 2 (Western states)	English as a subject for three years at the 'scuola media'.
Pupil preparation for CLIL	None	Additional EFL lessons	Additional EFL lessons

2 Many primary schools in the three countries have some form of early EFL learning in the lower grades, involving e.g. singing English songs and the first steps of speaking. By 'English as a subject' we mean the systematic teaching in a structured way of all four language skills: speaking, listening, reading and writing.

Table 1: *Continued*

CLIL teachers	Qualified to teach the subject matter	Qualified to teach the content subject as well as EFL as a subject	Qualified to teach the subject matter
---------------	---------------------------------------	--	---------------------------------------

As can be seen in Table 1, national practice as to CLIL and pupil demands vary. As to curriculum characteristics, the CLIL approach in the Netherlands comes closest to the immersion approach: 50 to 60% of the curriculum is taught with English as the language of instruction from the start of Year One at grammar school. CLIL practise in Germany varies as its sixteen states each have their own practise, but on the whole the German approach is cautious, with about 15 % of English-medium lessons in only two or three subjects, often supported by lessons in the mother tongue. As in the Netherlands, in Germany CLIL subjects are taught with English as a medium of instruction throughout the school year. Italy employs the modular approach, which implies that CLIL modules involving a series of lessons in alternating subjects are implemented in the course of one or more years; a subject is not taught completely in English throughout the school year. In the first years the English-medium modules take up about 10 to 20% of the curriculum; in later years the percentage gradually increases. A remarkable feature of Italian CLIL is the presence of languages as a CLIL subject; Italian, the pupils' mother tongue, and Latin are implemented in modules suitable for CLIL: history of literature and Roman civilization. The Netherlands offers almost all non-language subjects in the CLIL programme, including mathematics and science, while in Germany predominantly history, politics, geography and sometimes sports are found. In addition to English-medium lessons there are out of school activities in which the English language can be practised: Dutch, German and Italian CLIL pupils all take part in language excursions. In the Netherlands and Germany CLIL classes generally receive one extra EFL lesson on top of what is regular in the mainstream classes.

Teacher training varies across the three countries; not all content teachers are equally well equipped to teach their subject in English. In Germany teachers in secondary education usually have a dual qualification, which enables them to teach two subjects. In order to qualify for CLIL this has to be a combination of English and a non-language subject, providing a sound basis for English medium instruction. Dutch and Italian teachers on the other hand are generally qualified to teach only one subject; prospective CLIL content teachers have to acquire a sufficient EFL level by means of extra training and courses. In support of the CLIL lessons Italy has the team-teaching approach, in which the content teacher and the EFL teacher cooperate in the classroom, the former focussing on content, the latter on language.

In each of the three countries the financial position of CLIL teachers is the same as that of mainstream teachers. In the Netherlands and Italy they get compensation by means of extra time, a benefit that was abolished in Germany. In general the switch to CLIL did not result in employing new content teachers in any of the three countries; the existing staff received training for CLIL teaching. National CLIL practice follows certain rules. Dutch schools show considerable uniformity as they are under supervision of The European Platform for Education in the Netherlands, which has set certain standards that have to be met before a school is an acknowledged CLIL school. Italian schools are in regional networks under the supervision of their Local Educational Authorities and CLIL practice varies according to region. In Germany certain CLIL stipulations are laid down in curricular guidelines, such as the right of pupils to have CLIL contents at their disposal not only in the foreign language, but also in their mother tongue.

#### **2.4 The present study**

The CLIL approach as a way of training English as a Foreign Language has already proved its merits, as shown by research. A longitudinal study in the Netherlands (Admiraal, Westhoff & De Bot, 2006) showed that pupils in CLIL classes in the first four years of secondary school had higher EFL scores in terms of oral proficiency and reading comprehension when compared to a regular control group, while Huibregtse (2001) found similar results. In the Basque country Lasagabaster (2008) found an improvement of writing and pronunciation in favour of CLIL pupils, while in the context of German speaking countries Dalton-Puffer (2008) reported favourable effects of CLIL on receptive skills, vocabulary, morphology, fluency, creativity and affective outcomes whereas syntax, writing, informal language, pronunciation and pragmatics were unaffected or indefinite. More recently (2011) she found that writing skills have received increased attention in current studies, whose findings concur in that 'CLIL students have at their disposal a wider range not only of lexical but also morphosyntactic resources, which they deployed in more elaborate and more complex structures' (p. 186). In Berlin Zydati (2012) found that the linguistic competence of pupils in CLIL classes - especially regarding vocabulary and grammar and/or general proficiency - has a major impact on task performance in the content subject matter and on pupil motivation to persevere.

Even though almost every country in Europe has to a greater or lesser extent adopted the innovative approach and CLIL has become a much researched area, the implementation of CLIL and pupil progress in diverse contexts have not yet been compared. In the present study English-medium CLIL programmes and their effects on the acquisition of pupil EFL skills will be explored in the diverse contexts of the Netherlands, Germany and Italy. The research addresses the following question:

· To what extent does CLIL in the Netherlands, Germany and Italy produce better EFL results than mainstream teaching?

The study has a naturalistic design and aims to investigate how pupils in their own specific national setting perform as to EFL learning. In each of the three countries four groups take part: two groups in classes with a CLIL programme and two groups in mainstream classes. The groups are all at mainstream secondary schools that can be described by the British term 'grammar school', that is to say secondary schools preparing for an academic career, in which an excellent command of the English language has become essential. The study concentrates on vocabulary, grammatical accuracy and text comprehension, and is aimed at the effects of CLIL on pupil EFL acquisition in their first two years in the CLIL stream. The degree of progress is measured by evaluation of pre/post-test results.



## Method

### 2.5 Participants

The Netherlands has two types of education preparing for university: the classical *Gymnasium* and the grammar school section of what is known here as ‘comprehensive school’, a school which, apart from pre-university education, also offers general secondary and lower secondary professional education. Three schools, located in the southern part of the country, took part in the present study: two ‘comprehensive’ schools with both mainstream and CLIL grammar school departments and one non-CLIL classical *Gymnasium* with an enhanced EFL curriculum. The classical *Gymnasium* took part in the research with one class as a control group. One of the ‘comprehensive’ schools took part with a CLIL class, the other one with a CLIL class and a class from the mainstream department as a control group. Both CLIL schools were accredited by the Dutch European Platform.

Germany participated with two schools from different states, both *Gymnasium* schools with mainstream and CLIL departments. Each took part in the research with one CLIL class and one class from the mainstream department as a control group. One of the schools is situated in Lower Saxony, the second in Berlin, the only school in this study with pupils of various international backgrounds and CLIL experiences at primary school.

In Italy the participant schools are both *Licei Scientifici*, five-year academic mainstream upper secondary schools, which had CLIL modules in all school years. Both schools belong to the regional CLIL network for Friuli –Venezia – Giulia.

Table 2: *Distribution of pupils (N=263)*

Country	N	Initial age	CLIL		Control	
			boys	girls	boys	girls
<i>The Netherlands</i>	84					
School 1	20	12.4	9	11		
School 2	22	12.3			11	11
School 3	42	12.5	10	7	14	11
<i>Germany</i>	97					
School 1	50	12.7	16	13	12	9
School 2	47	12.6	8	13	13	13
<i>Italy</i>	82					
School 1	34	14.3	16	4	10	4
School 2	48	14.4	14	11	6	17

In each of the three countries there were four groups of pupils: two CLIL groups and two non-CLIL control groups, distributed as shown in Table 2. The great majority, 96%, were born and raised in their present country and speak its language as their first

language. None of them had English as a native language; all of them had received EFL training in previous school years, though not quite to the same extent.

In the Netherlands this previous EFL training took place at primary school. At most schools EFL is part of the curriculum during the last two years and is the same for all; this meant that pupils in the CLIL classes had had the same EFL training as those in mainstream classes.

Germany is made up of sixteen states, two of which are represented in the present study; in the state of Lower Saxony pupils start secondary education at the age of ten, as is general practice in most states in the western part of the country. From this moment on prospective CLIL pupils receive additional EFL lessons on top of the normal secondary school curriculum in preparation of the CLIL programme that starts two years later. In the state of Berlin primary education continues until the age of twelve; EFL teaching takes place in the last two years. Moreover, in this highly internationally populated area English-medium CLIL at primary school is not uncommon; consequently, EFL levels of the Berlin pupils in this study varied as several had had earlier CLIL training or an international background.

Italian pupils leave primary school when they are eleven and then go to the *Scuola Media* for three years, a type of middle school providing general education, the same for all. This means that Italian pupils are fourteen when they first start their pre-university learning at the *Liceo*. At the *Scuola Media* EFL is a compulsory subject while at some schools prospective CLIL candidates get extra EFL training on top of the normal curriculum.

The participating schools are selective as to their admission of pupils to the CLIL classes<sup>3</sup>. In the Netherlands there is a standardised achievement test (S.A.T.) at the end of primary school after which streaming takes place: only those with very good overall results go to grammar school and for a CLIL career pupils have to demonstrate sufficient motivation in personal interviews. In the German Lower Saxony pupils are admitted to the CLIL stream on grounds of motivation, whereas in Berlin there is an entrance test, except for those who enter from CLIL primary schools. In the Italian schools prospective CLIL pupils also had to complete an EFL entrance test.

The schools in this study show considerable differences as to years of experience with CLIL: the school in Lower Saxony started in the year 1991, one of the Italian schools in 2003. The number of lessons and subjects taught via CLIL were also very diverse, while some schools - but not all - had extra EFL lessons on top of the regular number. Apart from formal instruction in English all CLIL pupils engage in a variety of out of school language related activities<sup>4</sup>.

3 For details see appendix 1

4 For a detailed survey see appendix 2

## 2.6 Instruments

To measure pupil EFL skills 263 pupils completed a written English test in their third week at grammar school (pre-test, Test 1) and another more advanced test (post-test, Test 2) towards the end of their second year. Language competence is a multi-faceted concept, a broad term. Our test was to be in written form; in order to evaluate a range of skills in the time available we decided upon testing primarily receptive knowledge, but to a lesser degree also productive skills. Our keynote for the testing procedure was that L2 proficiency is an integration of knowledge of words, expressions, insight into the rules of the language and an ability to understand its written texts. We selected assignments of receptive Vocabulary, Grammar, Idioms and Text Comprehension, while the testing of productive skills in Grammar, Idioms and Vocabulary coincided in one assignment in which errors had to be corrected.

In view of the fact that pupils across the three countries had not had quite the same amount of EFL input uneven levels at the start of grammar school were taken into account, which is why the English tests were made up of assignments from various test levels. Based on interviews with EFL teachers in the three countries as to the general EFL levels of pupils entering grammar school and the testing formats they were familiar with we made use of diverse sources and levels; assignments were taken from exams at Preliminary, Elementary and Pre-Intermediate levels; for the second test from Intermediate and Proficiency Levels.

The assignments for grammar and text comprehension were part of standardized tests by the British Anglia Examination Syndicate (<http://www.anglia.nl/>), an organization which offers internationally accepted and recognised examinations and awards Certificates in English for Overseas Candidates at levels ranking from First Step till Masters. The first test comprised 66 and the second test 20 items for grammar, made up of both multiple choice items and blanks where pupils had to fill in the correct form of a word, e.g. plurals. Text comprehension assignments were made up of open and multiple choice questions and blanks: three open questions and five multiple choice questions in Test 1 and Test 2, and four additional blank items in Test 2, in which pupils are asked to fill in a synonym from the text to a given word. The 'English as a Foreign Language Vocabulary Test' developed by Paul Meara (1992) was used to measure receptive vocabulary. The test consists of a series of word strings with 40 real words and 20 pseudo-words per section. Pseudo-words were non-existing items that followed the phonotactic rules of English: they look like real words but are in fact non-existent. The real words were drawn at random from lists of words whose level of difficulty was based on the frequency of appearance in English texts, composed by Hindmarsh (1982) and Nation (1986). Pupils completed three sections of increasing difficulty, in which they were asked to put a cross before each word they knew; no hints were given as to the

presence of pseudo-words. 60 items of the highest level in Test 1 were used again in Test 2; the remaining 120 in Test 2 were of a more advanced level.

The assignments for idioms were taken from the ESL/EFL Test Net (<http://www.english-test.net/esl>), a testing source on the Internet for EFL teachers, which provides tests for various linguistic skills at levels ranging from elementary to advanced. In Test 1 and Test 2 there were six items for idioms, of elementary level in Test 1, of advanced level in Test 2. In order to assess not only receptive understanding, but also the productive use of grammar, idioms and vocabulary an assignment was added in the form of an error correction assignment at pre-intermediate level. Ten written sentences with a grammatical or idiomatic error were presented, and pupils were required to identify this error and replace it by correct forms. The assignment was presented to the pupils as follows:

*In these sentences one of the underlined parts is not correct. Choose the wrong item and correct the mistake. Note: There is only **one** wrong item in each sentence!*

Example: I like elephants and giraffes very much but I like not lions.

like .....

much: .....

X I like not: ..... ***I don't like*** .....

The error correction assignment was part of both tests.

## 2.7 Procedure

At the start of the school year 2007/2008 (Test 1) and again near the end of the school year 2008/2009 (Test 2) the tests were presented. The pupils were informed about their participation in the university research project. The test was introduced as a means to investigate the way in which young people at grammar schools in diverse European countries learned English. The participants were told that their test results only served research purposes, by no means would affect their marks at school reports and that participation was anonymous. Completion of the test was to take one lesson period; the assignments were self-explanatory. Subsequently the pupils completed the EFL test and handed them over to the test leaders, mostly their English teacher in cooperation with the university researcher.

## 2.8 Analysis

To get insight in the results several ANOVAs were conducted. To begin with, an ANOVA with Group (CLIL versus Control) and Country (the three participating countries) as between-subject factors was conducted after the pre-test (Test 1) to explore differences at the first moment of measurement. To find out the role of the CLIL intervention additional ANOVAs with Time as a within-subject factor and Group

and Country as between-subject factors were conducted to compare the results at the first with those at the second testing moment, both within and between countries.

## Results

### 2.9 Effects of CLIL across countries

In order to answer the research question the English test batteries Test 1 (pre-test) and Test 2 (post-test) were scored on the subparts: Vocabulary, Grammar, Idioms, Error Correction and Text Comprehension. The score for Vocabulary is a combination of the number of correctly marked real words and the correctly ignored pseudo-words. The Text Comprehension question results were subdivided into scores on open questions, for which pupils had to frame their answers in correct sentences, and multiple choice items and blanks combined. As the number of items in both tests was not identical for the subparts grammar and multiple choice text questions mean item scores were computed for the analysis.

Table 3: Test Scores for all Dutch, German, and Italian pupils.

Variable	CLIL group				Control group			
	Test 1		Test 2		Test 1		Test 2	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Vocabulary	129.55	29.67	142.84	20.61	118.18	19.63	124.65	18.66
Correcting errors	9.33	4.37	15.04	3.43	6.88	3.63	11.99	3.83
Grammar	0.65	0.15	0.67	0.18	0.50	0.14	0.44	0.18
Idioms	2.27	1.17	3.07	1.43	2.09	1.24	2.12	1.27
Text open questions	3.83	1.71	3.89	1.86	2.86	1.62	2.31	1.60
Text multiple choice	0.73	0.28	0.81	0.20	0.66	0.26	0.63	0.26

Table 4: Test scores for the pupils, itemized for country.

	CLIL group				Control group			
	Test 1		Test 2		Test 1		Test 2	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>The Netherlands</i>								
Vocabulary	136.65	16.63	156.38	10.51	118.66	24.98	134.62	18.50
Correcting errors	7.73	3.81	16.32	1.44	6.77	3.62	13.60	3.01
Grammar	0.57	0.14	0.72	0.09	0.47	0.15	0.53	0.17
Idioms	2.60	1.21	3.97	0.87	2.51	1.25	2.85	1.35
Text open questions	2.95	0.82	3.84	1.57	2.04	1.23	2.53	1.41
Text multiple choice	0.80	0.21	0.90	0.12	0.69	0.27	0.81	0.20
<i>Germany</i>								
Vocabulary	135.36	24.25	141.42	20.42	114.47	15.59	120.64	18.64
Correcting errors	10.46	4.98	13.60	3.72	6.98	4.07	10.70	4.23
Grammar	0.69	0.16	0.57	0.22	0.53	0.12	0.37	0.20
Idioms	2.12	1.22	2.60	1.54	1.83	1.19	1.77	1.15
Text open questions	4.42	1.34	3.02	1.85	3.34	1.58	2.51	1.73
Text multiple choice	0.84	0.17	0.70	0.25	0.70	0.19	0.62	0.21
<i>Italy</i>								
Vocabulary	117.27	38.78	133.29	21.46	122.30	15.77	117.08	12.88
Correcting errors	9.38	3.69	15.58	3.72	6.89	3.09	11.60	3.59
Grammar	0.67	0.13	0.74	0.15	0.48	0.14	0.41	0.09
Idioms	2.18	1.03	2.84	1.35	1.89	1.17	1.65	0.86
Text open questions	3.89	2.27	4.89	1.60	3.30	1.73	1.78	1.58
Text multiple choice	0.57	0.34	0.86	0.13	0.55	0.31	0.44	0.24

The ANOVA of the pre-test indicated several significant between-subjects effects, as can be seen in Tables 3 and 4. For Vocabulary significant test results were found for group:  $F(1,257) = 13.79$ ,  $p < .001$  and for the interaction between country and group:  $F(2,257) = 7.14$ ,  $p = .001$ . It appeared that in the Netherlands and Germany the CLIL group had higher Vocabulary scores than the control group, in contrast to Italy, where the reverse was observed. Also the CLIL groups in the Netherlands and Germany scored higher than the CLIL group in Italy. For Idioms significant differences between countries were found:  $F(2,257) = 6.191$ ,  $p = .002$ . Dutch pupils scored higher than their counterparts in Germany and Italy. The test results for Grammar, Error Correction and Text Comprehension showed significant differences both between countries (Grammar:  $F(2,257) = 10.052$ ,  $p < .001$ ; Error Correction:  $F(2,257) = 3.087$ ,  $p = .047$ ; Text Open Questions:  $F(2,257) = 18.883$ ,  $p < .001$ ; Text Multiple Choice:  $F(2,257) = 16.828$ ,  $p < .001$ ) and groups (Grammar:  $F(1,257) = 72.154$ ,  $p < .001$ ; Error Correction:  $F(1,257) = 31.983$ ,  $p < .001$ ; Text Open Questions:  $F(1,257) = 19.455$ ,  $p < .001$ ; Text Multiple Choice:  $F(1,257) = 7.285$ ,  $p = .007$ ). Dutch pupils appeared to score relatively low on Grammar,

Error Correction and Text Open Questions, whereas the Italian pupils scored relatively low on the Multiple Choice questions test. Overall the CLIL groups scored higher than the control groups.

The ANOVAs in Tables 5 and 6 show complex patterns of significant effects between countries and groups as well as within countries. Gain scores were highest in Italy and lowest in Germany, where several tested language skills showed negative effects for the CLIL groups. The scores of the CLIL groups were higher than those of the control groups at both testing moments. Pupils in the Netherlands and Germany started off with about equal results on Vocabulary, but the Dutch pupils showed better progress over time; the Dutch CLIL pupils more so than the control group. The Italian CLIL pupils started off with a slightly smaller vocabulary than shown by their mainstream peers, but they progressed considerably whereas the control group declined.

As to Grammar the German and Italian CLIL pupils started with better results than the Dutch, but at the second test the Dutch CLIL results were about equal to those of Italy. The Dutch CLIL group showed significantly better progress compared to the control group. The German Grammar results of both the CLIL and control groups declined, whereas in Italy this was so for the control group. Dutch scores on Idioms were highest at both tests for both CLIL and control groups; the Dutch CLIL group showed the best progress, while German and Italian test scores declined for their control groups.

The results of the assignments in which errors had to be corrected showed considerable improvement for all groups. The initial scores of Germany were highest in the Berlin class; their increase, however, was considerably less than that of the Dutch and Italian pupils, while the CLIL gain scores were about the same as those of the control group.

Text Comprehension scores for the open questions depended not only on providing a comprehensibly correct answer, but, as was indicated in the assignment, also on framing this answer in a correct, full sentence. Initially the German CLIL pupils were best at this, but their precision declined so that gain scores were negative. The Italian CLIL group on the other hand showed the highest gain scores and indeed many answers were impeccable. Answers framed in incomplete sentences that were correct from the view of comprehension were found in all groups, but the most in the Dutch classes. The multiple choice questions presented very few difficulties for the Dutch classes, but the German scores decreased over time. In Italy gain scores were divided: those of the CLIL classes developed very positively, those of the control classes negatively.



Table 5: ANOVA Repeated Measures Test 1 and Test 2

	df <sup>1</sup>	Vocabulary			Grammar			Idioms			Correcting errors			Text open quest.			Text m.choice		
		SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>
WS: Time	1	12419.69	.000	.173	0.24	.126	.009	23.57	.000	.075	3967.49	.000	.642	6.68	.060	.014	.107	.111	.010
BS: Group	1	30883.17	.000	.149	4.88	.000	.343	52.33	.000	.101	985.35	.000	.167	202.31	.000	.194	2.59	.000	.138
Time x group	1	2233.08	.002	.036	.283	.000	0.96	22.36	.000	.072	25.84	.084	.012	19.82	.001	.040	.403	.002	.036
BS: Country	2	16608.53	.000	.086	.125	.182	.013	86.94	.000	.158	41.47	.339	.008	35.72	.005	.041	3.07	.000	.159
Group x country	2	6151.34	.012	.034	.231	.043	.024	.752	.812	.002	52.62	.253	.011	25.36	.022	.029	.381	.050	.023
Time x country	2	4117.48	.000	.065	1.38	.000	.342	11.93	.005	.040	409.14	.000	.156	72.80	.000	.132	1.41	.000	.116
Time x group x country	2	2759.29	.003	.044	0.47	.102	.018	1.497	.515	.005	37.75	.113	.017	53.85	.000	.101	1.36	.000	.112

<sup>1</sup> df<sub>error</sub> = 257

Table 6: ANOVA Repeated Measures Test 1 and Test 2 per country

	df <sup>1</sup>	Vocabulary			Grammar			Idioms			Correcting errors			Text open quest.			Text m.choice		
		SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>	SS	p	η <sup>2</sup>
<i>The Netherlands</i>																			
WS: Time	1	13183.01	.000	.442	.437	.000	.342	30.58	.000	.283	2462.67	.000	.817	19.75	.000	.167	.512	.000	.178
BS: Group	1	16355.88	.000	.282	.892	.000	.267	15.05	.006	.087	141.12	.001	.117	50.53	.000	.223	.392	.013	.073
Time x Group	1	147.30	.397	.009	.085	.005	.092	11.15	.001	.126	32.24	.031	.055	1.68	.241	.017	.003	.764	.001
<i>Germany</i>																			
WS: Time	1	1811.90	.000	.170	.996	.000	.501	2.10	.192	.018	570.62	.000	.404	60.23	.000	.268	.629	.000	.167
BS: Group	1	21037.23	.000	.237	1.637	.000	.248	15.31	.008	.072	492.94	.000	.158	30.58	.005	.082	.559	.002	.099
Time x Group	1	.147	.968	.000	.029	.101	.028	3.58	.090	.030	4.12	.497	.005	3.94	.135	.023	.034	.314	.011
<i>Italy</i>																			
WS: Time	1	1185.50	.098	.034	4.83	.983	.000	1.82	.220	.019	1206.81	.000	.596	2.68	.324	.012	.320	.029	.058
BS: Group	1	1268.34	.224	.018	2.568	.000	.569	22.29	.000	.172	424.87	.000	.262	138.74	.000	.303	1.951	.000	.240
Time x Group	1	4579.47	.001	.119	.202	.000	.199	8.41	.010	.081	22.76	.139	.027	64.14	.000	.228	1.651	.000	.241

<sup>1</sup> The Netherlands: df<sub>error</sub> = 82; Germany: df<sub>error</sub> = 95; Italy: df<sub>error</sub> = 80

## 2.10 Discussion

The purpose of this study was to explore if the innovative CLIL method produces better EFL results than the mainstream approach and its findings suggest that this is the case, but not in each of the three countries, not for each tested language skill and not to the same degree. The best overall performance for both CLIL and control groups was found in the Netherlands. In both test batteries the Dutch CLIL pupils had the highest scores for receptive Vocabulary as well as for Idioms, and showed the best progress over time for these two skills. The Italian CLIL pupils showed the best progress for Text Comprehension, though they did not set out with the highest scores for this skill: the German CLIL pupils did, but their performance declined and they ended with the lowest text comprehension scores. German performance for receptive Grammar also developed negatively, for CLIL as well as control classes. The language skill in which CLIL classes developed best in comparison with the control classes was Idioms. As to the productive skills as tested in the Error Correction assignment, however, the CLIL groups in the three countries performed better than the control groups at both tests, and all CLIL and control groups showed progress over time.

CLIL as well as control class score developments were diverse and difficult to explain. The 'success of CLIL' does not seem to be proportional to CLIL content lessons: several gain scores were highest in Italy, which is surprising in view of their modest modular CLIL programme. Nor does it depend on teacher qualifications: the dually qualified teachers in Germany did not produce better results than did their Dutch and Italian colleagues with single qualifications.

Attributing the better results of the CLIL groups entirely to the English-medium content lessons is debatable; differences in performance between the CLIL and Control groups were significant for several tested skills before the CLIL intervention had started, a finding that is also mentioned in other research (Huibregtse, 2001; Admiraal, Westhoff & De Bot, 2006). This head start of CLIL groups makes it difficult to account for the final results: perhaps it would have been justified to expect a faster growth rate for the CLIL groups. Moreover, negative outcomes were found for the German schools. In his research conducted at grammar schools in the Berlin area Zydatiß (2012) discusses 'a double language threshold - a lower one and an upper one' (p. 27). A failure to meet the demands of the lower threshold may negatively affect not only pupil achievement as to content matter, but also their motivation to persevere: Zydatiß found evidence that CLIL learners turn their back on CLIL classes for strategic reasons, i.e. better grades when participating in non-CLIL classes.

Recent evaluation of research (Bruton, 2011) expresses doubts as to the effects of the English-medium content lessons per se, without the added extras of the CLIL curriculum and suggests that the effects of selection could very well have

more impact on the success of CLIL than hitherto acknowledged. Research into the domestic background of pupils opting for the CLIL grammar school in the Netherlands (Weenink, 2005) indicates that this type of school basically attracts the privileged, providing a modern alternative for the elitist classical *Gymnasium* with Latin and Greek on the curriculum. In this scenario, reflects Weenink (p. 212), it is feasible that the CLIL grammar school will serve to prepare the most motivated pupils for admission to selective, prestigious English-medium taught programmes in higher education and first class universities, whereas regular, non-CLIL schools will increasingly have to cater for the less motivated, and prepare them for the more average studies. Even though in the three countries in the present study selection criteria are diverse, in order to be admitted to a CLIL classes all pupils had to demonstrate high motivation, good overall performance or EFL proficiency, factors that are likely to contribute to successful perseverance.

In his evaluation of research Bruton (2011) further criticizes the phenomenon of having control groups from the same school as the experimental group; such control groups, he says, are ‘the remnants from the (selected) CLIL groups’ (p. 529). The authors of the present study presume that this does not always transpire. Not every linguistically talented pupil opts for CLIL or other enhanced language learning, but prefers the mainstream learning path, and why not the mainstream class of a CLIL school? A CLIL school may have a very good name and attract more pupils anyhow. To sum up, whether mainstream classes of the same school as the CLIL classes are inadequate as control groups remains speculative; nevertheless the possible effect of CLIL classes absorbing linguistic talent could account for the negative control class results in the present study, notably those in Italy. The Italian schools do not only offer English-medium but also German-medium CLIL, thus attracting a considerable part of L2 interested pupils to either programme. Following the same premises it could be felt least in the Dutch classical *Gymnasium* control class, which was part of a different school and in most respects comparable to its CLIL counterpart.

It should be kept in mind that the present study is small-scale and its findings are only an indication; the number of participant schools is limited and it is difficult to decide to what extent the CLIL approaches are representative for each specific country. If such a concept exists the CLIL schools in the Netherlands probably come closest on account of the effect of national coordination by the Dutch European Platform. Along the same lines the two Italian schools can be seen as representative for CLIL in their own specific region at the time; however, initiatives in other Italian regions were not evaluated and could have led to different results. The German school in Lower Saxony is fairly representative of what is common German CLIL practice; the Berlin school on the other hand is rather unique on account of its catchment area and international

population. However, each of the schools is representative for secondary mainstream pre-university education, with carefully developed CLIL streams.

Finally we would like to observe that the present study serves as a first step in the field of European cross-cultural CLIL research. More and more countries and schools have adopted CLIL since we initiated our study, which makes it possible for future research to find larger populations and more uniformity as to variables such as pupil age and previous EFL training, motivational factors, the implementation of CLIL in the school curriculum and the linguistic orientation of same-school control classes. In order to generalise qualitative research findings as to cross-cultural effects of CLIL on EFL performance such larger studies are necessary, in which differences will be smaller and likely to cancel each other out.

Appendix 1: *CLIL conditions in the Dutch, German and Italian schools*

	The Netherlands <sup>1</sup>		Germany		Italy	
	school 1	school 2	school 1	school 2	school 1	school 2
Pupil admittance is based on:						
Entrance test				x	x	x
Primary school reports	x	x		x		
Sufficient pupil motivation	x	x	x			
CLIL teacher training:						
Dual qualifications			x	x		
Extra CLIL and EFL training	x	x	x	x	x	x
CLIL teacher facilities:						
Peer counseling			x	x	x	x
Partner schools cooperation				x	x	x
CLIL network support					x	x
EFL teacher support					x	x
International study visits	x	x	x		x	x
Compensation of time	x	x			x	x

<sup>1</sup> The Dutch classical *Gymnasium* took part with a control group only and is not included in this table.

Appendix 2: *The implementation of CLIL subjects and EFL in the school curriculum.*

CLIL features	The Netherlands <sup>1</sup>		Germany		Italy	
	school 1	school 2	school 1	school 2	school 1	school 2
CLIL since	2000	1998	1991	1999	2003	1998
CLIL subjects:						
History	x	x	x	x	x	x
Geography	x	x	x	x		
Biology	x	x		x		x
Chemistry						x
Maths	x	x			x	x
Physics					x	x
Computer Science				x		
Latin						x
Italian					x	
Philosophy						x
Religious Education						x
Music	x	x		x		
Arts	x					x
Sports	x		x			
CLIL lessons <sup>2</sup> per week	16	15	5	8	Year1: 2 Year2: 5	Year1: 6 Year2: 9
% of the curriculum	53	52	17	22	10 to 25	20 to 30
EFL lessons <sup>2</sup> per week	5	3	4	5	3	3
Out of school activities						
Language excursions	x	x	x	x	x	x
Theatre visits	x	x		x	x	x
Public speaking contests	x			x		
Youth Parliament			x	x		
Pupil exchange projects	x	x		x		
International study visits	x	x		x		
International ICT projects	x			x		
EFL school projects					x	x

<sup>1</sup>The Dutch classical *Gymnasium* took part with a control group only and is not included in this table.

<sup>2</sup>A lesson is 45 minutes in the Netherlands and Germany, 50 and 55 minutes resp. in schools 1 and 2 in Italy.





# 3

## CHAPTER

### **Effects of Content and Language Integrated Learning in Europe:**

A systematic review of longitudinal experimental studies



## Abstract

The Content and Language Integrated Learning (CLIL) approach, an educational approach in which subject matter and a foreign language - predominantly English - are taught and learnt side by side has developed into a very popular educational innovation in most European countries. A host of research studies have shown the benefits of the approach, and discuss favourable effects especially as to L2 gains. However, critical voices have underscored the fact that CLIL attracts or selects mainly high-achieving learners. Hence, the question arises whether it is justified to attribute improved L2 performance mainly to the CLIL intervention, or to favourable learner characteristics. Several reviews of literature were published in the past, but due to a lack of longitudinal findings no conclusive evidence as to the added value of CLIL in the process of L2 learning could be produced. The present review aims to fill this void and has undertaken a search of two decades of longitudinal studies into the effects of CLIL on various linguistic skills in the field of English as a Foreign Language (EFL). The findings indicate that robust studies were undertaken in only a limited number of European countries, and that only few of them were large-scale. Yet, the conclusions provide clear indications as to the contexts in which CLIL leads to significantly better L2 results.

significantly better L2 results.

*Key words:* Selective CLIL, Longitudinal effects, EFL gains, Literature review.

### Reference:

Goris, J., Denessen, E., & Verhoeven, L.

Effects of Content and Language Integrated Learning in Europe: a Systematic Review of Longitudinal Experimental Studies. *Revised manuscript submitted.*

### 3.1 Introduction

In the past few decades more and more schools in almost all European countries have adopted innovative educational approaches which are meant to prepare young people better for the increasingly internationalised world of the 21<sup>st</sup> century. One innovation in particular – Content and Language Integrated Learning (CLIL) – has met with enthusiasm of teachers, parents and students alike and proved its success. The CLIL approach has become a driving force in various types of mainstream education, mostly secondary but increasingly also primary and vocational schools. Its dual focus aims to develop proficiency in a curriculum subject together with the language through which it is taught – nowadays almost invariably English as a Foreign Language (EFL), which is the target language – a foreign language for the learners under discussion in the present article. Achieving this twofold goal requires an integrated approach to instruction and learning practice. In CLIL education the subject teacher needs to adapt didactics in order to make both content and the L2 comprehensible, so that learning takes place in an interconnected way.

The effectiveness of CLIL soon became the focus of academic studies, primarily with respect to foreign language learning. Research outcomes have been very positive and unquestionably indicate significantly higher L2 levels for CLIL tracks when compared to conventional language classes. It has been convincingly made clear that CLIL learners are better L2 performers, more motivated and more linguistically and academically talented: we can safely say that CLIL selects or attracts the high achievers. However, this gives little or no information about the effectiveness of the intervention itself. The great majority of research studies have investigated all sorts of CLIL benefits by only one measure at a certain point, often with results in favour of CLIL, but a longer perspective is missing. Studies that follow the development of EFL skills of learners in a CLIL class over a certain period of time, and compare the progress made with that of a compatible mainstream, non-CLIL class, so that increased EFL performance can be attributed to the CLIL curriculum are still scarce.

In recent years several reviews of research focussing on the added value of CLIL for various categories of educational practice and motivational outcomes have been published (Dalton-Puffer, 2008, 2011; Pérez-Cañado, 2012; Ruiz de Zarobe, 2011). Even though findings were mainly positive the specific contribution of the CLIL intervention to the growth of EFL skills remained unclear, because only few longitudinal studies could be discussed. The scarcity of solid longitudinal findings has evoked criticism, and the CLIL literature pronounced a need of robust outcomes as to CLIL effects on L2 learning in order to evaluate its merits. A thorough systematic review of experimental longitudinal research, which allows the L2 outcomes to be attributed to CLIL practice is lacking to date. Therefore the present review evaluates the longitudinal studies that

have been published since the onset of CLIL as to the effects on EFL skills of the various CLIL approaches at primary and secondary education levels in varying European settings.

### 3.2 CLIL in European contexts

Even though the teaching of content through an L2 is not a new phenomenon or specific for a certain region, CLIL is very much associated with the European context of mainstream compulsory education as in secondary and more recently also in primary and vocational schools. The launch of the term 'CLIL' in Europe took place in 1994, when the approach was described as a dual focus methodology in which content and language are learnt together, in an integrated way (Marsh et al., 2001). Its roots are in the Canadian immersion approach, in which the curriculum is taught in both the L1 and L2 of the state (Westhoff, 1994).

The preparedness for CLIL was triggered by a general dissatisfaction with the current L2 learning practise around the 1980s and 1990s, which was considered ineffective, especially from a communicative point of view. The ongoing European integration and the ensuing need for L2 competencies (Marsh, 2002; Oonk, 2004) both in daily life and the workplace (Wolff, 2007) contributed to the appreciation of CLIL, and may have raised an awareness among young people that language skills are valuable in an internationalised world. The results of a EU poll in 2006 indicated that 68% of the respondents see English, the world's lingua franca, as the most important foreign language and this is reflected in present-day CLIL programmes. Even though its medium of instruction could in principle be any foreign language, the great majority of European CLIL programmes make use of English as a target language.

It is hard to catch the term CLIL in one single concept. Almost all European countries have introduced CLIL in some form, but as to its origins and implementation European nations have different backgrounds. As will be described in the Results sector of this article, each country has its own educational context and possibilities for implementation, which is displayed in the divergence of CLIL practice. Education is a complex and multi-faceted process, defined by local systems. Theories of second language acquisition (Krashen, 1982; Ellis, 1985) describe L2 development as a complex phenomenon in which progress cannot exclusively be attributed to an educational intervention: acquiring L2 competence is seen as a dynamic process in which a variety of factors interact and whose development is not always linear. For CLIL this is not different. CLIL scholars (Coyle et al., 2010) describe the phenomenon as an 'umbrella term' in which duality, the teaching of subject content through an L2, with the object of teaching and learning both the L2 and the content matter in an integrated way, is a core characteristic. In its broadest sense, the definition of CLIL could be interpreted

as incorporating learner scholastic ability, motivation and aptitude, interacting with teacher didactics, L2 skills and enthusiasm, in short: as more than the sum of its parts.

Apart from the positive effects of the CLIL approach, critical voices have also been heard, mainly from Bruton (2011; 2015) and Paran (2013) who see the CLIL phenomenon as glamourised and put shortcomings in research designs and outcomes under scrutiny. Notably the failure to match experimental and control groups in terms of language level and aptitude has evoked this criticism. In the specific case of CLIL its potential learners are often high performers, in academic as well as linguistic fields. At the start of secondary education some form of differentiation or streaming in accordance with individual learner characteristics such as verbal intelligence and academic ability is common practice in many European countries (Eurydice, 2012). However, for those who wish to study in a CLIL class additional criteria apply. Secondary schools generally make use of selection criteria for admittance to a CLIL class, involving enhanced motivation and above-average EFL skills, while in some countries prospective learners for CLIL streams receive extra EFL training in preparation (Authors, 2013). This may lead to significant discrepancies as to initial EFL level between experimental and same-cohort mainstream groups, a phenomenon that seems to be treated as an intrinsic design problem in CLIL research. It has been acknowledged that securing homogeneity can be a difficult task, e.g. by Lasagabaster (2008) who discussed the fact that some variables are hard to control in a study carried out in an authentic educational setting, often by individuals and with limited means. In the Results sector of this article we will discuss how the authors of the studies canvassed here have dealt with the issue of matching experimental and control groups.

### 3.3 Previous reviews of CLIL effects on linguistic gains

In past years several inventories of outcome-based classroom studies were presented. No unanimity existed as to the linguistic competencies that benefit from CLIL. Dalton-Puffer (2007; 2008; 2009) focussed on L2 gains in divergent European contexts, mostly German-speaking countries. Positive effects were reported for receptive language skills, vocabulary and morphology as well as creativity, risk-taking, fluency and speaking confidence. Positive but different findings e.g. with regard to writing skills were reported in Spain by Ruiz de Zarobe (2011), who contradicted previous studies in some respects. The present review aims to contrast and compare research conditions and CLIL contexts, and present more information as to how the linguistic gains came about in the divergent educational context of the countries in Europe that worked with CLIL over the past decades.

A similar goal was pursued in a meta-analysis conducted in 2012 by Pérez-Cañado, who presented a comprehensive, updated and critical review of the way in

which CLIL, the new educational approach was playing itself out on our continent. The survey pertained to countries in northern, central, eastern and southern Europe and discussed a broad field of cognitive, educational and affective variables affected by CLIL education, such as motivation, teaching practice, content learning and also linguistic merits, mainly concerning fields of EFL competence. The results unquestionably indicated that CLIL affects L2 language learning outcomes. However, longitudinal studies were scarce and moreover, no detailed discussion as to educational details of those included was presented. The conclusions (p. 329) stated that the last two decades had mainly seen a growing number of studies of a descriptive nature, focussing on the benefits of bilingual education, whereas solid empirical studies have been sparse, and underscored the lack of robust findings resulting from a pre-test / post-test design.

Six years have passed since, years in which CLIL has increasingly become a feature of 21<sup>st</sup> century education and placed itself in the spotlights of research into CLIL effects but also of criticism. It is generally acknowledged that CLIL is promotive to L2 skills; however, its selectiveness, the way in which experimental and control groups are compared and the absence of pre-tests have been criticized as factors biasing research outcomes. The present review aims to evaluate the longitudinal studies into CLIL effects on EFL progress that have been published since the early days of CLIL, present a detailed overview of the various linguistic skills that were tested and what the findings were, and compare variation as to learners and educational contexts. We also seek to make clear how research outcomes have developed over time, and discuss how topical CLIL issues such as initial differences and experimental-control group matching have been dealt with by the authors.

### **3. 4 The present review**

As we have discussed above the contexts in which European CLIL have taken shape are very much divergent, yet their goals converge: to provide better L2 learning conditions, either for a select group or for all. In the light of the complexity of introducing a new and far-stretching educational programme such as CLIL, its existence and development are hardly past infancy. The field of CLIL is very much in motion and therefore, the authors of the present study think it of value to the research field to investigate to what extent CLIL has fulfilled its promise of better L2 education. Our aim is to come to an evaluation of the contribution of CLIL to competencies in the most used target language, English, of mainstream primary and secondary school learners during the past 20 years of CLIL existence. To be more precise, the authors aim to provide a systematic review of longitudinal experimental research analysing the development of EFL proficiency skills of a certain cohort of learners over time and comparing results of CLIL groups with those of control groups in traditional classes.

## Method

### 3.5 Data collection

A systematic search for articles analysing the effects of CLIL education on EFL skills was conducted by the first author together with a university librarian in February 2018. Our main search principle was that L2 proficiency is an integration of knowledge of words, expressions, insight into the rules of the language and an ability to understand its written texts, which we defined as the research focus of the studies we were looking for. Therefore we combined the keyword “CLIL” and its earlier term ‘bilingual education’ with the search words “vocabulary, grammar, idioms and text comprehension” to find articles in Europe in the databases ERIC and PsycInfo. In total, 235 articles were found in the ERIC database and 172 in that of PsycInfo. This number was considered sufficient as a starting point for further selection. The screening of articles took place in several phases. To start with, the first author classified the titles and abstracts of the articles as relevant or irrelevant. This resulted in 142 studies being relevant for further inspection. In the second screening round a number of inclusion and exclusion criteria were applied. Studies were included if they answered to the conditions of CLIL as discussed in our introduction, contained a measure of one or more EFL skills, were conducted in the course of the past 20 years, written in English, had participants in mainstream primary or secondary education in a European country, made use of a control group and were longitudinal, i.e. had more than one measure in time of the same cohort, with statistical analyses of significant results. They were excluded if the English language used as the medium of instruction in bilingual forms of education was not a foreign language, as e.g. in English-Welsh programmes in Wales. Full-text articles were retrieved and further inspection by the first author using the same inclusion criteria resulted in 19 articles and two dissertations.

### 3.6 Presentation of the results

To systematically evaluate the retrieved texts, they were first subdivided into two categories: studies conducted at primary or at secondary schools. An overview was made of key characteristics for each study: title, year of publication, country, name of the author or authors, the number of times data were collected, CLIL content subjects, sample size for CLIL and control groups and starting age of the participants – their age at the first research measurement. Longitudinal CLIL effects – i.e. findings of a greater increase in EFL scores over time of experimental groups when compared to control groups, provided they were significant – were categorized as pertaining to: overall proficiency, which also may include oral skills, vocabulary, grammar, reading and listening skills. Table 1 presents the results of primary education, Table 2 those of secondary education. After drawing up overviews of the above data in the two tables,

we concentrated on describing the educational and research context for each of the studies. Our aim was to provide information as to the development and implementation of CLIL in the respective countries, selectiveness and classroom practice in the diverse educational settings, the tested linguistic skills and research outcomes, and the practice as to matching CLIL and control classes. If possible - i.e. if mentioned in the study under discussion - we also described effect sizes of significant results.

Table 1: Results for Primary Education

Country	Authors	Year	Primary Education				Longitudinal CLIL effects per measured scale									
			Publication	Data collecting times	CLIL subjects	N CLILN	Control	Age at T1	Overall proficiency	Vocabulary	Grammar	Reading fluency	Listening			
Netherlands	Van der Leij et al.	2010	Acquiring reading and vocabulary in Dutch and English	2 in 1 year	Reading	23	23	8	8	–	–	significant	–	–	–	–
Spain	Agustín-Llach and Alonso	2014	Vocabulary growth in young CLL and traditional EFL learners	3 in 3 years	Natural science	58	49	9/10	9/10	–	–	none	–	–	–	–
Spain	Agustín-Llach	2015	Age and type of Instruction in lexical development	idem	idem	68	61	9/10	9/10	–	–	none	–	–	–	–
Spain	Pladevall-Ballester and Vallbona	2016	CLIL in minimal input contexts: A longitudinal study of primary school learners' receptive skills	4 in 2 years	Science Arts & crafts	138	149	9/10	9/10	–	–	–	–	–	none	negative



## Results

### 3.7 Primary education outcomes

As outlined in Table 1, four primary school studies were evaluated. Starting in the Netherlands the only study (van der Leij et al., 2010) conducted in primary education is an early, small-scale study at a time when primary CLIL was by no means common ground. Tested scales were L2 vocabulary, word reading fluency and orthographic knowledge, the information stored in memory telling us how to represent spoken language in written form.

The two groups were comparable on Dutch origin, social-economic background, age and sex, and L1 skills that have shown to affect reading acquisition. Participants were assessed twice within one year halfway through in grades 2 and 3, when they were eight years old. The number of EFL lessons was the same as that of the control class, viz. 5x 20 minutes per week in grades 1 and 2. The experimental group received substantial bilingual (Dutch and English-medium) instruction in English reading. The results showed favourable effects for reading ability in both languages, as well as significant differences in vocabulary acquisition between the experimental and control group: the authors report an impressive difference in progress between the groups as to L2 vocabulary (effect size  $\eta^2$   $p = .12$ ). Even though this study is not about the L2-medium teaching of content which is a distinct feature of CLIL, it provides support for its didactics of being exposed to L2-medium instruction.

The majority of studies included in the present review, both in primary and secondary education, were conducted in Spain. Spain is made up of 17 autonomous communities and CLIL implementation may vary greatly from one community to another. CLIL is less of an elitist approach and introduced at broad educational levels, in which a divergence of CLIL policies exist. For example, CLIL is compulsory in all primary schools in Navarre, whereas only a few experimental programmes have been implemented in La Rioja. Moreover, CLIL contents and policy may differ greatly from one school to the next as schools can decide for themselves. Often there are no admittance criteria and admission to CLIL is voluntary. This still implies (self)selectiveness: the average ability and motivation to learn English is presumably higher, a learner characteristic seen in the great majority of present-day non-obligatory CLIL environments.

An example of non-selective CLIL at a broad level is the bilingual project that was launched in ten Spanish regions in 1996 as the result of an agreement between the Ministry of Education and Science and the British Council. In order to provide EFL education compatible with EU standards an enriched language learning model was implemented in state schools to provide better EFL education from infant school to secondary levels, for learners from all socio-economic backgrounds, who often had no access to foreign language learning. In one of Spain's northern regions a study

at primary level was conducted in 2014 (Agustín-Llach & Alonso) testing receptive vocabulary growth over time. All participants were from the same school, but some years apart. The variable age was controlled for. All 4<sup>th</sup>, 5<sup>th</sup> and 6<sup>th</sup> graders participated in the study, first when the school had only English as a school subject and later when CLIL tuition was introduced. At the start of the study participants were nine or ten years old and attended the 4<sup>th</sup> grade; at the end of the study they were between 11 and 12 and in the 6<sup>th</sup> grade. The experimental CLIL group had been receiving extra exposure to English in natural science lessons for two hours a week since grade 1, the control group had only received traditional EFL lessons. Data were collected three times by means of ten-minutes vocabulary tests during class time in three consecutive school years. CLIL learners showed higher figures for vocabulary sizes, and differences increased with grade until they became significant in the 5<sup>th</sup> and even more so in the 6<sup>th</sup> grade. As to longitudinal effects, CLIL learners showed slightly higher growth rates than non-CLIL learners, but no significant statistical effects were found: very similar patterns of lexical development in CLIL and traditional learners were observed. The authors mentioned age as a determining factor in receptive vocabulary knowledge: increasing exposure does not lead to lexical gains in young learners, but more benefit may be expected as they grow older.

A related study by the same first author (2015) gave similar results as to different aspects of vocabulary development of a similar, slightly larger group of CLIL and traditional EFL learners, measured along three years. The study controlled for the variable age. Both groups completed a letter-writing task, which was then scrutinized for L1 influence in the form of borrowings – which the author mentioned as typical for low level learners and generally an indication of overall lack of lexical knowledge in the L2 – and lexical creations. The results showed that vocabulary related to school and classroom activities, and management was frequent in both groups of learners. Both produced roughly similar numbers of words related to the field of science. CLIL learners produced fewer instances of borrowings than traditional learners and these tended to decrease with increasing proficiency: the number of borrowings produced by CLIL learners in 6<sup>th</sup> grade was significantly lower than that of traditional EFL learners, faintly suggesting CLIL benefits. As in the previous study, the author considered (young) age as a factor imposing a strong constraint in L2 lexical development, even more so than exposure time in itself.

The last primary school study (Pladevall-Ballester & Vallbona, 2016) was conducted in the Spanish bilingual region of Catalonia in which Catalan, the language of instruction, together with Spanish, are the majority languages. English is taught as the main foreign language in mainstream education. CLIL was initially the approach adopted to revitalize Catalan after the Spanish dictatorship. All subjects except the

Spanish language have since been taught through Catalan. Receptive skills (reading and listening) in English were measured by means of Cambridge language tests at four different times during two academic years and involved pupils aged nine or ten in the 5<sup>th</sup> and 6<sup>th</sup> grades. In order to guarantee comparability between the two groups, the total amount of exposure up to each testing time was kept the same. Four state-funded private schools participated with two groups, one exposed only to EFL sessions and the other one exposed to EFL sessions and an additional CLIL hour per week. CLIL was new: two of the schools had implemented CLIL in the science subject, and the other two used it in the arts and crafts subject, both as a first time experience. The results showed no significant differences between the groups with regard to reading skills, while the control group significantly outperformed the CLIL group as to listening skills. As an important factor underlying this negative CLIL outcome the authors observed that the results of the comparison between the arts and crafts group and its control group at the third measure were significantly higher in favour of the control group, which had almost certainly an effect on the overall results of the sample. Discrepancies in instructional practice between the two CLIL groups as to their content subjects were observed. The listening competence of the young learners in arts and crafts could have been affected by instructional practices that relied heavily on visual inputs and gestural support in comparison to the science group. As a result, the arts and crafts learners did not necessarily have to confront the demands of the language, so did not have to develop the strategies necessary to improve their listening abilities as much as the science learners did. In addition, the language used in the arts and crafts lessons was mainly based on instructions, and did not require as much cognitive effort on the part of the learners as the language used in the science lessons, which was much more complex and demanding, thus favouring the learners' listening comprehension development. The authors concluded by mentioning that the CLIL implementation was new to the teachers, which probably affected their instructional skills.

### **3.8 Secondary education outcomes**

Table 2 presents an overview of studies conducted in secondary education. In the northern part of Europe we found only one study: a Swedish dissertation (Sylvén, 2010). In spite of the popularity of the English language – according to EU figures about 95% of the Scandinavian population have EFL proficiency to some degree – longitudinal CLIL studies are thin on the ground in the Nordic area. The author mentioned the fact that in grade 10, when CLIL starts, the year groups are split up into several different groups for elective reasons and sometimes optional courses, which hampers longitudinal research studies. Selection of students based on proficiency tests is not allowed, but in practice CLIL is self-selective in that it attracts students who are already

fairly proficient in English. The study analysed the testing of four different types of vocabulary in three rounds spread over two years and showed favourable results for the CLIL groups when compared to the same-school mainstream control groups, matched as to age. A possible bias lay in student motivation to participate in the project, about which the author remarked: “whereas the CLIL students were happy enough to show their skills, the control group found it hard to find the motivation. Their general feeling was best described as: ‘This is only meant to show how badly we perform in comparison with the CLIL group’.” (p.49).

Vocabulary growth was a frequent focus of CLIL research as the increased exposure to the L2 is seen as crucial for the acquisition of new words. In Austria Gierlinger and Wagner (2016) investigated CLIL-based vocabulary growth and conducted a study in non-selective, rather low-achieving mixed-ability classes in lower secondary education. A popular Austrian CLIL form in lower secondary and primary education nowadays is the teaching of content in a number of modules, interspersed with mother tongue teaching. The participating schools offered chemistry, history and geography in around five to seven modules of up to four weeks each, representing 60-80 hours of CLIL class time, spread over the school year.

Table 2: Results for Secondary Education

Country	Authors	Year	Secondary Education			Longitudinal CLIL effects per measured scale									
			Publication	Data collecting	CLIL subjects	N CLIL	N Control	Age at T1	Overall proficiency	Vocabulary	Grammar	Reading Compreh.	Listening		
Sweden	Sylvén	2010	Teaching in English or English Teaching?	3 in 2 years		99	264	15/16	—	—	—	—	—	—	—
Austria	Gierlinger and Wagner	2016	Revisiting CLIL-Based Vocabulary Growth in Sec. Education	2 in 1 year	geography history chemistry	39	48	14	—	—	—	—	—	—	—
The Netherlands et al.	Admiraal et al.	2006	Evaluation of bilingual secondary education in the Netherlands	16 in 6 years	history geography	584	721	12	—	—	—	—	—	—	—
The Netherlands et al.	Verspoor et al.	2015	The effects of English bilingual education in the Netherlands	3 in 1 year	50% of the curriculum	Grade1: 83 Grade3: 74	Reg. 84 Contr. 49	12	—	—	—	—	—	—	—
Netherlands Germany	Goris et al.	2013	Effects of the CLIL approach to EFL teaching: A comparative study	2 in 2 years	50% of cur history geography	Neth: 37 Germ: 50	Neth: 47 Germ: 47	12 13	—	—	—	—	—	—	—
Italy Germany	Dallinger et al.	2016	The effect of CLIL on students' English and history competences	2 in 1 year	history modules	Italy: 45 483	Italy: 37 NonCLIL1: (SS) 354 NonCLIL2: (DS) 444	14 13	—	—	—	—	—	—	—

Table 2: Continued

Country	Authors	Year	Secondary Education		Data collecting subjects	N CLIL	N Control T1	Age at T1	Longitudinal CLIL effects per measured scale				
			Publication	Data collecting subjects					Overall proficiency or idioms	Vocabulary	Grammar	Reading Compreh.	Listening
Germany	Rumlich	2017	CLIL theory and empirical reality	2 in 2 years	1 increasing to 3 subjects	503	NonCLIL1: (SS) 473 NonCLIL2: (DS) 182	12/13	none	–	–	–	–
Spain	Pérez-Vidal and Roquet	2015	CLIL in Context: Profiling Language Abilities	2 in 1 year	science	50	50	12/14	–	sign.	sign.	sign.	none
Spain	Roquet and Pérez-Vidal	2015	Do Productive Skills Improve in CLIL Contexts?	2 in 1 year	science	50	50	12/14	–	none	none	–	–
Spain	Gené-Gil et al.	2015	Development of EFL Writing over 3 Years in Sec. Ed.	4 in 3 years	science or social science	30	20	13	–	none	sign.	–	–
Spain	Gené-Gil et al.	2016	A methodology for longitudinal research on EFL written production	4 in 3 years	science	30	30	13	–	none	none	–	–
Spain	Alonso et al.	2008	Plurilingual education in secondary schools: Analysis of results	2 in 2 years	20-25% of the curriculum	Gr.1: 67 Gr.3: 44 Bac: 48	Gr.1: 20 Gr.3: 20 Bac: 30	12 14 16	undecided	–	–	–	–
Spain	Ruiz de Zarobe	2008	CLIL and Foreign Language Learning	3 in 3 years	social science Engl.lit.	CLIL1:24 CLIL2:36	29	14/15	none	none	none	–	–

Table 2: *Continued*

Country	Authors	Year	Publication	Data collecting	CLIL subjects	N CLIL	N Control	Age at T1	Longitudinal CLIL effects per measured scale				
									Overall proficiency or idioms	Vocabulary	Grammar	Reading Compreh.	Listening
Spain	Merino and Lasagabaster	2017	The effect of CLIL programmes' intensity on English proficiency	2 in 1 year	CLIL- : 3.4 CLIL+: 8.4	208 CLIL+108	77	11/12	–	–	–	–	–
Spain	San Isidro and Lasagabaster	2018	The impact of CLIL on pluriliteracy development	3 in 2 years	social science	20	24	14/15	–	–	–	–	–
Spain	Pérez Cañado and Lancaster	2017	The effects of CLIL on oral comprehension and production	3 in 1 year and 6 months	2 subjects	12	12	15/16	–	–	–	–	–
Spain	Pérez Cañado	2018	CLIL and Educational Level: A Longitudinal Study on the Impact of CLIL on Language Outcomes	2 in 1 year and 6 months	arts, maths, science, PE technology soc/natural sciences	1,033	991	Prim: 11/12; Sec: 15/16	–	–	–	–	–

The study was conducted before and after the CLIL intervention within the school year 2010-2011. Participants were from four different schools and formed two CLIL and three control classes, matched as to age, in which the CLIL learners were not preselected but formed part of a whole-class of mixed ability policy who participated in their school-wide CLIL enrichment project. The authors found no significant effect on EFL vocabulary scores for the CLIL treatment, nor for CLIL exposure over time. Moreover, their investigation of CLIL teacher classroom language indicated that teachers mainly used the band of 1,000 high-frequency words in the content lessons, which may have restricted more advanced and broader word learning.

In Dutch secondary education three studies (Admiraal et al., 2006; Verspoor et al., 2015; Goris et al., 2013) were conducted at pre-university level with pupils aged 12 at the start, being the start of secondary school and also CLIL. The predominant type of Dutch CLIL is about 50% of the curriculum being taught in English, with the main focus on the content rather than the language. Admittance to a CLIL class is offered mainly at mainstream pre-university secondary education and subject to selection criteria: good overall academic performance, motivation to persevere and an above average interest to learn the English language. The 2006 study (Admiraal et al.) was started in 1993 when the CLIL approach was new in mainstream education in the Netherlands. The authors used the abbreviation BE (Bilingual Education, a term still used to denote CLIL). The study tested receptive vocabulary, reading comprehension and oral proficiency; however, only receptive vocabulary was measured longitudinally with 16 tests spread over six years. The other two scales had one measure only, and both showed higher scores for the BE (CLIL) groups. Vocabulary mean scores at the start of the programme were significantly higher for the CLIL group than for the control group, also after student characteristics – gender, general ability, home language, language contact, motivation to learn English, introduced as covariates – were taken into account. As to score development on the EFL vocabulary test, there was no significant difference between the experimental and control groups; the CLIL learners did not acquire English words faster than the control group. The population and CLIL practice of this early study is not representative for the later more developed CLIL in the Netherlands. Students from the BE (CLIL) programme differed significantly on student characteristics such as nationality and native language, which may be explained from the fact that three out of the five experimental schools were schools for International Education, accommodating more proficient EFL speakers. The other two were mainstream schools populated by almost 100% of Dutch native speaker students, offering a mix of a bilingual programme and regular secondary education. The number of CLIL lessons remained unclear. The control groups were partly from these two, and partly randomly sampled from two other, non-CLIL schools.



The second study (Verspoor et al., 2015) conducted in the Netherlands is of a more recent date and fully represents the Dutch CLIL approach as accredited by the European Platform for Education in the Netherlands<sup>1</sup> which coordinates and monitors all CLIL schools. The authors tested receptive vocabulary and full linguistic repertoire by means of a productive informal writing task. There were three measurements spread over one year – in October, January and June – at four secondary schools, with participants in grade 1 and grade 3. The authors were of opinion – and agreed with Bruton – that comparing CLIL classes only to non-CLIL classes would give a distinct bias as in their experience non-CLIL classes have on average a lower scholastic aptitude, less motivation and lower initial EFL proficiency. They included students of the – prestigious and selective – Dutch *Gymnasium* as a better match to act as a control class for CLIL and therefore incorporated three streams in the study: CLIL, regular (same school non-CLIL) and *Gymnasium* (as control). The results indicated that at the first test in grade 1 the CLIL as well as the control classes had significantly higher scores than the regulars, with no such differences between CLIL and control. However, CLIL learners were significantly more motivated than regular and control peers and gained more: at the second and third tests the CLIL scores were significantly higher than those of both the regular and control groups, while between the latter two there were no significant differences. In the third grade – a different cohort – CLIL did not continue to gain more. The authors concluded that, when all covariates were taken into account, CLIL only outperformed the regulars, not the control classes, so that CLIL maintained the lead rather than gaining more than the other groups. Initial proficiency was found to be a significant contribution to final proficiency. The authors wondered whether similar proficiency results could be accomplished by better or more EFL teaching in regular programmes.

The reason for the initial better performance of the CLIL groups remained unclear. Prospective learners for CLIL streams in the Netherlands receive no special preparation. In principle they have the same EFL primary school lessons, even though practice between schools differs: some start earlier and do more than the obligatory lessons. Both studies tested mainly vocabulary. It is possible that learners with an inclination towards the English language pick up more words from its omnipresence in daily life. Nor was it clear why vocabulary gain seemed to be initial and did not develop in a linear fashion. A possible ceiling effect was mentioned, or the fact that vocabulary tests may have limitations in that they make use of frequency bands, testing the most common words whereas CLIL vocabulary gain is likely to be found in subject-specific or academic use of language.

---

1 In 2015 Nuffic merged with the European Platform and took over its coordinating role.

The third study (Goris et al., 2013) was conducted between 2007 and 2009 and tested more language skills apart from vocabulary. The results in three countries were compared: the Netherlands, Germany and Italy. The experimental groups were representative for national CLIL practice in their region at the time. Tested skills included receptive vocabulary, idioms, grammar, idioms and text comprehension. In each country four classes from pre-university secondary schools took part in the study: two CLIL groups and two control groups. There were two measurements: the first one at the beginning of secondary school, also the starting point of CLIL and the second one two years later, at the end of the second grade. With one exception in the Netherlands – where one control group from a classical *Gymnasium* took part – the control groups were from the same schools as the experimental groups. Pupil age, scholastic aptitude, SES and demographic background, national majority native language, EFL contact and preparation for CLIL were the same for each experimental-control match per country, even though between countries differences existed as to age and CLIL preparation. The results indicated that the initial test scores were on the whole highest in the Netherlands and lowest in Italy. Furthermore, the CLIL groups in all three countries had higher – though not always significantly – initial scores than the control groups on all scales, apart from vocabulary in Italy where the control groups knew slightly more words. The second test round showed higher scores on all scales for the CLIL groups. As to gain scores there were differences between countries. In the Netherlands CLIL gain scores were significant for grammar and idioms, in Germany no significant CLIL gains were found, while in Italy such scores were found on all tested scales, which was surprising in view of the modest modular CLIL approach. A possible explanation for the successful outcomes in Italy might lie in a Hawthorne effect, the positive influence of taking part in an event. Both the CLIL approach and the EFL presence on the curriculum were very new here and the international university research met with great enthusiasm of teachers and pupils alike. The opposite effect could have occurred in Germany, where the strict anonymity required for participation in research possibly contributed to a disinterested attitude – or the fact that the English language was not abundantly popular in society at large.

In Germany two robust studies (2016, 2017) were retrieved. The 2016 study (Dallinger et al.) investigated skill development of 1806 German CLIL and non-CLIL eighth-graders in English and history, controlling for a wide range of student, classroom and teacher characteristics such as student EFL interest. The study made use of two kinds of non-CLIL control groups: cohorts from the same school as the experimental CLIL groups (non-CLIL 1) and cohorts from different schools without CLIL programmes (non-CLIL 2). There were two measurements: one at the start and one at the end of grade 8. In Germany, parents may decide for or against a CLIL school when registering their child for

secondary education – mostly a *Gymnasium*, an academic type, preparing for university. This type of school accommodates mainly learners from high-SES homes, in which one or both parents followed academic education. Admittance is decided together with the teacher and school admittance criteria. The German educational system in most states provides extra EFL education in preparation for CLIL, which results in substantial differences in initial EFL proficiency between CLIL and non-CLIL: the former are likely to possess higher prior knowledge in English.

In the study under discussion CLIL classes accommodated more students with immigration backgrounds and higher socio-economic status than non-CLIL. The CLIL classes' initial general English skills – as measured by a C-test<sup>2</sup> – and listening comprehension were significantly better than those of the non-CLIL classes. This was also the case at the second test. As to the longitudinal pre-test/post-test effects the results of multilevel modelling confirmed that CLIL classrooms showed significantly greater increases in English listening comprehension but not general English skills than non-CLIL classrooms. Controlling for systematic differences by means of multilevel regression analyses greatly diminished the CLIL effect on English skills for both domains, and rendered that on general English skills insignificant. The authors mentioned prior achievement as the strongest predictor of future learning.

Similar findings were discussed by Rumlich (2017). The article summarizes the author's dissertation on CLIL streams at German secondary schools of the *Gymnasium* type in North-Rhine Westphalia, the most populous German state with the largest number of CLIL schools. These schools have the most intense form of CLIL implementation in mainstream German education, and offer continuous forms of CLIL with a duration of at least one school year. At the start one content subject is involved, in grades 8 and 9 two or at most three subjects. In preparation of this high-intensity version students receive additional EFL lessons in grades 5 and 6 before the start of CLIL in grade 7 at ages 12-13. Over 1,000 learners were involved in a two-year pre-test / post-test research study into the effects of CLIL on EFL proficiency as measured by C-tests. As in the earlier mentioned German study three groups participated: CLIL classes and two kinds of non-CLIL control groups: cohorts from the same school as the

---

2 by Lucy Katona and Zoltan Dornyei  
from: Forum English Teaching

The C-test is an integrative testing instrument that measures overall language competence, very much like the cloze test. It consists of four to six short, preferably authentic, texts in the target language, to which "the rule of two" has been applied: the second half of every second word has been deleted, beginning with the second word of the second sentence; the first and last sentences are left intact. If a word has an odd number of letters, the "bigger" part is omitted, e.g., proud becomes pr-. One-letter words, such as I, are ignored in the counting. The students' task is to restore the missing parts. In a typical C-test there are 100 gaps-that is, missing parts. Only entirely correct restorations are accepted.

experimental CLIL groups (non-CLIL 1) who are seen as ‘a negatively selected group of students with below-average EFL-proficiency’ (p.115) – thus inappropriate as controls and cohorts from different schools without CLIL programmes (non-CLIL 2), whose students are unselected and unprepared, thus seen as neutral and more appropriate controls. The author observed that prospective CLIL and non-CLIL general proficiency in English diverged greatly in favour of the former even before the implementation of CLIL, partly due to the preparatory lessons and partly to selection effects. The findings as to EFL development indicated that all groups advanced significantly over time, with little difference between the CLIL and non-CLIL 2 controls and a slightly weaker progress for the non-CLIL 1 group. It became evident that there was no detectable influence of CLIL on general EFL proficiency. The author concluded with a strong claim for longitudinal evaluations that screen for selection, EFL preparation and class composition effects.

At secondary level Spanish studies also had the lion’s share. In Barcelona (Pérez-Vidal & Roquet, 2015) investigated the effects of a newly introduced CLIL programme which was carefully designed by a team of university experts, in order to identify which areas of L2 competence benefitted the most from CLIL instruction. Two different groups of Catalan / Spanish bilingual learners were analysed longitudinally over one academic year. The productive skill of writing and receptive skills of reading and listening were investigated, as well as lexical-grammatical ability, the skill to use vocabulary and grammar correctly. Written development was analysed quantitatively as to syntactic and lexical complexity, accuracy and fluency, and qualitatively as to task fulfilment, organisation, grammar and vocabulary. The experimental group, who had previous experience with CLIL since they were ten in grade 5. received formal EFL lessons as a school subject, in addition to an English-medium science subject taught with a CLIL approach. The control group received formal instruction of English as a subject only. Data collection started at the end of their first year of secondary education, in grade 7. The CLIL-group was not matched for age with the control group as this would have given them the advantage of more EFL exposure, but on the ground of a similar total number of hours of EFL education. This meant that the control group included learners who were a year older and one grade higher than the CLIL group. The results of two measurements within one school year for receptive skills showed that the CLIL group improved their reading competence significantly more than the control group, but not their listening competence. CLIL students’ lexical-grammatical ability also increased significantly more. As for the productive skill of writing, there was a significant improvement for the CLIL group in their abilities to write more accurate and syntactically complex texts, and a general improvement in the whole set of qualitative measures (task fulfilment, organisation, grammar and vocabulary). The positive results for writing skills appear to be in contrast with Dalton-Puffer’s findings (2008), who classified writing as

one of the areas of linguistic competence likely to remain unaffected by CLIL instruction as content teaching is conducted almost completely without writing activities. However, the Spanish authors claimed a transfer of writing skills to the CLIL context from the mainstream EFL lessons, in which writing is often practised.

Even though the results for writing skill development turned out to be positive for the CLIL group, only a part was significant. The same authors – with a reversal of first and second authorship (Roquet & Pérez-Vidal, 2015) – produced a more detailed discussion of the participants' written production outcomes. The study singled out the composition assignment – writing a dialogue based on a picture – that was part of the original testing battery and used the same criteria to evaluate the development of aspects of written abilities. The results of the qualitative and quantitative measures indicated that as to syntactic and lexical complexity as well as fluency in writing the CLIL group did not progress significantly more than the control group; this was only the case in the domain of accuracy. As to qualitative results in the field of task fulfilment, organization, grammar and vocabulary there were no significant differences between the progress of the two groups. Overall, the authors could not confirm the superiority of CLIL as to writing skills.

A further study into fine-grained aspects of writing competence was conducted in a different part of Catalonia, on the Balearic Islands, by Gené-Gil et al. (2015). They evaluated complexity, accuracy, and fluency (CAF) in a CLIL programme called 'European Sections', introduced in 2004. Schools joined the programme on a voluntary basis and participation of students was optional. Participants were two groups of Spanish / Catalan bilingual secondary students, an experimental CLIL group learning science or social science through the medium of English and a comparable non-CLIL control group with English as a subject only. All participants answered a profile questionnaire enabling the researchers to rule out any important differences in language background and extra-school exposure to the target language. CLIL and non-CLIL students' average age was 13 at the first test at the start of CLIL in 2008. Both CLIL and non-CLIL groups were asked to write a timed composition (25 minutes) in English at every data collection time, four times spread over three school years. The results indicated no significant differences at the first test: CLIL and non-CLIL participants' onset level was equivalent. Over the course of three school years, CLIL students attained significantly improved performance in every domain analysed (written syntactic and lexical complexity, accuracy and fluency), whereas in the case of their non-CLIL counterparts improvement was restricted to the lexical complexity and accuracy domains.

In a later article the authors (2016) critically examined their methodology. Their major concern was to describe EFL writing in all its complexity and multidimensionality, a difficult task as fine-grained effects of writing skills are difficult to assess, let alone

evaluate and compare statistically over time and between groups. They held the pre-test / post-test design with repeated measures to be the only way to control for differences in task complexity, despite the risk of carryover effects. In a study related to the previous one they used micro-analytical measures of diversity in use of vocabulary or lexical complexity and accuracy. Participants were two groups of comparable adolescent learners of EFL: an experimental CLIL group and a non-CLIL comparison group. The former received three hours a week of EFL instruction and had started studying science in English for three hours a week in grade 8 on a voluntary basis. The latter only received EFL instruction. Data about written production were gathered through two communicative writing tasks, designed for this study: a general, interpersonal task and subject-specific, yet sufficiently general task related to science, the CLIL subject in the participating schools. The results indicated that CLIL learners showed improved writing competence, particularly in micro-analytical measures of diversity in vocabulary use and accuracy, and that they incorporated some specialized lexis into their subject-specific compositions. The authors proposed a methodological framework with a greater focus on evaluation procedures of such details, “in order to capture all its richness” as most writing activities simply support grammar and vocabulary learning, and to encourage more investigations of EFL writing development in CLIL contexts in which progress often lies in nuances. They confirmed that more exposure to the target language in itself does not necessarily lead to enhanced written competence.

Several Spanish studies were published in the bilingual Basque country. The educational system in the Spanish Basque Autonomous Community (BAC) offers linguistic models with either Spanish or Basque as the language of instruction and the other one a curriculum subject, or with Basque and Spanish-medium teaching side by side. At the start of the millennium FL medium teaching, using foreign languages – mostly English – in addition to the two national languages, was introduced. The Department of Education initiated a ‘Plurilingual Experience’ (PE) in twelve schools in 2003, six of them participating in a case study analysing its effects. The PE programme required that at the compulsory level at least seven hours a week be taught in a foreign language – other than Basque or Spanish. In the post-compulsory or baccalaureate years, when students are 16-18 years old, the number of L2 medium subjects should be at least 20-25 % of all lessons. The experimental schools offered the opportunity to study the curriculum subjects in three different languages: Basque, Spanish and English (L3 for the students). Participants had to answer selection criteria of academic performance and motivation. There were two measures at three levels (grades 1, 3, and baccalaureate): the first one in October 2004, at the start of PE, the second in May 2006 – after two years. The control groups were in the same grade as those of the experimental groups but did not take part in the PE. The evaluation looked at English

proficiency – listening and reading comprehension, written and oral production and grammatical knowledge using validated Cambridge tests as well as qualitative data concerning opinions from students and teachers.

The report on the results of the PE project “Trilingual students in secondary education” was published by the Basque Institute of Educational Evaluation and Research in 2007, and the research team, Alonso et al. (2008) summarized the study. The findings showed that the English-medium experimental group achieved considerably better results than the control group, advantages that increased even more in the course of time. Comparisons and longitudinal growth were calculated and described in percentages – not very common in statistical analyses. To estimate the accumulated gains, which were quite substantial, the grading for each test was distributed on a common scale so that the differences between each phase could be summed. For this, the approximate and provisional results obtained in earlier research on the empirical validation of the CEFR (Common European Frame of Reference<sup>3</sup>) scales were used. The research team stated that this analysis is a theoretical simulation of scores that refer to the hypothetical differential gain that would arise if the data were presented on a common scale, so that results have to be taken with caution. For the present review it means that the question whether the effects were statistically significant remains unanswered. We felt that this landmark study had to be included in our discussion, the more so as it initiated prolific CLIL practice and research within the Basque Country.

Ruiz de Zarobe (2008) conducted a study to investigate if including more English-medium content subjects enhances CLIL effects on L2 proficiency (the more-content-is-better hypothesis) and tested EFL speech production of learners in two different CLIL programmes. Adequacy as to pronunciation, vocabulary, grammar, fluency and production of content were compared with those of a control group enrolled in a traditional EFL programme only. The first CLIL group, CLIL 1, entered CLIL at 14. One content subject, social sciences, was taught through English for three or four hours a week. The second CLIL group, CLIL 2, entered a CLIL programme with two English-taught content subjects: social sciences (three to four hours a week) and modern English literature (two hours a week). The control group only received the conventional three hours of EFL per week, the same as the CLIL groups. Learners were in grade 3, aged 14-15 at the first test and in the last year, the pre-university year or baccalaureate at the last test. By then, the number of participants supplying data had decreased: the control group consisted of seven students, the CLIL 2 group of fourteen and data of the CLIL 1 group could no longer be collected at the time. The results showed that

---

3 The Common European Framework of Reference for Languages (CEF or CEFR) was put together by the Council of Europe as a way of standardising the levels of language exams in different regions. It is very widely used internationally and all important exams are mapped to the CEFR.

at the first measurement in grade 3 the CLIL groups significantly outperformed the control group in all scales, without important differences between the CLIL types. In the 4<sup>th</sup> grade CLIL 2 once again scored highest, with significant differences in all scales except vocabulary. The participants in the pre-university grade belonged to two groups: CLIL 2 and control. The differences between the two were significant for vocabulary and grammar. Although the CLIL groups' scores were higher than those of the control group throughout the grades, there seemed to be no significant increase in proficiency throughout the years. The question if more content is better could be answered affirmatively: a positive relationship was present between the amount of EFL exposure and the linguistic outcomes.

Merino and Lasagabaster (2017) addressed a similar issue: the role played by intensity in CLIL programmes on overall proficiency in English: speaking, reading, listening and writing. Participants were 393 secondary education students from three different autonomous communities: the Basque Autonomous Community, an officially bilingual community in Basque and Spanish; La Rioja, a monolingual Spanish community, and Cantabria, also a monolingual community. They were spread over two experimental groups and one control group. The control group consisted of 77 learners from eight schools with Basque as the means of instruction for all subjects except Spanish. The first experimental group (CLIL-) was made up of 208 CLIL learners from the same eight schools, who had 3.4 CLIL sessions per week. The second experimental group (CLIL+) comprised 108 CLIL learners from five high schools in Cantabria and La Rioja with 8.4 CLIL sessions per week. In two test rounds in the school year 2010 - 2011, the first one in the final term of grade 7 and the next one at the end of grade 8, Cambridge ESOL test (Key English Test) were administered. The findings showed that a higher amount of CLIL produced a greater improvement in the L2. At the first test, both CLIL groups showed significantly higher scores than the non-CLIL group. No differences were found between the scores attained by CLIL- and CLIL+. However, at the second test the contrast between CLIL- and CLIL+ had increased and showed that the evolution of CLIL+ students was significantly higher than that of their CLIL- counterparts. The latter made progress in an almost identical degree to the control group. As in the previously discussed study (Ruiz de Zarobe, 2008), the question if more content is better could be answered affirmatively.

A small-scale study into general EFL as measured by proficiency was conducted by San Isidro and Lasagabaster (2018) in a rural multilingual school in Galicia, a Spanish north-western autonomous community with specific linguistic and cultural hallmarks. Two official languages are spoken: Galician and Spanish. The study investigated the impact of CLIL on



multiple language learning (L1, L2 and EFL) and content learning. The first measurement of the two-year longitudinal study was at the start of the third grade of secondary education in September 2012, when learners were 14-15 years old. The final test was at the end of the fourth grade. EFL scales had three measures using Cambridge tests and included the four skills: reading, writing, speaking and listening. Participants were two comparable groups without any prior selection, and homogeneous before the implementation of CLIL. The results indicated that both groups improved their competence in English after two years, while the CLIL cohort made significantly greater progress. Another interesting finding was that the CLIL students also outperformed their non-CLIL counterparts in both Spanish and Galician over the two school years, whereas content learning was not negatively affected.

A small-scale study on listening and oral production skills (Pérez-Cañado & Lancaster, 2017) was published on the outcomes of a 4th grade secondary school sample. The pre-tests took place in September 2012, the post-tests were carried out in June 2013, the second post-tests were completed in January 2014. The CLIL and control groups had initially been matched on the pre-test, when no statistically significant differences were found on listening skills. Over the period of one academic year and six months following the conclusion of the intervention programme, the post-tests were applied. The listening tests were group-administered in one sitting under the same conditions each time. In turn, in the oral production tests, the students were examined in pairs, with individual subtasks lasting up to five minutes. On the post-test statistically significant differences emerged in favour of the CLIL group, for listening skills as well as for oral production skills: all of the tasks in the oral production test (spoken interaction in an interview and in individual speaking) evinced statistically significant differences in favour of the CLIL group, demonstrating that these students are able to communicate more effectively. At the final stage of the investigation, both groups had levelled out on oral comprehension competence: no statistically significant differences were detected for any part of the test or for the test as a whole. As to oral production skills, however, at the second post-tests the scores of the CLIL students significantly surpassed those of their non-CLIL peers, with statistically significant differences for the overall test and each of its tasks and skills.

The last study to be discussed here is a large-scale investigation into EFL proficiency (Pérez-Cañado, 2018) involving 1,033 CLIL students and 991 EFL learners in 53 public, private, and charter schools<sup>4</sup> in three of the monolingual communities in Spain which have the least tradition in bilingual education: Andalusia, Extremadura and the Canary Islands. Tested skills were grammar, vocabulary, reading, listening and

---

4 Charter schools are state-financed schools, most of which have a religious orientation.

speaking. At the start of the academic year 2014 – 2015, the experimental and control groups of 11-12 year-olds were matched on a pre-test in terms of socio-economic status, English level, verbal intelligence and motivation. Thus, homogeneity between both cohorts was initially guaranteed. At the end of the same academic year, in June 2015, when 828 learners (aged 11-12) were finishing the last grade (grade 6) of primary school and 1,196 (aged 15 -16) were about to complete the last grade (grade 4) of compulsory secondary education – so all of them were on the switch to the next level - the English language post-tests were administered. Six months later, in December 2015, the delayed post-test was applied to the same participants who were previously in 4<sup>th</sup> grade of secondary education and had continued in the 1st grade of the baccalaureate, where CLIL instruction stopped. The results of the English language post-test at the end of primary school showed statistically significant differences on all the linguistic components and skills sampled, invariably in favour of the CLIL group, and particularly marked for the productive speaking skill. As to the post-test at the end of the 4<sup>th</sup> grade significant differences had increased and were also found on both oral production and comprehension skills.

Medium effect sizes were discerned for vocabulary (Cohen's  $d = -0.619$ ) and the five aspects of speaking ( $-0.858$ ). After four additional years of participation in CLIL education, at the post-test at the end of the 4<sup>th</sup> grade, the differences in EFL competence had increased. Statistically significant differences were found in favour of the CLIL cohorts on all the linguistic aspects sampled, at high confidence levels and with large effect sizes. The latter were particularly considerable for use of English (Cohen's  $d = -1.160$ ) and speaking ( $-1.230$ ), especially as to lexical range ( $-1.442$ ) and task fulfilment ( $-1.482$ ).

Time turned out to be a crucial factor to ascertain the effects of CLIL on foreign language attainment; the longer the students had been benefitting from bilingual education, the greater the differences with their non-bilingual counterparts. For the group in the 4<sup>th</sup> grade of secondary school, who had had four additional years of participation in CLIL programmes, the differences in EFL competence were further reinforced when compared to the younger cohort, and statistically significant differences emerged in favour of the CLIL groups on all the linguistic aspects. The effects pervaded and became even stronger six months later, when the former 4<sup>th</sup> graders were in the first year of non-compulsory secondary education. Statistically significant differences continued to be discerned in favour of bilingual streams on all the linguistic components and skills sampled, at high confidence levels, and with even larger effect sizes, especially for speaking (Cohen's  $d = -2.671$ ), and except for reading, which had the comparatively lowest effect size ( $-0.868$ ).

At this level, it turned out that productive skills had been more positively affected by CLIL as opposed to receptive. Interesting data were found in the follow-up of the 4<sup>th</sup> graders on aspects of oral competence that need more time to develop in order to be significantly improved, viz. pronunciation and fluency. Also at this point, type of school yielded interesting results: the non-bilingual charter schools were catching up with the public and private bilingual ones, as there were no statistically significant differences between them and both bilingual types of schools on the use of English.

### 3.9 Conclusion

The aim of the present review was to evaluate the findings of longitudinal experimental research into the effects of the CLIL approach on EFL proficiency, conducted during the past twenty years. Our findings as comprised in tables 1 and 2 do not provide unequivocal support for the hypothesis that learners in a CLIL class will develop more EFL proficiency over a certain stretch of time than their mainstream counterparts: the majority of studies produced null effects. Furthermore, it is striking that studies with a longitudinal perspective have been undertaken in only a limited number of European countries, and mainly during the last five years. The number of participants greatly varies. Nevertheless, even though an overall picture of longitudinal CLIL effects in Europe is lacking, all findings – both large-scale and small-scale - are valuable and helpful in analysing this massive educational innovation, seen by many as best practice for the future.

As to the analysed skills the picture shows a broad diversity: a variety of EFL competences was investigated. The most frequently tested skill appeared to be vocabulary. The three Dutch studies as well as the Swedish and Austrian ones all tested receptive vocabulary, for which no significant growth over time was found. The Goris et al. study (2013) also tested other skills, and found significant results in the Netherlands for idioms and grammar, an indication that CLIL affects certain linguistic skills more than others. The same study did not find significant effects for any linguistic skill in Germany. Several studies tested overall proficiency, for which no significant effects were found in the Netherlands or Germany where spoken skills were not part of the tests.

Productive writing was tested in several Spanish areas. The authors discussed various aspects of writing skills – for the sake of convenience we listed them under ‘grammar’ and ‘vocabulary’ in Table 2 –, some of which clearly indicated superiority of CLIL learners. Spoken fluency was tested in Spain where five studies addressed this skill as part of overall proficiency in receptive and productive skills. Three of them found significant results in favour of CLIL for spoken fluency, the skill commonly believed to be the most favourably affected because of the increased opportunity for authentic communication.

A topical issue in many European educational systems is whether an early start of L2 learning is better than postponing it until a later date, or whether it is wise to include CLIL courses before any teaching of the target language has taken place. In most European countries CLIL at elementary level was hardly practiced until recently, and as yet no comprehensive literature discusses early mainstream CLIL results. Apart from one primary school study in the Netherlands, which presented positive effects on vocabulary, mainly as a result of teacher instruction rather than a content subject, three studies were found in two different regions in Spain. No convincing evidence in favour of an early introduction of CLIL emerged, while one study even indicated negative results for listening skills. The authors assumed that (young) age is a constraining factor in L2 development. Studies undertaken with somewhat older primary school learners provided more positive outcomes, as in Pérez-Canado (2018) where a large group of 11-12 year-old CLIL learners already obtained higher EFL gain scores than their mainstream peers after one year.

Another recurrent issue in the CLIL literature is the question whether a higher amount of CLIL leads to more positive L2 effects. Two Spanish studies in the present review addressed this question and both studies provided affirmative answers. It seems best to consider these results as context-specific, for across countries they do not always materialise. The results in the Netherlands and Sweden, where a large percentage of the curriculum is offered in the target language, were not strikingly better than those e.g. in Germany, which practises a moderate approach by including two or three subjects. Several authors provided indications that CLIL results grow in the course of time; that learners profit more after a longer period of learning in a CLIL track.

It is hard to interpret our findings without discussing the practice of matching experimental and control groups and the initial CLIL head start, issues under criticism by Bruton (2011). The most reliable practice from a statistical point of view would be to match pupils – ideally a random sample, as in Admiraal et al. (2006) who included a mix of same-school and random-sampled participants from non-CLIL schools in the control group – on crucial variables such as age, SES, intellectual and linguistic capacity, as well as initial EFL levels. Notably this last feature presented problems and was not always observed in the studies under review, the main reason being that in many countries pupils receive extra EFL preparation if they aspire to CLIL, as in Germany and Italy. In Sweden CLIL is self-selective and introduced for those interested in upper secondary education, which implies that learners are already fairly proficient. This inevitably leads to an EFL head start when compared to mainstream counterparts. At the same time, the studies in the Netherlands showed that such head starts are also possible without special preparation. Therefore, the results of these studies must be interpreted with this limitation in mind.

The CLIL initial head start was even found to be significant in several studies, which makes it difficult to explain the progress these pupils make in the following years: does it result from CLIL or the obviously better L2 skills? The use of a control group with lesser skills is thought to present a bias in the research findings. Suitable options to match experimental and control groups were found in the Netherlands as there are two types of pre-university education, the *VWO* - the mainstream type of preparing those with sufficient academic talent for university studies - and the classical *Gymnasium*, of old a more prestigious school type for those that excel, comparable to CLIL candidates. The study by Verspoor et al. (2015) involved groups from both types, and their study made clear that the difference in EFL progress between CLIL and non-CLIL *Gymnasium* learners was not significant, whereas it was between CLIL and the 'regular' *VWO* learners from the mainstream departments of the participating CLIL schools. The former appeared to be an appropriate match for the CLIL learners, in contrast with the so-called same-school control groups which the authors described as having on average a lower scholastic aptitude, less motivation and lower initial EFL proficiency.

In Germany, where schools preparing for university do not know a similar divide as in the Netherlands, the authors (Dallinger et al., 2016; Rumlich, 2017) expressed similar views and Rumlich described same-school control groups as 'a negatively selected group of students with below-average EFL-proficiency'. Cohorts from different schools without CLIL programmes, where students are unselected and unprepared were included as more appropriate controls in both German studies. They, too, provided evidence for the same-school experimental-control group mismatch, as did student discouragement to participate in a school research project mentioned by Sylvén (2010). In this respect the Austrian study presented a more appropriate research context: the CLIL students were average learners and participated in their school-wide CLIL enrichment project. They were not preselected but formed part of a whole-class of mixed ability policy. Suitable research contexts were also present in Spain, where experimental-mainstream matches were found without initial differences in EFL skills, even though the problem was acknowledged here, too, in the case of a one-year older control group which is not ideal, either.

What our review has not made clear is the specific impact of various CLIL target language content subjects, if such effects exist apart from the classroom language used by all CLIL teachers. It is hard to imagine that a language-rich subject such as e.g. history influences the target language to the same degree as a subject depending on complex cognitive explanations such as e.g. mathematics. Few authors elaborated on this issue even though Pladevall-Ballester (2016) explained that more complex language favoured the learners' listening comprehension development. For the rest, our study produced

no indications that the target language of some content subjects affects EFL skills more or differently than others.

Another issue not dealt with in the present review is the impact of content teacher skills, both from a didactic and linguistic point of view. Great varieties exist as to levels of L2 skills. As an example we mention that in Germany content teachers are qualified to teach both the L2 and the subject. In the Netherlands they are only qualified to teach their subject, while a B2 level for the target language is required for CLIL teaching. In addition, classroom practice diverges across European countries. Sometimes native speaker teachers and language assistants are employed, while the role of the EFL teachers in the CLIL process also diverges. We feel that the impact of teacher skills is beyond the scope of this review, as it is a complex and many-faceted phenomenon that needs further study.

### 3.10 Discussion

Our findings produced interesting contradictions between Spain on the one hand and the rest of Europe on the other. Whereas the latter were found to produce mainly null effects, in Spain significant effects were more frequent. In order to contribute to possible explanations we must look at the history of CLIL. In most countries CLIL came about bottom-up, inspired by parental or educational demands. It was introduced almost without exception as a selective option at pre-university level, intended to prepare high SES students for international careers and studies. In Spain, on the other hand, CLIL had its origins in the opposite direction: it came about top-down as a joint initiative of educational authorities. Its ideology was to provide better EFL learning opportunities for all, as from the early years. Spain, just as other southern-European countries e.g. Italy – included in the Goris et al. (2013) study, who found several significant effects that were hard to explain from the analyses of results – used to be and still is a society with low EFL proficiency. According to the Special Eurobarometer on Europeans and their languages, issued by the European Commission in 2006, Spain and Italy represented the countries where the smallest percentage of citizens could hold a conversation in language apart from their national L1. This may account for the positive effects of a massive L2 learning innovation: there was much more room for improvement than e.g. in Sweden or the Netherlands, which were among the top 8 of countries in which nine out of ten inhabitants could speak at least two languages according to the EU survey and where English was spoken quite well by the great majority of people.

The focus of the present review was to investigate if CLIL has met its promise of providing a better EFL learning approach. The answer is by no means negative, but the degree to which it is positive varies. High EFL-proficiency countries with elitist and

highly selective CLIL such as the Netherlands and Germany have gained only little on the tested scales. In Spain, on the other hand, a low EFL-proficiency country, the CLIL approach was planted on fertile soil. In the first place because of the country's dire need of improved EFL teaching in an increasingly internationalised market place, but also because there was ample experience with the teaching of content through two languages in Spanish bilingual regions. In the Basque Country, for instance, educational practice entails teaching content through Spanish and Basque, two completely unrelated national languages, which is claimed to be conducive to further L3 acquisition, in the case of CLIL the English language.

The final thought of the present review may be summarised by saying that the research published in the past two decades as to the benefits of CLIL has presented many valuable and robust findings, predominantly during the last five years and conceivably in response to criticism and reviews. Our most conclusive finding is that CLIL is profiled best in the divergent contexts of Spain, which sets a positive precedence for other low-EFL countries in the EU that are still in less advanced stages.







# 4

## CHAPTER

**The contribution of CLIL to learners'  
international orientation and EFL  
confidence**

## Abstract<sup>1</sup>

As the target language in Content and Language Integrated Learning is invariably English, many see CLIL as a way of helping learners develop an optimal command of English as a foreign language. The focus of many prior research studies has been on gains in language proficiency but the aims of CLIL reach well beyond this. The present study concentrates on whether, and to what extent, CLIL also contributes to building pupils' confidence as EFL users, who are well-prepared for life in an increasingly internationalised world. Specifically, it looks at the impact of two years' CLIL study on two constructs: 'EFL confidence' and 'international orientation'. The study was undertaken with eleven groups of 12 to 15-year-olds at mainstream 'grammar' schools (i.e. for pupils intending to enter university) in the Netherlands, Germany and Italy. The study involved 231 pupils: 123 pupils following CLIL streams and 108 mainstream pupils. The results indicate that even at the start of CLIL classes pupils already showed greater confidence in their EFL skills and had a stronger international orientation than their mainstream counterparts. However, all pupils, both CLIL and mainstream, showed a positive development on our two variables during their first two years at grammar school. The CLIL intervention thus seemed to produce only a small added value.

*Keywords:* CLIL practice in the Netherlands, Germany and Italy, EFL learner confidence, international orientation, English-medium CLIL.

### Reference:

Goris, J., Denessen, E., & Verhoeven, L. (2017).

The contribution of CLIL to learners' international orientation and EFL confidence.

*The Language Learning Journal*, doi: 10.1080/09571736.2016.1275034

---

1 Data for this chapter were collected between 2007 and 2009.

#### 4.1 Introduction

Content and Language Integrated Learning has been introduced in many European countries since the 1990s onwards. This innovative approach aims to improve language learning opportunities through the use of a target second language in the teaching of a range of subjects in the school curriculum. It has been strongly promoted in Europe where considerable value has been placed on knowledge of foreign languages but conventional teaching methods have been considered inadequate for meeting learners' future communicative needs in a changing, globalised society where English has in effect become the lingua franca (Eurydice Report, 2006). Even though in principle any foreign language may be used as a medium of instruction in CLIL programmes, English is the most widely implemented target language for CLIL in Europe (Dalton-Puffer, 2011).

Warschauer (2000), among others, has described how the industrial societies of the past are giving way to a new economic order based on global manufacturing, and the use of new technologies. These developments have brought about extensive international trade negotiations and cooperation, and have led the European Union to promote education for multilingual and multicultural citizens in a globalised context, where travelling, studying at a foreign university or building a career in a foreign country are within reach of all. Thus, attention to foreign language education has been promoted in European school curricula. Furthermore, globalisation may also be contributing to an awareness among young people that language skills are valuable in an increasingly internationalised marketplace and integrated Europe and the rest of the world. The extensive language input in CLIL classes provides additional opportunities for learners to process and use a foreign language. CLIL pupils are expected to not only learn more language for the purposes of social communication, but also to develop a broader range of academic language proficiency that could potentially lay a foundation for TL use in further study or employment (Graz group, 2013).

The implementation of CLIL has been the focus of research ever since it was first introduced and results suggest broadly positive outcomes. Many studies have found higher EFL proficiency levels for pupils enrolled in CLIL classes in addition to mainstream language classes (e.g. Huibregtse, 2001). This was the case even when pupils were at the very early stages of CLIL. Admiraal, Westhoff and de Bot (2006) found significantly higher scores for EFL reading comprehension, general oral proficiency and pronunciation after two years of CLIL. The authors of the present study (2013) found similarly positive results for vocabulary, grammar, idioms and text comprehension in three European countries: the Netherlands, Germany and Italy, each of which have rather different CLIL approaches. Dalton-Puffer (2008) also compared research findings from CLIL in different European countries: she found positive effects for receptive language skills, vocabulary and morphology as well as creativity, risk-taking, fluency

and speaking confidence. In particular, spontaneous oral production was the area where the difference between CLIL and mainstream learners is most noticeable (Dalton-Puffer 2011). Studies focussing on affective outcomes such as motivation, positive attitudes towards language learning, satisfaction and increased confidence have also showed positive results in favour of CLIL learners. CLIL learners tend to be more motivated to learn the foreign language than their non-CLIL counterparts and have more positive attitudes towards language learning (Doiz et al., 2014; Pérez-Cañado, 2012). They also develop better communication skills and experience satisfaction when they succeed in mastering the content subjects in the foreign language, a feeling that enhances their motivation and linguistic confidence (Dale & Tanner, 2012).

Research has also found that pupils following CLIL programmes had higher EFL proficiency scores before starting CLIL (Rumlich 2013). Verspoor, de Bot and Xu (2015) found initial proficiency to be a strong predictor for later EFL proficiency results, an effect still present after three years of study in a CLIL class, interacting with motivation and attitudes. Otwinowska and Foryś (2015) see the initial presence of favourable attitudes towards CLIL learning as a prerequisite to positive outcomes. There are questions therefore as to whether CLIL 'works' because CLIL classes tend to be 'selective', and thus the pupils following them tend to be more motivated and more proficient in the first place, or whether the CLIL approach itself can be considered the key factor in bringing about the positive outcomes highlighted by research.

The aim of the present study is to investigate how pupils in the Netherlands, Germany and Italy starting CLIL in 'grammar' schools - i.e. secondary schools preparing learners for university - differ from their peers in mainstream classes in these schools in terms of their international orientation and their perceived confidence for using their EFL skills, and how these variables develop in the two groups of pupils after two years.

## 4.2 The aims of CLIL

CLIL is considered by some to be based on egalitarian principles and appropriate for a broad range of learners (Marsh, 2002; Wolff, 2002). In several European countries (e.g. Spain, Latvia and Estonia and lately in Italy), CLIL programmes have been implemented with a wide range of pupils. However, in practice, CLIL still is predominantly a selective programme, for which schools tend to apply strict admittance criteria: pupil selection typically depends on above-average academic performance, an adequate level of EFL and motivation to persevere (Bruton, 2011). These criteria were applied by the schools participating in the present study. Several studies point out that CLIL pupils are often from internationally orientated homes (Mehisto, 2007; Weenink, 2005) and have a strong desire to learn English, seeing the CLIL programme as a challenge. This suggests they have what Gardner (1985) describes as integrative motivation, implying that they take pleasure in language learning, and have favourable attitudes towards language learning, L2 native speakers and international cultures. Integrative motivation has repeatedly been found to be conducive to successful second language learning (e.g. Dörnyei et al., 2006).

CLIL schools have typically also fulfilled a pioneering role in internationalisation in secondary education from several social perspectives. The Eurydice report (2006) noted 'preparing pupils for life in a more internationalised society and offering them better job prospects on the labour market' among the aims of CLIL, as well as the socio-cultural aim of 'conveying to pupils values of tolerance and respect vis-a-vis other cultures'.

## 4.3 CLIL practices in the Netherlands, Germany and Italy

In the Netherlands, of 642 secondary schools, there are at present 130 schools with English-medium CLIL programmes, most of them 'grammar schools' for academically gifted pupils. Similar figures were not available for Italy or Germany. In terms of the subjects taught through CLIL in these three countries, these are determined by national or regional guidelines. There is also considerable variation in the proportion of lessons taught using CLIL in different countries. The authors of the present study (2013) found that in the Netherlands, CLIL can comprise up to 50 to 60% of the curriculum, while in Germany, the percentage is about 20%. Italy has a modular approach, which means that CLIL subjects are not offered throughout the school year but in a set number of modules, increasing in number per year.

In addition to curricular subjects, CLIL often covers out-of-school activities such as English-oriented language excursions and theatre visits, international pupil exchanges and ICT projects. In this respect, schools in the three participating countries decide their own policy. CLIL school coordinators from the Netherlands and Germany

taking part in this study mentioned pupils participating in public speaking contests or the European Youth Parliament as extra opportunities for EFL practice. This was not mentioned by the Italian schools.

Recent developments in CLIL practice relate to the introduction of English-taught programmes into a broader range of school types. A development we wish to mention here is the project launched by the Italian Ministry of Education to make CLIL mandatory in some form in the last year of secondary high schools (*Licei* and *Istituti Tecnici*) in order to provide not only the selected CLIL groups but all pupils with extra EFL practice in preparation of life in the 21<sup>st</sup> century (Langé, 2014). This development took place after the data for the present study were collected and does not affect our research findings.

#### 4.4 Motivation in L2 learning

Considerable research on motivation for second language learning has been conducted by Robert Gardner and associates (Gardner & Lambert 1972; Gardner 1985). This research initially distinguished between two orientations: integrative and instrumental. Integrative motivation comprises the desire to belong to, and engage with, the target language culture as well as having favourable attitudes towards the learning situation, the teacher and the course. Instrumental orientation encompasses the wish to develop favourable career prospects, and is utilitarian in nature. Socio-linguistic research has confirmed that positive attitudes towards the target language community are linked to high levels of language confidence (Noels & Clément, 1996; Hummel, 2013). MacIntyre et al. (1998; 2001) stress the fact that the major motivation to learn a foreign language is the development of communicative relationships with target language speakers. MacIntyre et al. define the learner's 'willingness to communicate' as 'readiness to enter into discourse at a particular time with a specific person or persons, using a L2' (1998: p.547). This is a non-cognitive variable, which was found to contribute to L2 achievement, along with linguistic confidence - the trust in one's own ability to use the foreign language adequately. This feeling of confidence and the absence of anxiety are mentioned as a positive influence on a person's willingness to use the foreign language (Dörnyei & Skehan, 2003).

In later versions of his work on L2 motivation, Gardner (2010) looked at motivation in classroom learning and noted that it was affected by 'the teacher, the class atmosphere, the course content, materials and facilities, as well as personal characteristics of the student'

(2010: p.3). This aspect was integrated in Dörnyei and Hadfield's 'L2 Motivational Self' (2013) and relates to the impact of success in the language classroom on the development of confidence in one's L2 skills.

#### 4.5 The present study

As we have discussed, present-day educational policy in Europe and CLIL in particular, typically aims to promote in young people a positive attitude towards other cultures and an interest in global affairs, together with confident advanced language skills which would enable them to take up studies at a foreign university if they so choose, or start a career abroad. To what extent CLIL adds value to the achievement of these aims is the key question motivating this study. The specific research questions for the study are as follows:

- Are pupils who have chosen to follow CLIL in grammar schools in the Netherlands, Germany and Italy more internationally orientated and more confident in their EFL skills than their mainstream peers at the outset of the CLIL programme?
- Does CLIL contribute more to pupils' international orientation and EFL confidence than mainstream education in the course of the first two years at grammar school in these three countries?



## Method

### 4.6 Participants

The participants were basically the same as in our 2013 study (Authors, 2013), in which we studied their cognitive EFL proficiency results. There were 231 pupils comprising 123 CLIL and 108 mainstream pupils, aged between 12 and 15. We restricted ourselves to secondary schools preparing pupils for university studies so as to have participants with broadly similar intellectual abilities and career outlook. We refer to these schools as ‘grammar’ schools or by their local terms, *Gymnasium* in the Netherlands and in Germany, and *Liceo* in Italy. Three Dutch schools took part along with two German schools, and two Italian schools. Table 1 shows the distribution of participants.

Table 1: *Distribution of pupils (N=231)*

Country	N	Initial age	CLIL		Control	
			boys	girls	boys	girls
<i>The Netherlands</i>	84					
School 1	22	12.4			11	11
School 2	20	12.3	9	11		
School 3	42	12.5	10	7	14	11
<i>Germany</i>	86					
School 1	45	12.8	14	10	12	9
School 2	41	12.7	7	11	12	11
<i>Italy</i>	61					
School 1	20	14.2	16	4		
School 2	41	14.4	14	10	5	12

All participants had learnt some English, starting at primary school. In Dutch primary schools at least one hour per week is obligatory for the last two years when pupils are 11 or 12. In Germany, the 16 individual states decide the details of the school curricula. In most German states primary school pupils have two compulsory lessons per week in Years 3 and 4, at ages eight and nine. However, both in the Netherlands and in Germany many schools teach additional hours and also in lower grades. In Berlin, where one of the study schools was located, children have the opportunity to go to CLIL primary schools, where they practise the target language from Year 1 (aged six). The Berlin pupils start secondary education at age 12, the other German pupils at age ten. Thus, the latter spend Years 5 and 6 at secondary school, with two or three hours of EFL teaching on the curriculum before the CLIL lessons start in Year 7.

The situation in the Netherlands and Germany was largely as described above when the data were collected for the present study. In Italy there have been changes

since then. The country undertook a reform in 2010 to make English a compulsory subject with a specific weekly timetable throughout all primary education until the end of secondary school. At the start of the present study Italian primary schools had some form of foreign language education and many of them had started teaching English, but in an informal way. There were considerable differences between schools. Formal EFL lessons, generally two or three per week, were on the curriculum at the start of the *Scuola Media*, which lasts three years and starts at age 11. In short, before the pupils came to their present class at grammar school they had had very diverse EFL learning paths but they can all be seen as beginners.

For the present study, we selected as our 'experimental' groups pupils at the start of secondary CLIL EFL programmes, with their counterparts in mainstream classes in the same schools – with the exception of one mainstream class in a different school in the Netherlands - comprising the control groups. Given the diverse educational contexts of the Netherlands, Germany and Italy, there were inevitably some differences in age and in initial levels of EFL between the different national groups. In the Netherlands and Germany, participants were aged 12 while in Italy, in the first year of the *Liceo*, they were aged 14. The majority of participants (96%) were born in the country they were studying in, and had the national language as their mother tongue.

#### 4.7 Instruments

The participants completed a questionnaire measuring their international orientation and their confidence as EFL learners. These constructs were based on Gardner's (1985) concept of 'integrative motivation' and MacIntyre's (1998) concept of 'willingness to communicate'. International orientation was considered to encompass a broad, integrative orientation, i.e. a general interest in other languages and cultures, a sense of interest in, and identification with, target language speakers and cultural products, as well as a desire to use the target language for international communication, study, and work. 21 items were drawn up, reflecting aspects of integrative motivation, and a further three items referred to more utilitarian motives such as the wish to have a better job, reflecting Gardner's concept of 'instrumental orientation'. The 24 items had high reliability (i.e. they correlated strongly at  $\alpha=.88$  for the pre-test and  $\alpha=.85$  for the post-test) and thus appeared to refer to one single underlying construct.

'Confidence as EFL learners' was based on those aspects of Gardner's concept of integrative motivation that pertain to the learning situation, with the addition of the concepts of linguistic confidence and willingness to use the foreign language (Dörnyei & Skehan, 2003). It was measured by five items (with a reliability of  $\alpha=.75$  for the pre-test and  $\alpha=.81$  for the post-test). All items were evaluated on a six-point Likert scale. The full list is given in the Appendix. Pearson's correlations between the constructs of

international orientation and EFL confidence were  $r=.199$ ,  $p=.00$  for the pre-test and  $r=.151$ ,  $p=.02$  for the post-test. Thus, the two constructs were found to be significantly coherent. Pupils with high scores for international orientation will also have high scores for EFL confidence.

#### **4.8 Procedure**

At the start of the school year in which CLIL classes commenced, participant pupils completed the questionnaire a first time (pre-test). They repeated the questionnaire (post-test) at the end of the following school year, i.e. more or less two years later. The test administrator, a university researcher, explained to the pupils that they were taking part in a university research project and their answers would be anonymous and for research purposes only. Both the pre-questionnaire and the post-questionnaire were completed in a single lesson period.

## Results

### 4.9 Initial scores

An ANOVA of the pre-questionnaire was conducted to show initial variation on international orientation and language confidence. This was followed by ANOVAs of repeated measures of the pre-questionnaire and post-questionnaire.

Table 2: Scores for pupil scales, itemized for country.

	CLIL group				Control group			
	Begin year 1		End year 2		Begin year 1		End year 2	
<i>The Netherlands</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
International orientation	3.86	.54	4.07	.50	3.55	.57	3.75	.61
EFL confidence	4.49	.76	4.67	.78	3.84	.86	4.03	1.01
<i>Germany</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
International orientation	4.39	.57	4.22	.57	3.77	.67	4.06	.65
EFL confidence	4.26	1.05	4.41	.99	4.15	1.07	4.01	1.02
<i>Italy</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
International orientation	3.68	.65	3.92	.59	3.34	.55	3.79	.57
EFL confidence	3.80	1.04	3.79	1.12	3.20	.84	3.55	.80

The descriptive statistics shown in Table 2 indicate that for all three countries, the CLIL groups had higher initial scores than the control group on both international orientation and EFL confidence: CLIL pupils started out with more interest in the international cultures, foreign languages and the use of English for future careers, and greater confidence in using the target language. This is confirmed by the ANOVA of the pre-test, shown in Table 3, which indicated a significant initial advantage for the CLIL groups on international orientation,  $F(1,225) = 25.16, p < .01$ , and EFL confidence,  $F(1,225) = 11.61, p = .001$ . There were also significant differences between countries both for international orientation,  $F(1,225) = 16.05, p < .01$  and for EFL confidence,  $F(1,225) = 9.71, p < .01$ . The German CLIL participants had significantly higher initial scores on international orientation, while the Dutch CLIL pupils had significantly higher scores for EFL confidence. The Italian scores on both variables and for both CLIL and control groups were significantly lower than those of the corresponding Dutch and German groups.

### 4.10 Score development

Table 3 shows the development over time of EFL confidence and international orientation. Over the two years, all pupils became significantly more interested in the international culture, foreign languages and the use of English:  $F(1,225) = 23.52, p < .01$ . Significant interaction effects for group were found,  $F(1,225) = 6.59, p = .011$  and for

country  $F(1,225) = 3.5, p = .032$ . As can be seen in Table 2, there are differences in score development between the groups and countries over time: the German CLIL responses on international orientation actually decreased (from an average of 4.39 to 4.22 on the scale of 1-6), in contrast to participants from the other two countries, who gave more positive responses on post-test. Interestingly, however, the Italian control group scores for international orientation increased more than those of the CLIL group. Scores for EFL confidence increased for both Dutch groups, but confidence decreased in the German control group and very slightly in the Italian CLIL group. Although overall, the scores on EFL confidence tended to increase over the two years, this was not found to be a significant improvement:  $F(1,225) = .072, p = .788$ .

Table 3: ANOVA Repeated Measures Beginning year one and End year two (pre-test / post-test).

	International Orientation				EFL Confidence				
	Df*	SS	p	$\eta^2$	F	SS	p	$\eta^2$	F
WS:Time	1	4.31	.000	.095	23.52	1.47	.070	.015	3.32
Time x country	2	1.28	.032	.030	3.50	.82	.398	.008	.924
Time x group	1	1.21	.011	.028	6.59	.03	.788	.000	.072
Time x group x country	2	1.11	.050	.026	3.03	1.67	.154	.016	1.88
BS: Country	2	13.64	.000	.104	13.11	31.84	.000	.089	11.02
Group	1	9.94	.000	.078	19.10	19.68	.000	.057	13.62
Group x country	2	.38	.694	.003	.365	3.17	.336	.010	1.10

\*Df error = 225

A more detailed analysis of score development per country is shown in Table 4. This suggests that both Dutch and Italian pupils developed significantly in international orientation during the course of their two years at grammar school:  $F(1,82) = 12.36, p = .001$  and  $F(1,59) = 12.71, p = .001$  respectively. However, in neither case were there significant differences between the CLIL and control groups.

The German results showed a somewhat different picture: taking scores for international orientation from both CLIL and control participants, there was no evidence of significant growth over time for either group. There was, however, a significant difference between the CLIL and the control group:  $F(1,84) = 11.64, p = .001$ . The German control groups developed significantly more interest in international culture, foreign languages and the English language for their future purposes, than the CLIL group.

Table 4: ANOVA Repeated Measures Beginning year one and End year two (pre-test / post-test) per country.

	International orientation					EFL confidence				
	<i>Df</i> *	<i>SS</i>	<i>p</i>	$\eta^2$	<i>F</i>	<i>SS</i>	<i>p</i>	$\eta^2$	<i>F</i>	
<i>The Netherlands</i>										
WS: Time	1	1.74	.001	.131	12.36	1.44	.039	.051	4.39	
BS: Group	1	4.09	.005	.093	8.43	17.09	.000	.150	14.43	
Time x Group	1	.00	.969	.000	.002	.004	.908	.000	.013	
*df error: 82										
<i>Germany</i>										
WS: Time	1	.18	.335	.011	.94	.00	.976	.000	.001	
BS: Group	1	6.47	.001	.119	11.31	2.81	.184	.021	1.80	
Time x Group	1	2.21	.001	.122	11.64	.849	.223	.018	1.51	
*df error: 84										
<i>Italy</i>										
WS: Time	1	2.95	.001	.177	12.71	.71	.204	.027	1.65	
BS: Group	1	1.34	.105	.044	2.71	4.31	.110	.043	2.64	
Time x Group	1	.254	.300	.018	1.09	.82	.171	.032	.824	
*df error: 59										

The scores for the development of EFL confidence also showed differences between groups and countries. In the Netherlands, all pupils developed significantly in their EFL confidence over time, but no significant differences were found between the CLIL and control groups:

$F(1,82) = 4,39, p = .039$ . In Germany, the CLIL groups became somewhat more confident while the control groups decreased in confidence. Conversely, in Italy, the control groups increased in EFL confidence but not the CLIL groups. There was thus no significant evidence of either German or Italian pupils as a group increasing in EFL confidence over the two years; however, a more fine-grained analysis showed some growth in confidence over time for the German CLIL group and for the Italian control group.

#### 4.11 Discussion

The aim of the present study was to explore whether the CLIL approach, as compared to the mainstream curriculum, adds value in enabling young European learners to develop into confident EFL speakers, prepared for life in a global world. Our research found that the CLIL programmes investigated in Germany, the Netherlands and Italy attracted pupils not only with higher initial EFL proficiency than mainstream

learners, as shown by our previous research (2013) but also with an above-average linguistic confidence and interest in the international world. This was the case in all three countries, even though there was some variation between groups and countries in the detail of results. At pre-test, the scores of Italian pupils on EFL confidence and international orientation lagged behind those of pupils in Germany and the Netherlands. This may be because their native language is of Roman, rather than Germanic, origins; but further, the introduction of EFL as a common curriculum subject in Italian primary education is of a relatively recent date as compared to the other two countries, where primary EFL lessons have been generally adopted over the past few decades. The Dutch CLIL pupils were the most confident EFL users, but the German CLIL classes also started with relatively high scores in EFL confidence and also had the highest initial scores on international orientations.

Our results suggest that the EFL CLIL approach in these three countries did not produce a significantly greater increase in learners' international orientation and language confidence than the mainstream approach: CLIL pupils developed positively, but so did mainstream pupils, and largely to the same degree. Assuming there was a general inclusion of a global dimension in the curriculum (Mannion et al., 2011), CLIL learners did not develop any particular advantage over non-CLIL pupils as far as international orientation was concerned; rather, it appears there was a positive development across the grammar schools on this construct.

In this respect, only a few differences between groups (CLIL v. mainstream) and between countries could be discerned. In the Netherlands and Italy, all groups - both CLIL and control - developed along the same lines: they all became significantly more positively orientated towards the international culture, foreign languages and the role of the English language. In Germany, however, the control group increased in international orientation while the CLIL group decreased. One explanation may be found in the fact that one of the two participant schools was located in Berlin, where CLIL primary schools with a variety of target languages are common practice in order to accommodate the multicultural population. Primary CLIL learners are immersed in international thinking from a very early age, which may well account for German pupils' high initial scores on international orientation. It may be that this, then, did not leave much room for growth. The mainstream pupils on the other hand started with lower scores, and it seems likely that they benefitted more from the general international orientation of their surroundings and contacts with CLIL school mates. As to EFL confidence, the variation in development between groups and countries was even smaller: no significant interaction effects were found.

In interpreting our results, several limitations should be borne in mind. In the first place, the number of participants was small and moreover, they are spread across

several nations, resulting in limited numbers of participants per group. The comparison of countries has led to interesting results, yet it is very hard to generalise them to a larger population. More comprehensive studies are necessary to come to robust value of CLIL. A further limitation lies in the fact that the CLIL groups started with higher initial scores than the control groups. It would have been useful to compare CLIL groups with control groups of mainstream learners having the same pre-test scores but who were studying in similar, but non-CLIL, schools. Even non-CLIL learners sometimes prefer schools offering a CLIL stream because of the international orientation this reflects; it is conceivable that the learning experience of these pupils in such schools is in fact influenced by the 'side-effects' of CLIL on the general school climate. Thus, the differences between the CLIL and mainstream results in the present study must be interpreted with caution.

Despite the fact that the results of the present study do not suggest conclusively that CLIL adds value in terms of promoting international orientation and EFL confidence, it should be noted that the CLIL programmes studied did generally provide an inspiring learning environment for the select group of the most motivated and confident EFL learners. The CLIL approach was generally successful in pushing high initial scores on international orientation and EFL confidence even higher.

Our research findings give rise to suggestions for future research. It would be helpful to investigate further the development of the non-cognitive variables discussed in the present study over a greater length of time. As the age of learners increases, their educational focus may shift. Their international orientation may be affected by what goes on in the world; their cultural interests may develop in a different direction, and likewise their confidence as an EFL user.

A notable outcome of the present study is that, while mainstream learners have far less contact with target language speakers than CLIL learners, this apparently does not mean that their EFL confidence develops very differently. This raises the question of how the wider range of mainstream learners would perform in a non-selective CLIL context. Marsh (2013) sees CLIL as a particularly appropriate educational approach for today's young language learners, and likely to benefit a broad range of learners, not just a privileged group. As mentioned earlier, this has been a recent initiative in Italy where EFL through CLIL has now been made mandatory in the last year of secondary school. Such initiatives call for further studies to investigate the extent to which the positive outcomes for CLIL, both in language proficiency and in international outlook, are maintained with non-selective implementation. A further issue would be to identify the strategies CLIL content teachers use to implement CLIL effectively with more diverse learner groups, and what attitudes and beliefs support them.



## **Appendix: Full list of questionnaire items International Orientation and EFL Confidence**

### *International orientation:*

1. I am good at languages
2. If I know English I can get to know other cultures and peoples
3. I need it for later studies
4. I want to know more about the lives of the English speaking nations
5. I want to write letters and e-mails to friends in foreign countries
6. I would like to work in a foreign country
7. If I know English I can learn more about what is happening in the world
8. I want to be like the English or Americans
9. I would like to make friends with foreigners
10. It will help when I am on holiday in a foreign country
11. I will get a better job if I can speak English
12. I would like to learn as many foreign languages as possible
13. I want to read English books and newspapers
14. It must be wonderful to live in America
15. It is interesting to learn more about English and American people
16. I would like to live in England
17. Most of my friends also want to learn English
18. I think America is a wonderful country
19. English people are friendly
20. Some of the most important people in our town are from England or America
21. On the whole you can trust English people
22. Later, after I have left this school, I will go on studying English
23. I would like to get to know more American people
24. On the whole I like English and American people

### *EFL confidence:*

1. I can easily write a text or a small story in English
2. Our English lessons are difficult
3. I never feel quite sure of myself when I speak English in the classroom
4. I feel uneasy whenever I must read or write an English text
5. I always feel that the other children in my class are better at English than I am

*Six-point Likert scale answering possibilities:*

1. This is absolutely true for me
2. This is almost true
3. This is a bit more true than untrue; more than half true
4. This is a bit more untrue than true; less than half true
5. This is almost untrue
6. This is absolutely untrue for me



# 5

## CHAPTER

**Determinants of EFL learning success  
in Content and Language Integrated  
Learning**

## Abstract<sup>1</sup>

Research has found that Content and Language Integrated learning (CLIL) programmes often select pupils who are already competent L2 learners. The present study investigates if CLIL learners' high scores for English as a Foreign Language (EFL) can be attributed to three specific learner variables: 'EFL aptitude', 'EFL confidence' and 'international orientation'. Additionally, the effect of out-of-school EFL exposure was taken into account. The study was undertaken in eight CLIL classes at secondary schools spread over four European countries: the Netherlands, Germany, Italy and Hungary. The results showed significant effects, both initial and after two years. We found that learners' EFL confidence has a stronger influence on L2 proficiency results than language learning aptitude, involvement in the international world or the presence of English in society at large.

*Keywords:* English-medium CLIL, EFL confidence, EFL aptitude, L2 exposure, international orientation.

### Reference:

Goris, J., Denessen, E., & Verhoeven, L.

Determinants of EFL Learning Success in Content and Language Integrated Learning.

*Revised manuscript submitted.*

---

1 Data for this chapter were collected between 2007 and 2009.

## 5.1 Introduction

Content and Language Integrated Learning (CLIL) has become a well-known educational approach in the present millennium. Almost all European countries have introduced it in some form, almost exclusively with English as the target language. Favourable effects mentioned in research frequently relate to EFL proficiency (Admiraal, Westhoff & De Bot, 2006; Verspoor, De Bot & Xu, 2015; Wolff, 2007).

Academic ability, together with the aptitude to learn a foreign language, take up a prominent position in the discussion of favourable CLIL results. Research confirms that CLIL learners are not only academically able but also above average L2 learners. Wolff (2007) claims that 'they process the FL more deeply and learn it more proficiently than traditional language learners' (p.21). Research findings as to L2 effects from CLIL in the European context were compared by Dalton-Puffer (2008). In a later study (2011) she found spontaneous oral production as the most noticeable CLIL effect. It has also been emphasized that pupils selected for CLIL already have better EFL skills than their mainstream peers before they start learning in the CLIL class (Huibregtse, 2001; Rumlich, 2017). As schools in most European countries apply selection criteria as to academic performance, linguistic skills and EFL motivation for admittance to CLIL classes, questions were raised as to the effect of this selection. If CLIL provides extra L2 learning opportunities for a select group of high achievers the effects of CLIL on pupil EFL performance may be overestimated due to a lack of control for selection effects (Bruton, 2011; Küppers & Trautmann, 2013; Paran, 2013).

The focus of the present study is the question if the better EFL performance of pupils selected for CLIL in four diverse educational settings is affected by certain learner variables. Specifically, we have concentrated on the effect of three constructs: 'EFL aptitude', a natural ability to learn the English language, 'EFL confidence', the confidence with which pupils use the L2 in the classroom and 'international orientation', which was considered to encompass a general interest in other languages and cultures, a sense of identification with target language speakers and a desire to use the target language for international communication, study and work. Additionally, we have investigated the effects of the presence of the English language in everyday life, notably in the media, which is often seen as promotive to L2 skills of young people. We selected four European countries with different native languages and diverging English-medium CLIL approaches: the Netherlands, Germany, Hungary and Italy. In the four countries the degree to which the target language is present in the media also differs, with the implication that out-of-school EFL exposure is not the same for all learners involved. We have analysed the effects of the three learner variables as well as the effect of the environmental factor on cognitive EFL proficiency results at the start of CLIL, and

evaluated how these effects had developed after two years' CLIL study at secondary schools in the four countries.

## 5.2 English as a CLIL target language

In the previous century EFL teaching used to be limited to two or three lessons a week - that is to say, in the schools that had English on the curriculum, which was not always the case throughout the European context. Moreover, English was generally not used as a vehicular language, needed to master subject competences: the foreign language was studied as a subject in itself in order to reach future L2 competence. Practice of spoken skills was limited and highly pre-structured. This changed in the 1990s, when Content and Language Integrated Learning - studying subject content by means of a foreign language - was introduced in mainstream education, mostly in secondary schools preparing learners for university. The position of English as a lingua franca had become worldwide and affected L2 education at schools, where it became the most studied foreign language as well as the most-used vehicular CLIL language. In CLIL education learners are encouraged to look across borders, in order to understand themselves and others in different cultures (Coyle et al., 2010). CLIL seems capable of a leading influence when it comes to developing learners with positive attitudes towards cultural diversity, who become aware of the responsibilities of global as well as local citizenship (Bentley, 2010).

In CLIL classes content and language are taught and learnt together, and the foreign language takes on an instrumental role. Learners have to develop what Marsh (2013) calls 'language awareness'; they have to move from viewing language learning as an object of study towards explicit understanding of how language is used in a variety of contexts. The language is the vehicle through which CLIL learners set out to master the contents of a variety of subjects. They have to master a large amount of subject-specific vocabulary and grammatical structures, and to develop communication skills in order to express thoughts and feelings and to interpret facts and data.

In Finland Seikkula-Leino (2007) investigated how successfully pupils had learnt content in CLIL, with a view to motivation, self-esteem and confidence in language learning. She found that if CLIL involves language that is still beyond the pupils' current competence, mastering subject matter becomes a demanding experience with the inherent risk that learners feel incompetent and doubtful as to their L2 skills. Attitudes and feelings towards learning and their effects on outcomes have been discussed in recent studies (e.g. Otwinowska & Forýš, 2015; Prüfer, 2013). The results indicate that positive feelings towards CLIL and the target language strongly interact with learning motivation and positive results.

### 5.3 CLIL in the Netherlands, Germany, Hungary and Italy

The countries in the present study each have their own approach to CLIL, and share characteristics with the views mentioned above in different ways. They all have English-medium CLIL - at the time of our data collecting almost exclusively at secondary school, primary CLIL was still an exception - in which language and content are learnt together, but the intensity, the dual focus and CLIL teacher training differ. In the Netherlands the English-medium programme comes close to the immersion approach, nearest to being content-driven. Dutch CLIL programmes show considerable uniformity as schools offering them are under supervision of The European Platform for Education in the Netherlands - merged with Nuffic in 2015 - which has laid down certain standards that have to be met before a school is an acknowledged CLIL school. The content lessons with English as a target language take up at least 60% of the curriculum as from the start of the first year of grammar school when pupils are generally twelve and beginning EFL learners. They have no formal preparation other than the usual English lessons in the last two years of primary school, which are the same for prospective CLIL and non-CLIL learners alike.

The German CLIL concept can be characterized as a cautious approach (Wannagat, 2007) even though regional variants differ. Germany consists of 16 states, each with its own educational policy and CLIL practice. Certain CLIL stipulations are laid down in curricular guidelines, such as the right of pupils to have CLIL content lessons not only in the foreign language, but also in their mother tongue. In most states in the western part of the country English is on the curriculum as from grade 3 at primary school when children are eight. Primary school leaving age is ten, and during the first two years at secondary school prospective CLIL pupils receive extra EFL training in preparation of the CLIL lessons that start when they are twelve. Only a small number of content subjects is involved in CLIL, while supportive lessons in the native language, German, are also offered: mostly history, politics and geography, but also sports lessons. A different type of German CLIL practice is found in Berlin, a city state in the East participating in the present study. CLIL is practised here on a larger scale and with more extensive programmes to accommodate the vast international community. CLIL classrooms with a variety of target languages exist both at primary and secondary mainstream education. English-medium content lessons generally take up more than 50% of the curriculum, a percentage that is not equalled in other German states. Primary CLIL pupils mostly continue their education at secondary schools of the same type, alongside pupils from monolingual schools and backgrounds, which accounts for an EFL gap in the first years (Zydatib, 2012).

Hungary needs special mention because of its L2 policy. As discussed by Dörnyei, Csizér and Németh (2006) political developments in Hungary had a major



effect on foreign language teaching. Russian used to be the compulsory foreign language at schools in Hungary, until the change of political regimes and the Education Act of 1985 granted more freedom and students massively opted for other modern European languages, mainly German and English. Hungary joined the EU in 2004 and in the following years English not only became the first L2 in the school curriculum but also the language of instruction for a number of subjects in secondary education alongside German (Farkas & Kniezsa, 2002). Primary school curricula include foreign languages, mostly but not necessarily English, which is why CLIL grammar school classes sometimes accommodate pupils without any prior knowledge of English. In order to overcome initial EFL gaps and to train the target language thoroughly most CLIL schools offer a preparatory year, popularly known as 'zero year'. The normal class of 36 pupils is divided into three groups of twelve that study the target language in 16 to 20 lessons per week, in which a native speaker language teacher is involved. Unlike in the CLIL approaches in the other three countries the target language is learnt separately, which is seen as a necessary preparation both as to general proficiency and the jargon needed for the content subjects. English-medium content teaching, often by native speaker teachers, starts in the following year and takes up 50% of the curriculum.

Italy also has its own approach. Italian schools are in regional networks under the supervision of their Local Educational Authorities and CLIL practice varies according to region. Recent developments in Italian CLIL programmes relate to legislation as from the year 2010, which obliges all secondary high schools (*Licei* and *Istituti Tecnici*) to teach a non-language subject through the medium of a foreign language in the final year. At the time of the first measurements of the present study, however, the Italian CLIL programme was in its initial stages, had a modular form and was highly selective. The teaching of a content subject by means of the target language was limited to a set number of modules which took up about 20% of the curriculum, a percentage that was increased in the following years. The preferred solution for teaching in the CLIL mode consisted in team-teaching, which means that two teachers – the subject teacher and the EFL teacher – were cooperating in the classroom, distributing their focus of attention according to a planned procedure for the development of the lesson and according to both the linguistic and content needs of the pupils as these arose (Coonan, 2012, p.119). The pupils do not receive special EFL training in preparation of CLIL, they are beginning EFL learners with three years of English lessons at the usual *Scuola Media*, or, to use more recent terminology, *Scuola secondaria di primo grado*, the lower three-year secondary school.

#### 5.4 The presence of English in society at large in the four countries

Language exposure, or the degree to which the foreign language is present in society at large, is seen as a factor of influence on the L2 learning process as it provides extra input in addition to formal learning at school (Doiz, Lasagabaster & Sierra, 2014). Even though the English language has the status of the world's lingua franca and is also the foreign language which is learnt the most in education, its presence in everyday life varies across the four countries. A major study by Berns, de Bot and Hasebrink (2010) mentions the fact that English is the daily language in most international companies in the Netherlands, and that "job announcements implicitly assume potential employees' English skills and only mention English proficiency when very special skills or near-native command is necessary" (p.20). The Special Eurobarometer, a survey co-ordinated by the European Commission in 2012 mentions that 90% of the respondents in the Netherlands 'speak the English language well enough in order to be able to have a conversation' and 57 % is able to follow television or radio news in English. English TV broadcasts are subtitled and dubbed programmes are unheard of in the Netherlands (whereas they are common practice in Germany, Hungary and Italy). In these countries the English language is much less present. For Germany the Eurobarometer indicates conversational skills in English for 56% of the population, while English-spoken films are dubbed in the German language and television broadcasts exclusively in English are rare occasions. As Berns, de Bot and Hasebrink discuss in their study (2010), the younger generation has a preference for music in the English language on the radio, but the debate on the share of German language music continues to be in favour of quotas.

Further away from the anglicised society of the Netherlands are Italy and Hungary, geographically but also in other respects. From a linguistic point of view the Italian and notably the Hungarian language are remoter from English than Dutch and German, which are both of Germanic origins. The Eurobarometer mentions Hungary and Italy as two of the countries where respondents are least likely to be able to speak any foreign language. The survey shows that in Hungary 20% of the respondents had conversational skills in English, while in Italy this is 34%. English is hardly present in the media in either country. In Hungary restrictive laws for the use of English in the media apply and Hungarians live in a relatively dubbed world as the main TV channels that are available for everyone run shows which are translated into Hungarian (Pétery, 2011). A similar limited presence of the English language exists in Italy (Doiz et al., 2014). English-medium broadcasts are scarce in the media and television programmes are dubbed. If people in countries with dubbing practice want to watch a film in English, they have to look for it on the Internet or on DVDs, an effort that is not always taken for granted by EFL learners.

### 5.5 The present study

As we have discussed, adequate EFL proficiency is necessary for overall school results in CLIL classes. In a previous study (Authors, 2013) we found that pupils selected for CLIL in the Netherlands, Germany and Italy had significantly higher EFL scores than mainstream learners, both initially and after two years. In a second study involving the same population (Authors, 2017) we discussed the fact the CLIL learners had a greater confidence in their ability to use the English language than the mainstream learners, which also applied to more positive attitudes towards the international world. In the present study we have studied CLIL learners in four diverse educational contexts. The research focus is on the benefits for EFL learning resulting from two affective learner variables, 'EFL confidence' and 'international orientation'. In addition, we have studied the effects of the language learning aptitude that is often associated with CLIL learners, 'EFL aptitude' and the presence of English in the media in the Netherlands, Germany, Hungary and Italy.

The research questions for the present study are as follows:

- What are the effects of the learner variables EFL aptitude, EFL confidence, international orientation and the environmental variable out-of-school exposure on CLIL pupil entry levels of EFL proficiency?
- What are the effects of these variables on EFL proficiency growth in the first two years of CLIL in classes at secondary schools preparing for university?
- To what extent are there differences across the four countries?

In order to answer the research questions we analysed the data of two measurements in eight CLIL classes in the four countries.

## Method

### 5.6 Participants

The eight schools participating in the present study can be classified as academically oriented schools preparing students for study at university. In the Netherlands the acronym *VWO*, the term *Atheneum* and *Gymnasium* are used for this type of school, in Germany they are known as *Oberschule* or *Gymnasium*, in Italy the *Ginnasio* and various types of *Licei* exist and in Hungary the *Gimnázium*. The schools all have classes with a CLIL programme for which they apply admittance criteria. Each of the schools took part in the present research with an English-medium CLIL group. Table 1 shows the distribution of the participant pupils.

Table 1: Distribution of CLIL pupils (N=162)

	N	age	boys	girls
<i>The Netherlands</i>	37	12.3	19	18
<i>Germany</i>	42	12.6	21	21
<i>Hungary</i>	39	15.0	15	24
<i>Italy</i>	44	14.3	30	14

As the duration of primary school varies across countries, so does pupil age at the start of secondary education. In the Netherlands children generally start secondary school at twelve, which is also the start of CLIL. In German states in the western part of the country they start secondary education at ten, while the CLIL programme starts two years later. In the states situated in the East so also in Berlin primary school leaving age is twelve; pupils from CLIL primary schools - mainly found in Berlin - are already experienced CLIL learners, whereas for those from regular primary schools the start of CLIL coincides with the start of secondary education. In Italy pupils are fourteen when they enter the type of upper secondary education of their choice, after completion of the *Scuola Media*, while in Hungary the admittance age of pupils to the *Gimnázium* schools is fifteen. Table 2 presents an overview of CLIL conditions and EFL teaching in the participating schools.

Table 2: CLIL conditions in the Dutch, German, Hungarian and Italian schools

	Netherlands		Germany		Hungary		Italy	
	school 1	school 2	school 1	school 2	school 1	school 2	school 1	school 2
CLIL since:	2000	1998	1991	1999	1988	1987	2003	1998
<b>Admittance</b>								
Entrance test				x	x	x	x	x
Prim.school reports	x	x		x	x			
Motivation	x	x	x					
<b>Subject taught by:</b>								
Subject teacher	x	x	x	x	x	x		
EFL+ subject teacher							x	x
<b>Subject teacher L2 training:</b>								
Dual qualifications (EFL + subject)			x	x				
Extra EFL training	x	x			x	x	x	x
Native speaker					x	x		
<b>CLIL lessons per week</b>								
Year 1	16	15	5	8	0	0	2	6
Year 2	16	15	7	10	12	14	5	9
<b>EFL lessons per week</b>								
Year 1	5	3	4	5	19	16	3	3
Year 2	4	2	4	5	2	5	3	3

## 5.7 Instruments

In order to answer the research questions several tests and questionnaires were presented to the participants. A comprehensive EFL proficiency test was completed at the start of the grammar school CLIL programme. For the EFL testing procedure our main principle was that L2 proficiency is an integration of knowledge of words, expressions, insight into the rules of the language and an ability to understand its written texts. We selected assignments of receptive vocabulary, grammar, idioms and text comprehension for the two tests. In order to discriminate between the various language skills we chose to conduct a separate analysis for each sub part of the EFL tests corresponding to these notions. They consisted of a number of pen-and-paper assignments from standardized tests, designed for international use at secondary school. To measure pupil EFL proficiency growth a second test of similar construction was presented after two years.

Questionnaires for the language learning ability EFL aptitude, the non-cognitive learner variables EFL confidence and international orientation and the contextual variable out-of-school L2 exposure were completed by the end of the first term at

secondary school. These questionnaires were presented in English, which was a second language for the participants. In view of the fact that they were all beginning EFL learners we took great care in formulating the questions in simple language, and discussed learner comprehensibility with their respective English language teachers before administering them to the pupils.

The questionnaire we used to measure EFL aptitude was based on Carroll and Sapon (1959), the authors of the Modern Language Aptitude Test (MLAT). They defined language aptitude tests as referring to 'the prediction of how well, relative to other individuals, someone can learn a foreign language in a given amount of time and under given conditions'. The related MLAT-Elementary test was developed for the purpose of measuring the language aptitude of American children in grades 3 – 6, where it is an instrument to help teachers determine a child's readiness to learn a foreign language. Carroll (1959) demonstrated that foreign language aptitude comprises four cognitive abilities. These abilities are reflected in one way or another in the tests that were developed subsequent to Carroll's research. In order to measure our participants' aptitude to learn English we used parts of the Modern Language Aptitude Test – Elementary (MLAT-E). We were well aware of the fact that this test was originally designed to be given in English to English speaking children. Our decision to use it for beginning EFL learners with various native languages other than English was based in the first place on the fact that the Elementary version was in simple language, of a level that all participants mastered. Secondly, we based our choice on research into similarities between L1 and L2 acquisition (e.g. Ipek, 2009). We also found support in the discussion of language aptitude by Skehan (1989) who argues that language aptitude tests are supposed to measure - at least partly - an underlying language and learning capacity which is similar in L1 and L2. We used parts that could be presented as a pen-and-paper test to help us identify EFL aptitude of the L2 learners in our study ( $\alpha = .91$ ). The tested parts related to phonetic coding ability, the ability to associate sounds and symbols (Hidden Words); to the ability to discriminate between speech sounds (Finding Rhymes) and to grammatical sensitivity, the ability to recognize the function of a lexical element in a sentence (Matching Words). For 'Hidden Words' there were 20 items of disguised words, e.g. 'smmr', with four multiple choice answers, in which the correct answer had to be chosen from 4, in this case : 1. An animal 2. Good 3. Season 4. To drink. The 20 items for 'Finding Rhymes' had the same structure; for this task a rhyming word had to be selected for the prompt from 4 possibilities, e.g. 'rain': 1. Vine 2. Lane 3. Keen 4. Fine. For 'Matching Words' ten pairs of sentences were presented with a bold and underlined word or word group in the first sentence. In the second sentence the word or word group with a corresponding grammatical function had to be underlined, e.g. 'Our English teacher gave **us** a very difficult test' / 'I sent my friends a postcard

when we were on holiday'. In this case 'my friends' had to be underlined as its function corresponds to the dative 'us'.

EFL confidence and international orientation were measured by means of questionnaires in English with statements asking for responses on a six-point Likert scale. The confidence the participant pupils felt when using the English language was measured by means of responding to five statements ( $\alpha=.75$ ) on a six-point Likert scale, e.g. 'I feel uneasy whenever I must read or write an English text'. The statements were based on Gardner's concept of integrative motivation (1985) with addition of the concept of linguistic confidence of Dörnyei and Skehan (2003). In the same manner, they responded to items pertaining to international orientation; there were 24 statements on a six-point Likert scale ( $\alpha=.88$ ). Contents were also based on Gardner (1985) together with MacIntyre et al.'s definition of 'willingness to communicate' (1998).

Out-of-school exposure to the English language was measured in terms of its presence in the media in the form of TV broadcasts and pop songs. Even though the English language is present in all four countries, a major divide seemed to be between countries where films and TV programmes are dubbed versus subtitled in the national language, along with the presence of pop songs in English. Pupils completed a list with questions about the average time they usually spent on watching English-medium TV and listening to English-sung pop songs in minutes per day, such as 'Do you listen to pop-songs in English? How many minutes per day, on average?' For a full description of the questionnaires for EFL confidence and international orientation as well as the contextual variable out-of-school L2 exposure see the appendix.

### 5.8 Procedure

The data were collected in two rounds of measurement. At the start of the school year 2007/2008 the participant pupils completed the first EFL proficiency test as well as the EFL aptitude test, the questionnaires as to EFL confidence, international orientation and out-of-school exposure to the English language. In June 2009 they took the second EFL proficiency test. The test leader, a university researcher, informed the pupils about their participation in the university research project. They were told that taking part was anonymous and that their answers only served the research purpose. Completion of the proficiency test and questionnaire took one lesson period each.

### 5.9 Analysis

The mean scores and standard deviations of the pupil variables EFL aptitude, EFL confidence and international orientations and the environmental factor out-of-school L2 exposure were processed for each country. Likert type items were assumed to be assessed on (quasi-)interval level to enable the calculation of total scale scores. The

same was done to get the results for the pupil EFL scores for the EFL skills vocabulary, grammar, idioms and reading comprehension, for each of the four countries and for both measurements, test 1 and test 2. After that ANOVAs were conducted to measure the interaction effects at entry level: to what degree did the pupil variables and the environmental variables of out-of-school exposure affect cognitive EFL results at the start of the CLIL programme at grammar school? After that, ANOVAs of the repeated-measures design were conducted to measure the above mentioned effects of the pupil and environmental variables on EFL proficiency scores growth, the increase in test scores for the categories of the language skills after two years' time. Finally, a regression analysis was conducted for more information about significant effects.



## Results

### 5.10 Scores for construct descriptors

Descriptive statistics for EFL aptitude, EFL confidence and international orientation for each of the four countries are shown in the top part of Table 3. Here we need to explain the origins of the numerical divergence. EFL aptitude scores were based on the 50 items in the test, so had a maximum of 50 and a minimum of 0. The Likert-scale questions for EFL confidence and international orientation gave six possible options, which we counted as a minimum of one and a maximum of six. The bottom part of Table 3 shows the scores for environmental factors, the practice of watching television or listening to pop songs in English. The scores represent the average of minutes per day pupils spend on these activities.

Table 3: Initial scores for pupil scales, itemized for country.

	<b>Netherlands</b>		<b>Germany</b>		<b>Hungary</b>		<b>Italy</b>		<b>F</b>
<i>Pupil factors</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>
EFL aptitude	36.19	5.04	33.81	10.79	34.82	8.72	33.02	8.95	.98
EFL confidence	4.49	.76	4.26	1.05	3.86	1.06	3.80	1.04	4.36
Intern. orientation	3.14	.54	2.61	.57	3.05	.58	3.32	.65	11.09
<i>Environmental</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>
Watching TV	80.27	61.71	17.24	36.28	45.33	48.35	13.68	21.58	19.72
Listening to pop	90.24	68.71	70.12	65.22	78.21	65.41	54.64	67.65	2.04

As the overview makes clear, EFL aptitude scores did not vary greatly across the four countries: all pupils had about the same talent for EFL learning:  $F(3,158) = .98$ ,  $p = .404$ . The Dutch scores were slightly higher than those in the other three countries, but there were no significant differences. As to EFL confidence the Netherlands and Germany seem to be reasonably matched, which is also true for Hungary and Italy. The ANOVA results showed significant differences between the countries:  $F(3,158) = 4.36$ ,  $p = .006$ . By conducting post-hoc Bonferroni tests these differences were found to be these differences were found between the Netherlands and Hungary ( $p = .038$ ) and the Netherlands and Italy ( $p = .013$ ). As to international orientation significant effects were found:  $F(3,158) = 11.09$ ,  $p < .001$ . The post-hoc tests showed significant differences between Germany and the Netherlands ( $p = .001$ ), between Germany and Hungary ( $p = .006$ ) and between Germany and Italy ( $p < .001$ ). The Italian learners were more internationally orientated than their peers in the other three countries, of which the German learners scored lowest.

As to the out-of-school L2 exposure several differences can be seen, within and between countries as well as within groups of learners. As could be expected in view

of the fact that English-medium television broadcasts are frequent in the Netherlands, the time spent on watching TV was highest here. As to the other three countries - with fewer English-medium TV broadcasts - the Hungarian figures were surprising, as they were higher than those in Germany and Italy. A possible explanation is the effect of the increased interest for the English language in Hungarian society, which is also marked by the intensive L2 programme of the zero year at grammar schools. The ANOVA showed significant effects:  $F(3,158) = 19.72, p < .001$ . The post-hoc tests showed these significant differences to be between the Netherlands and Germany ( $p < .001$ ), the Netherlands and Hungary ( $p = .004$ ) and the Netherlands and Italy ( $p < .001$ ) and also between Hungary and Germany ( $p = .026$ ) and Hungary and Italy ( $p = .007$ ). Listening to pop-songs in English scored high in all four countries, but again highest in the Netherlands where other foreign languages or the native language are hardly present in popular music. The differences we found between the countries were not significant.

### 5.11 Scores for EFL proficiency test 1 and test 2

The EFL test scores in each of the four countries at the start of the CLIL intervention (Test 1) and again after two years (Test 2) for vocabulary, grammar, idioms and reading comprehension are shown in Table 4. As not all assignments of each subpart had the same number of items in Test 1 and Test 2, we could not measure progress numerically for each subpart. For vocabulary it was straightforward: each test consisted of 180 items, which means a score range from 0 to 180 and a numerical growth. This was also the case for idioms: the idioms tests 1 and 2 had six items each, so a score range from 0 to 6. For grammar the proceedings were different. Grammar test 1 had 66 items, test 2 had 20 items; therefore we calculated score means - for test 1 the group mean divided by 66, for test 2 the group mean divided by 20. The score growth is based on these means. As with grammar, for reading comprehension we had to consider the fact the scores that could be reached did not run parallel in the two tests, which presented a problem for numerical comparison of progress. Moreover, test 1 had five multiple choice and three open questions, test 2 had nine multiple choice and three open questions. As scoring multiple choice questions differs from scoring open questions we preferred to use Z-scores to measure progress. For this reason some results for reading comprehension are negative when the score means of the country are below the population average.

As can be seen in Table 4, the test results varied considerably, both within and across countries. The initial test scores showed significant differences between groups for vocabulary:  $F(3,158) = 3.50, p = .017$ ; for grammar:  $F(3,158) = 10.35, p < .001$ ; for idioms:  $F(3,158) = 5.69, p = .001$  and for reading comprehension:  $F(3,158) = 4.48, p = .005$ . Both the Dutch and German scores for vocabulary were significantly higher than those in Italy

( $p=.029$  and  $.047$  respectively), while for grammar the Dutch scores were significantly lower than those in Germany ( $p=.002$ ), Hungary ( $p<.001$ ) and Italy ( $p=.016$ ). For idioms the Dutch scores were significantly higher than those in Hungary ( $p<.001$ ) while in Germany reading comprehension scores were significantly higher than in Italy ( $p=.012$ ). Also the score growth after two years showed significant effects between groups: for vocabulary:  $F(3,158) = 3.88, p=.010$ ; for grammar:  $F(3,158) = 20.59, p<.001$ ; for idioms:  $F(3,158) = 5.95, p=.001$  and for reading comprehension:  $F(3,158) = 16.23, p<.001$ . In Hungary the score growth for vocabulary was significantly greater than in Germany ( $p=.006$ ). In the Netherlands score growth for grammar was significantly greater than in Germany and Hungary ( $p<.001$  and  $.014$  respectively) while the Hungarian and Italian grammar scores showed a significant growth when compared to those in Germany (for both  $p<.001$ ). Significant differences in score growth were also found for idioms; in Hungary they increased more than in Germany ( $p=.002$ ) and in Italy ( $p=.009$ ). As to score growth for reading comprehension this was significantly more in the Netherlands, Hungary and Italy than in Germany ( $p<.001$ ).

Table 4: EFL proficiency scores for the first and second tests, itemized for country.

EFL scores	The Netherlands						Germany						Hungary						Italy						
	Test 1		Test 2		Test 1		Test 2		Test 1		Test 2		Test 1		Test 2		Test 1		Test 2		Test 1		Test 2		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Vocabulary	136.65	16.63	156.38	10.51	134.93	24.68	140.90	20.55	128.97	35.57	155.36	13.24	117.07	39.21	133.09	21.67									
Grammar	.57	.14	.72	.09	.69	.15	.58	.21	.75	.14	.79	.14	.67	.13	.74	.15									
Idioms	2.59	1.21	3.97	.87	2.05	1.27	2.57	1.48	1.49	1.21	3.31	1.36	2.16	1.03	2.86	1.36									
Reading comprehension	-.33	.96	.31	.93	.55	1.07	-1.27	1.91	.39	1.95	.29	1.21	-.59	2.27	.70	1.33									

Table 5: Comparative survey of effects

	Vocabulary		Grammar		Idioms		Reading Comprehension												
	Initial	Growth	Initial	Growth	Initial	Growth	Initial	Growth											
df	SS	F	SS	F	SS	F	SS	F											
<i>Pupil factors</i>																			
EFL aptitude	1	2624.85	3.65	.53	187.89	.53	20.83*	.03	2.36	3.45	2.71	.00	.00	7.42	2.78	3.47	1.78		
EFL confidence	1	14225.77	19.80*	4.27*	1505.98	4.27*	.161	12.42*	.02	1.51	.90	.71	3.58	3.07	21.88	8.21*	2.89	1.49	
International orientation	1	53.02	.07	.23	80.87	.23	.04	3.15	.00	.01	1.67	1.31	1.54	1.32	.60	.23	.74	.38	
<i>Environmental factors</i>																			
Watching TV	1	371.49	.52	1.24	436.35	1.24	.02	1.20	.00	.02	.57	.45	.01	.01	6.61	2.48	5.02	2.58	
Pop songs	1	1214.91	1.69	.32	112.32	.32	.04	2.95	.02	1.42	.74	.58	4.01	3.44	1.98	.74	6.37	3.27	
Country	3	210.44	.10	.25	266.04	.25	.03	.67	.09	2.78*	3.33	.87	2.36	.68	6.21	.78	13.36	2.29	
<i>Interactions</i>																			
<i>Pupil factors* country</i>																			
EFL aptitude	3	1472.65	.68	1.05	1109.17	1.05	.06	1.4	.04	1.05	5.81	1.52	11.44	3.27*	8.25	1.03	1.79	.31	
EFL confidence	3	6330.91	2.94*	3.21*	3394.99	3.21*	.01	.24	.01	.42	15.37	4.02*	7.63	2.18	6.11	.76	16.48	2.82*	
International orientation	3	2369.18	1.10	.78	822.06	.78	.02	.38	.03	.76	1.36	.36	2.73	.78	3.03	.38	.99	.17	
<i>Environmental factors* country</i>																			
Watching TV	3	4035.34	1.87	1.60	1690.54	1.60	.02	.50	.01	.38	9.39	2.46	5.18	1.48	2.66	.33	4.90	.84	
Pop songs	3	1578.94	.73	.96	1015.86	.96	.10	2.44	.05	1.53	1.84	.48	4.74	1.36	3.04	.38	5.30	.91	
Df error:138																			
R2-model		.38					.52				.28						.24		

\* p < .05

### 5.12 Effects on EFL proficiency results and growth

To answer the first research questions an analysis of variance (ANOVA) was conducted to measure the effects of the constructs on the EFL proficiency scores for vocabulary, grammar, idioms and reading comprehension of the first test, which pupils performed at the start of CLIL. A second ANOVA, to answer the second and third research questions, measured their effects on EFL proficiency growth in the first two years of CLIL at grammar school and the differences between countries. The ANOVAs display the variety of parameters for reasons described above, which makes for numerical variety. However, this does not affect the outcomes of effects. The results are shown in Table 5.

For the initial measurement a number of significant results were found. EFL aptitude had a significant effect on grammar; aptitude for language learning benefits mainly grammar, likewise in all four countries:  $F(1,161) = 20.83, p < .001$ . There was a slight but significant linear regression:  $\beta = .039, p < .001$ .

The confidence learners had in their ability to use the English language, EFL confidence, produced several significant initial effects for EFL proficiency results and greater linear regression. A significant effect was found for vocabulary scores:  $F(1,161) = 19.80, p < .001$ . Effect sizes were small and varied per country:  $R^2$  for the Netherlands was .222, for Germany .029, for Hungary .420 and for Italy .101. Linear regression analysis also indicated significance:  $\beta = .42, p < .001$ . More confidence leads to higher vocabulary scores in the four countries, even though there were significant differences:  $F(3,138) = 2.94, p = .036$ . Regression results for the Netherlands were:  $\beta = .47, p = .003$ ; for Germany:  $\beta = .17, p = .28$ ; for Hungary:  $\beta = .65, p < .001$ ; for Italy:  $\beta = .32, p = .035$ . In the Netherlands, Hungary and Italy learners with more confidence in their English language skills reach higher vocabulary scores than in Germany.

A significant initial effect of EFL confidence was also found for grammar:  $F(1,161) = 12.42, p = .001$ ;  $\beta = .28$  and for reading comprehension:  $F(1,161) = 8.21, p = .005$ ;  $\beta = .29$ . The confidence of learners in their EFL skills influenced their scores for grammar and reading comprehension largely to the same degree in the four countries and with average effect sizes of  $R^2 = .073$  for grammar and  $.083$  for reading comprehension. As to the relation of EFL confidence and scores for idioms, however, there were significant differences between the countries:  $F(3,138) = 4.02, p = .009$ . The effect sizes varied and were smallest in the Netherlands ( $R^2 = .019$ ) and Germany ( $R^2 = .001$ ) and slightly larger in Hungary ( $R^2 = .162$ ) and Italy ( $R^2 = .133$ ). Regression results also varied: the Netherlands:  $\beta = -.14$ ; Germany:  $\beta = .03$ ; Hungary:  $\beta = .40, p = .01$ ; Italy:  $\beta = .37, p = .015$ . In Italy and Hungary EFL confidence has significantly more influence on scores for idioms than in Germany and the Netherlands, while in the Netherlands we even found a negative effect: more confidence leads to lower scores for idioms. A possible explanation is that learners in

the Netherlands feel over-confident: they can understand pop songs and TV broadcasts, hence feel less inclined to spend time studying idioms.

The pupil variable international orientation and the environmental variables did not produce significant effects: the fact that pupils in CLIL classes feel attracted to the international world, listen to pop songs or watch English TV had no significant effect on their EFL scores. However, correlation analysis of the constructs showed several significant results. Five correlations were significant, even though small:

EFL Aptitude and EFL confidence ( $r = .18, p < .05$ )

EFL Aptitude and Watching TV ( $r = .19, p < .05$ )

International orientation and Listening to pop songs ( $r = .19, p < .05$ )

EFL confidence and Watching TV ( $r = .33, p < .01$ )

Listening to pop songs and Watching TV ( $r = .23, p < .01$ )

The figures show that learners with more EFL aptitude and confidence also watch slightly more English TV broadcasts, while an international orientation goes hand in hand with listening to pop songs.

The second ANOVA was aimed at analysing the effects of the pupil and environmental variables on EFL proficiency growth in the first two years of CLIL - the score increase for the English language assignments of Test 1 in comparison with Test 2. There was a significant difference across countries as to the degree to which grammar scores developed:  $F(3,161) = 2.78, p = .044$ . In the Netherlands they increased significantly more than in Hungary and Italy, while in Germany the scores for grammar decreased.

The pupil and environmental constructs showed various effects. International orientation and the environmental variables did not produce significant effects on the growth of English proficiency scores in two years' time; they mainly reinforce each other somewhat, as shown by the correlation figures, rather than lead to cognitive benefits for EFL learning. Results were highly similar across the four countries, irrespective of differences in CLIL concepts or the presence of English in society at large.

EFL confidence proved to be a much stronger influence. We found a significant effect on the score growth for vocabulary:  $F(1,138) = 4.27, p = .041$ . The effects varied across countries:  $F(3,138) = 3.21, p = .025$ . The effect sizes were for the Netherlands:  $R^2 = .071$ , for Germany:  $.001$ , for Hungary:  $.315$  and for Italy:  $.000$ . Linear regression analysis also showed variety and indicated for the Netherlands:  $\beta = -.27$ ; for Germany:  $\beta = .03$ ; for Hungary:  $\beta = -.56$  and for Italy:  $\beta = .02$ . This means that in Hungary and the Netherlands the regression results were negative: pupils with more confidence showed a decrease of scores for vocabulary. This may be because in these two countries vocabulary scores were very high at the second test so that further growth might have been unlikely. EFL confidence also showed significant differences in effects per country for reading comprehension score growth:  $F(3,138) = 2.82, p = .041$ . The effect sizes, however, were

small:  $R^2$  varied from .085 for the Netherlands, .024 for Germany, .001 for Hungary and .089 for Italy. Linear regression analysis indicated for the Netherlands:  $\beta = -.29$ ; for Germany:  $\beta = .16$ ; for Hungary:  $\beta = .02$ ; for Italy:  $\beta = -.30$ . Dutch and Italian results were negative: more confidence led to a smaller score increase, similar as in cases of over-confidence. This is not easy to explain. Pupils may lose interest in reading texts as they grow older, find the subject boring or take comprehension questions too lightly.

The learner variable EFL aptitude, the ability to learn the English language, affected the increase of scores for idioms with significant differences across the four countries:  $F(3,138) = 3.27, p = .023$ . Again, effect sizes were small:  $R^2$  for the Netherlands was .005, for Germany .030, for Hungary .057 and for Italy .001. Regression for the Netherlands was:  $\beta = -.07$ ; for Germany:  $\beta = -.17$ ; for Hungary:  $\beta = .29$ ; for Italy:  $\beta = .03$ . In Hungary an aptitude for EFL learning related stronger to increasing scores for idioms than in the other three countries. A possible explanation lies in the fact that the Hungarian pupils spent considerably more time in the EFL classroom – during the intensive zero year – and thus had to study a greater amount of idioms, unrelated to their own native language, for which they may have had to rely more on their language aptitude than pupils in the Netherlands and Germany.

### 5.13 Discussion

The aim of the present study was to analyse if the specific aptitude for and attitudes towards EFL learning of starting CLIL learners are related to their EFL proficiency results, and if this is similar in divergent CLIL contexts. Our results suggest an affirmative answer mainly for EFL confidence, the confidence with which pupils use the English language. The language learning ability EFL aptitude was found to have a significant effect on increased knowledge of idioms only. EFL confidence, on the other hand, was found to have significant effects on EFL proficiency scores in more respects. Learners with more confidence in their language skills acquired higher scores for vocabulary and reading comprehension, which seems an adequate contribution of CLIL to a future career or study in an international context. Confidence in one's EFL skills was found to be of more weight for score growth than an aptitude for language learning, and also more relevant than the degree to which learners feel attracted to the international world or the degree of presence of English in society at large. No effects were found for these constructs which may be explained by the fact that society as a whole is becoming more internationalised, which makes this less of a prerogative for CLIL schools. As we have discussed, the environmental factor of out-of-school exposure showed positive correlations and may be seen as a side contribution to learner confidence and aptitude.



To conclude, the main finding of the present study is related to the effect of EFL confidence, the confidence with which pupils use their English language skills which we found to have effects both initially and after two years. Even though there were certain differences between countries, learner EFL confidence appeared to be a more important influence on cognitive proficiency results than the aptitude for learning English. The extent to which the pupils possessed this language learning ability was probably more or less what was required for admittance to CLIL classes, which could also explain why there was no convincing contribution to EFL score growth. Even when CLIL is selective and relies on language learning aptitude, our findings suggest that learners with confidence in their own EFL skills profit the most from the CLIL environment: the CLIL classroom considerably enhances the effect of such confidence.

We must mention the fact that the present study was not without limitations in its execution, notably as to measuring aptitude specifically for English as a Foreign Language in an international context of beginning learners. At the time of our data collection the traditional language aptitude tests that were available had been developed in the 1960s and 1970s - of which the MLAT is a well-known example. They were intended - either in their original English forms or in translations - for participants with the same native language. The development of tests based on a novel conceptualization of foreign language aptitude, in which participants are expected to learn elements of a new, artificial language - not resembling any one language or linguistic group, see Grigorenko et al. (2000) - was still fairly recent at the time. Little was known about the findings of such tests in educational practice, after a long period of limited interest in language aptitude research. Moreover, our specific focus was on aptitude for learning English. We therefore preferred using parts of the English version of the MLAT-E as an instrument, and chose to maintain their original English language in order to prevent bias which might occur in the case of translations into the four native languages of our participants. Our findings relating to learner EFL aptitude should therefore be interpreted with the above limitations in mind. Looking at the present State-of-the-Art which has witnessed a 'renewed enthusiasm across multiple disciplines of educational psychology, second language acquisition and cognitive neuroscience' (Wen et al., 2017) for aptitude research, the development of an instrument specifically for EFL aptitude in an international context to be validated in future research could be a welcome contribution to the CLIL research field.

Finally, we must bear in mind that the present study is a small scale one and therefore the results cannot be generalised. Moreover, there were no control classes so we cannot compare the effects of our constructs in the CLIL classes with mainstream classes. However, four very diverse CLIL contexts were investigated. Across the countries CLIL was found to provide a positive learning environment for learners with

confidence in their EFL skills, and to enhance its effect on their EFL proficiency scores in the course of time. The specific role of CLIL as to the role of language confidence raises the question what can or should be done for learners who may be very talented but lack this confidence. In view of the fact that CLIL is increasingly being considered for a larger group of learners this seems a relevant concern.

## Appendix

### *1. Full list of questionnaire items International Orientation and EFL Confidence*

#### **International orientation:**

1. I am good at languages
2. If I know English I can get to know other cultures and peoples
3. I need it for later studies
4. I want to know more about the lives of the English speaking nations
5. I want to write letters and e-mails to friends in foreign countries
6. I would like to work in a foreign country
7. If I know English I can learn more about what is happening in the world
8. I want to be like the English or Americans
9. I would like to make friends with foreigners
10. It will help when I am on holiday in a foreign country
11. I will get a better job if I can speak English
12. I would like to learn as many foreign languages as possible
13. I want to read English books and newspapers
14. It must be wonderful to live in America
15. It is interesting to learn more about English and American people
16. I would like to live in England
17. Most of my friends also want to learn English
18. I think America is a wonderful country
19. English people are friendly
20. Some of the most important people in our town are from England or America
21. On the whole you can trust English people
22. Later, after I have left this school, I will go on studying English
23. I would like to get to know more American people
24. On the whole I like English and American people





# 6

## CHAPTER

### Conclusion and Discussion

In this final chapter the conclusions of our research will be outlined. Before we start, we think it relevant to outline recent developments in the CLIL field in view of the fact that our data were collected some time ago. Our conclusions should be read with this information in mind. We will then discuss outcomes as to the effects we found of the innovative CLIL approach, the teaching and learning of English and subject matter in an integrated way at mainstream schools in various European quarters. We will first look at L2 outcomes, to see to what extent the CLIL target of providing better L2 learning opportunities has been realized. After that, we will discuss our findings as to the CLIL contribution to benefits for individual learners and preparation for the future, and see to what extent the success of learning in a CLIL class may be attributed to learner characteristics. Finally we will discuss the overall merits of CLIL education that we found in our research.

### **6.1 Some preliminary remarks**

The aim of the present thesis is to investigate CLIL and its effects in the first stages of its introduction - in the highly selective context of secondary schools preparing for university or higher education in various European quarters. It should be borne in mind that the data for the experimental part of the research were collected between 2007 and 2009. In the years that went by since then, the CLIL innovation has been developing in several respects.

As discussed in the introduction, CLIL is slowly being introduced at more educational levels, available for a broader group of learners - so a little less elitist. Admittance to a CLIL class seems no longer the exclusive prerogative of a small group of the best performing pupils with high motivation for a study at university. The CLIL approach is carefully moving into the direction of a trajectory for equipping young people with skills for the 21<sup>st</sup> century as described by Marsh (2013). Recent research has shown that CLIL can have a positive influence on completely unselected learners of a junior vocational level (Denman et al., 2018), even though still very rare in European CLIL. In addition, CLIL learning in primary education has gained ground. An increasing number of secondary CLIL learners is not new to the approach, and in all probability have more advanced EFL skills at the start of secondary school when compared to the population in our research. In the second place it is likely that, in education in general but especially in the initial stages of this innovative CLIL programme, a lot has happened in ten years' time. Teacher training courses have become adjusted to newly perceived needs, teachers have become more and more familiar with the new demands, new teaching materials have been introduced and didactics keep on developing. The mainstream departments, too, probably have invested time and energy to keep up with present demands. They may also have introduced enriched EFL learning in the course

of time. For the interpretation of the outcomes of the present thesis this implies, that the research conditions are emphatically placed in the first decade of the 21<sup>st</sup> century, when the CLIL innovation was in its early stages. As Denman et al. (2018) point out, the effects of CLIL can be very different as a result of context and interaction effects. Therefore it seems likely that the results discussed in the present thesis may not be fully generalizable to the present times of 2019.

## **6. 2 Effects of Content and Language Integrated Learning on proficiency in English**

In this thesis a major innovation in European education was investigated, more specifically as to its effects on English language education. The study has provided an insight into the question whether the combination of language and content lessons as practised in English-medium CLIL leads to better EFL skills across divergent contexts and educational levels than mainstream teaching with language lessons only. We found considerable differences as to teacher skills, the CLIL curriculum, learner preparation and the presence of the English language in society at large. In our research across three countries as discussed in Chapter 2 we compared CLIL and mainstream learner EFL results. CLIL learners showed better scores than mainstream learners but not in each of the three countries, not for each tested language skill and not to the same degree. The best overall performance for both CLIL and control groups was found in the Netherlands. As to development over time, CLIL as well as control class showed diversities which were difficult to explain. The language skill in which CLIL classes on the whole developed best in comparison with the control classes was knowledge of idioms. The Italian CLIL pupils showed the best progress for text comprehension, though they did not set out with the highest scores for this skill: the German CLIL pupils did, but their performance declined and they ended with the lowest text comprehension scores. Negative outcomes were found for the German schools. In the Berlin area, Zydatiß (2012) found conditions under which CLIL learners at grammar schools turned their back on CLIL learning for strategic reasons, i.e. better grades when participating in non-CLIL classes. In German CLIL as well as control classes receptive grammar also developed negatively.

The ‘success of CLIL’ did not seem to be proportional to CLIL content lessons: several gain scores were highest in Italy, surprising in view of their modest modular CLIL programme. Nor did it depend on teacher qualifications: the dually qualified teachers in Germany did not produce better results than their Dutch and Italian colleagues with single qualifications. Our evaluation of longitudinal research throughout Europe of the past twenty years into the effects of CLIL on EFL proficiency with a pre-test / post-test design and the use of control groups as discussed in Chapter 3 seemed to reveal a divide between high and low EFL proficiency countries. In the first, the overall picture of experimental research involving a variety of linguistic skills may be called



slightly disappointing: in most cases the - initially already higher - CLIL learner scores of the tested EFL skills had developed more than those of mainstream learners, but in most cases not significantly. More positive results appeared in Italy and the majority of Spanish areas, which can be seen as low EFL proficiency starters with much room for growth at the time but steadily catching up. This conclusion provides a hopeful prospect for other low-EFL proficiency areas considering CLIL implementation, which could benefit from ample Spanish practice.

We investigated the effects of a divergence of European CLIL approaches on EFL skills, with a variety of learners as to age, nationality, native language, additional home languages, school levels, prior EFL knowledge and exposure to English. What all contexts had in common is the fact that CLIL can be seen as selective: their classes are populated by the most motivated and highest achievers. Even in educational areas with a liberal approach and accessibility for all, as in Spain, the programme attracts learners with particular fields of interest and the abilities that go with them. This puts them at an advantage over the groups that are usually monitored as control groups. Critics have drawn attention to the fact that this may account for distorting factors in research findings as to the actual effects of the CLIL programme itself (Bruton, 2011) as possible differences could be ascribed to other variables. In her review of CLIL research Pérez-Cañado (2012) underscored the need expressed by researchers in the field of CLIL to secure the homogeneity of experimental and control groups when comparing CLIL and non-CLIL performance. However, it has also been acknowledged that this may be a difficult task, e.g. by Lasagabaster (2008) who discussed the fact that some variables are hard to control in a study carried out in an existing educational context, such as in the Basque schools where English is the third target language.

### **6.3 Effects of CLIL on pupil identity**

A major issue discussed in the present thesis is the question whether CLIL contributed to its aim of educating confident L2 speakers and citizens for life in today's internationalised world. Proficiency in the lingua franca can be seen as a prerequisite for this target, and studies in various European have shown positive L2 effects for receptive language skills, vocabulary and morphology as well as creativity, risk-taking, fluency and speaking confidence (Dalton-Puffer, 2008). Still, CLIL has been described as more than L2 learning and also has non-cognitive aims such as enhancing motivation, fostering positive attitudes towards other cultures and foreign language learning, as well as contributing to linguistic confidence of its learners. In the present thesis we have tackled these issues and investigated if and to what extent CLIL enhances pupils' confidence in their EFL skills and involves them in the international world.

In Chapter 2 we discussed that, as could be expected in view of CLIL selection criteria, CLIL pupils in the Netherlands, Germany and Italy started out with better EFL skills than their mainstream peers and even though they also progressed more, their score growth was not always significant. In Chapter 4 we discussed that the same group of learners had more interest in international cultures, and also in studying English and other foreign languages. Moreover, they had greater confidence in their target language skills, findings that had been confirmed by other research (Dale & Tanner, 2012; Doiz et al., 2014; Pérez-Cañado, 2012). However, after two years' time the CLIL environment had not enhanced these feelings significantly more than the traditional mainstream EFL curriculum. All learners - in CLIL as well as in mainstream classes across the three countries - developed more international involvement and language confidence, and more or less to the same degree. Even if this does not suggest a significant contribution of CLIL, a small success may be seen in the increase of CLIL learner scores that were already very high at the outset, both for international orientation and EFL confidence. Each of the CLIL programmes, irrespective of its intensity or teaching practice, succeeded in maintaining learner involvement and may as such be seen as a positive innovation.

#### **6.4 Variation in CLIL effects on the individual learner**

The present thesis analysed if the specific learners' talent for and positive attitudes towards EFL learning which were generally required for admittance to a CLIL class are related to EFL proficiency results in divergent contexts. Our results suggested an affirmative answer. As discussed in Chapter 2 learners in CLIL classes started out with better EFL skills than their mainstream peers, in Chapter 4 we saw that they also had more interest in international cultures, in studying English and other foreign languages, moreover, they had greater confidence in their target language skills. In Chapter 5 we found that an aptitude to learn the English language and notably the confidence with which pupils use it, EFL confidence, had significant effects on EFL proficiency scores while the presence of English in society at large was found to be a side contribution. Aptitude and confidence proved to be of more weight than the degree to which learners feel attracted to the international world.

As one of the central individual differences in language learning we investigated language aptitude: the specific talent to learn languages, believed to be very much present in CLIL classes. CLIL generally attracts not only academically talented but also above average L2 learners (Wolff, 2007). Aptitude scores have been found to be predictable from social class, vocabulary development and parental education (Skehan, 1989, p.33), learner characteristics that often show similar high levels in CLIL classes, as a side effect of selection. It is therefore not surprising that the EFL aptitude scores we discussed in Chapter 5 did not vary greatly across the CLIL contexts we investigated in

four countries: the Netherlands, Germany, Italy and Hungary. The role of this variable appeared to have mainly initial effects on EFL scores. Aptitude for language learning did not convincingly lead to an increase in scores, even though in the intensive EFL learning context of Hungarian CLIL it appeared to be supportive. A more meaningful finding was its relation with the learner variable language confidence, the confidence with which learners use the English language, a feature that turned out to contribute considerably to initial skills as well as to progress in language learning. Learner aptitude and confidence together seemed to form an interconnected construct promotive to making the most of the CLIL environment, and moreover an incentive to maximize profit from the presence of the English language in their daily life.

Even when CLIL is selective and relies on a talent for language learning, our findings suggest that learners with confidence in their own EFL skills profit the most from the CLIL environment: the CLIL classroom considerably enhances the effect of such confidence.

### **6. 5 The positive contribution of CLIL to the learning environment**

In our investigation of schools in divergent contexts the benefits of CLIL on a wider scale became apparent. Schools of all levels are faced with the challenge of how to keep up with the demands of 21<sup>st</sup> century education, in which CLIL is expected to play a prominent role (Marsh, 2013). Moreover, in the present day and age it becomes increasingly important for schools to distinguish themselves. The schools in our research project had all developed the know-how to answer fully to the demands necessary for CLIL which is associated with an atmosphere of looking beyond borders, new ways of teaching and attracting capable and motivated learners to its CLIL classes. For teachers the introduction of CLIL entailed the challenge of teaching their subject in English (Dale & Tanner, 2012) involving improved language skills, the mastering of new didactics, renewed interaction with colleagues, motivated students, possibilities to be involved in international school outings and meeting new colleagues. In relation to CLIL practice De Graaff et al. (2007) argued that not only CLIL content teachers can profit from effective pedagogical approaches to teaching and learning through an additional language, but foreign language teachers too, in that they can give their lesson a wider scope by stimulating content-based language learning activities in foreign language curricula.

For CLIL learners one of the purported benefits of CLIL introduction into European mainstream education was to provide more opportunities to practise real, appropriate language to communicate with others in the classroom and learn by interaction. After investigation of diverse educational contexts in various European quarters we can safely say that this aim has been realised. Content and language integration, focussing on authentic materials and cooperation, turned out to be

promotive to language development as well as to a sense of confidence in using the foreign language. Learners in CLIL classes, even in its most modest CLIL implementation as in Italy and also Germany, get hours more practice of spoken skills by interaction per week than would normally be the case and on the whole take part actively in classroom communication. Their enhanced L2 skills also facilitated out-of-school involvement in the language, such as practising speaking English at home or with friends, reading English books and sites on the internet and watching English programmes on TV.

At the same time, a wider goal was addressed. In 2006 the European Parliament recommended eight key competences for lifelong learning. These competences, which combine knowledge and skills appropriate for life in the 21<sup>st</sup> century, include communication and foreign languages; learning to learn; social and civic competences; sense of initiative and entrepreneurship; cultural awareness and expression. CLIL has been seen to be a key lever in realizing some of the eight key competences (British Council, 2014). CLIL learners across contexts in the present study had the possibility to participate in a number of out-of-school activities such as exchange projects, language excursions, theatre visits, public speaking contests and participating in the European Youth Parliament. As our study made clear, the CLIL approach has introduced its learners to a wider cultural context, prepared for internationalisation, improved language competence and confidence, developed multilingual interests and attitudes and increased learner EFL proficiency.

## 6.6 Limitations and future directions

We found CLIL effects across countries, however, it should be acknowledged that the number of participants was relatively small and our findings can only be generalised with great caution. Apart from that, as mentioned before, the data were collected when CLIL was still new. As could be expected in an innovative programme, developments have taken place, whether scheduled or through common experience throughout the years. The outcomes will most likely not be the same if the research were to be repeated.

As a downside to the divergence in national educational systems and the start of CLIL the students could not entirely be matched for age: in the Netherlands secondary school starts at twelve and coincides with the introduction of CLIL, in Germany secondary school starts at ten but CLIL two years later, while in Italy the middle school trajectory has to be completed first before embarking on CLIL at age fourteen and in Hungary the entire first secondary school year is spent on studying the target language at age fifteen.

Interestingly, we found CLIL effects on EFL performance at selective grammar schools to be dependent on language confidence and aptitude of learners. As CLIL is

increasingly being introduced into a wider array of schools, taking learner diversity into account and finding the individual learner variables that contribute to successful learning should become a more intensive area of research (Cenoz et al., 2013; Dalton-Puffer et al., 2014). In this respect also teacher skills and didactics need a prominent place on the research agenda (Coyle, 2007). Teacher training programmes need increasing focus on CLIL didactics, in which it is necessary to change established habits which are used in the L1 into teaching the same content in L2 (Papaja, 2013). Spratt (2017) discusses the importance and centrality of teacher language in CLIL, demonstrating the role of classroom interaction in the L2 for the comprehension and expression of higher level thinking skills such as evaluating, comparing, hypothesising. A related area is studying the effect of affectivity and teacher beliefs. A positive attitude, motivation and commitment of both learners and teachers to learn through an L2 have been found promotive to cognitive gains (Otwińska, 2013).

To answer the question to what extent the CLIL target of preparing students adequately for future studies is reached we need to consider how English-taught programmes in higher education have developed in Europe and the world at large. Professional development of lecturers in international classrooms in higher education, involving students with diverse nationalities and from different cultures, has been reported as an area in need of more systematic training programmes (Lauridsen, 2017). In Norway Hellekjær (2006) reported that most Norwegian students in upper-secondary education are highly proficient in basic interpersonal communicative skills, but that their cognitive academic language proficiency is not sufficient to tackle academic content in English-taught programmes. Dalton-Puffer (2013) emphasises the fact that learning content in an L2 needs to be supported by classroom discourse in the specific language of the subject, so-called academic language, which is seen as distinct from interpersonal language even though their learning paths are similar and both develop through social interaction (p.226).

The contribution of CLIL to student mobility needs further investigation in the first years of higher education, to provide an insight into the extent to which students feel prepared for the demands of higher education and life in a foreign culture. The educational innovation of integrating the learning of content and a foreign language is increasingly gaining ground throughout all levels of education. In view of my research findings learner language skills play an important role. Consensus as to the linguistic and didactical needs of content teachers and students seems a key factor for the future, in which decisions have to be made as to further implementation of CLIL. The issue of approaches to the integration of language and content has only recently appeared on the agenda. As Nikula (2017) argues: 'there has been a shift in emphasis in research from studies orienting to effects of CLIL on language learning outcomes to

studies that point towards the need to adopt a truly integrated view on language and content and to explore the potential that CLIL has in supporting the development of subject literacies.’(p.1). A core characteristic of CLIL is the use of an L2 to teach subject matter, as discussed in the present thesis. However, as Van Kampen et al. (2018) have investigated, there are mixed views in the countries of Europe and beyond as to the systematic inclusion of the L2 - the CLIL target language - in the teaching of content. In response to their findings, Rumlich (2018) argues that if CLIL content teachers rely largely on incidental L2 acquisition, this may turn out to be inefficient and detrimental to subject related competence. In view of the wider implementation of dual learning this should receive attention on the CLIL research agenda.

Our overall conclusion is, that in order to arrive at a fuller understanding of the individual variation of CLIL effects, there is a need of large-scale studies with different countries and educational levels being involved, in which learning effects and pupil identity can be examined. These should be related to learner and teacher characteristics, as well as CLIL teaching practice, and follow a longitudinal multilevel design.

## References

- Admiraal, W., Westhoff, G., & De Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation* 12(1), 75-93. doi:10.1080/13803610500392160
- Agustín Llach, M. P. (2015). Age and type of instruction (CLIC vs. traditional EFL) in lexical development. *International Journal of English Studies* 16(1), 75-96. Spain: University of Murcia.
- Agustín Llach, M. P., & Alonso, A. C. (2014). Vocabulary growth in young CLIL and traditional EFL learners: Evidence from research and implications for education. *International Journal of Applied Linguistics* 2(2), 1-13. doi:10.1111/ijal.12090
- Alonso, E., Grisaleña, J., & Campo, A. (2008). Plurilingual education in secondary schools: Analysis of results. *International CLIL Research Journal* 1(1), 3. Retrieved from <http://www.icrj.eu/11/article3.html>
- Artieda, G., Roquet, H., & Nicolás-Conesa, F. (2017). The impact of age and exposure on EFL achievement in two learning contexts: formal instruction and formal instruction + content and language integrated learning (CLIL). *International Journal of Bilingual Education and Bilingualism*. doi: 10.1080/13670050.2017.1373059
- Ball, P., Kelly, K., & Clegg, J. (2015). *Oxford Handbooks for Language Teachers: Putting CLIL into practice*. U.K.: Oxford University Press.
- Basque Institute of Educational Evaluation and Research. (2007). *Trilingual students in secondary school: A new reality*. Retrieved from <http://www.isei-ivei.net>
- Bentley, K. (2010). *The teaching knowledge test (T.K.T.) course: CLIL module*. U.K.: Cambridge University Press.
- Berns, M., De Bot, K., & Hasebrink, U. (2010). *In the presence of English. Media and European youth*. New York: Springer-Verlag New York Inc.
- Breen, M., & Candlin, C. (2001). The essentials of a communicative curriculum in language teaching. In D. Hall, A. Hewings (Eds.), *Innovation in English language teaching: a reader* (pp. 9-26). New York: Routledge.
- Breidbach, S., & Viebrock, B. (2012). CLIL in Germany: Results from recent research in a contested field of education. *International CLIL Research Journal* 1(4), 1. Retrieved from <http://www.icrj.eu/14/article1.html>
- British Council. (2014). *CLIL policy and practice: Competence-based education for employability, mobility and growth*. Como, Italy. Retrieved from: <http://www.britishcouncil.org/sites/britishcouncil.uk2/files/clil>
- Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System* 39, 523-532. doi:10.1016/j.system.2011.08.002
- Bruton, A. (2015). CLIL: Detail matters in the whole picture. More than a reply to J. Hüttner and U. Smit (2014). *System* 53, 119-121. doi:10.1016/j.system.2015.07.005
- Carroll, J., & Sapon, S. (1959). *The Modern Languages Aptitude Test*. San Antonio, Texas: The Psychological Corporation.

- Cenoz, J., Genesee, F., & Gorter, D. (2013). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 1-21.
- Cimermanova, I. (2017). CLIL- a dialogue between the language and subject teachers. *Scientia et Eruditio* 1, 1-15.
- Coonan, C. M. (2012). The foreign language curriculum and CLIL. *Synergies Italie* 8, 117-128. Retrieved from [https://gerflint.fr/Base/Italie8/carmel\\_mary%202.pdf](https://gerflint.fr/Base/Italie8/carmel_mary%202.pdf)
- Coyle, D. (2007). Content and Language Integrated Learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism* 10(5), 543-562. doi:10.2167/beb459.0
- Coyle, D., Marsh, D., & Hood, P. (2010). *Content and Language Integrated Learning*. U.K.: Cambridge University Press.
- Dale, L., & Tanner, R. (2012). *CLIL Activities. A resource for subject and language teachers*. U.K.: Cambridge University Press.
- Dallinger, S., Jonkmann, K., Hollm, J., & Fiege, C. (2016). 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 Publishing Company.
- Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe. In W. Delanoy & L.Volkmann (Eds.), *Future Perspectives for English Language Teaching* (pp.139-157). Heidelberg: Carl Winter.
- Dalton-Puffer, C. (2009). Communicative competence and the CLIL lesson. In Y. Ruiz de Zarobe & R.M. Jiménez Catalán (Eds.), *Content and language integrated learning. Evidence from research in Europe* (pp.197-214). Bristol: Multilingual Matters.
- Dalton-Puffer, C. (2011). Content and Language Integrated Learning: From practice to principles? *Annual Review of Applied Linguistics* 31, 182-204.
- Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics* 1(2), 216-253. doi:10.1515/eujal-2013-0011
- Dalton-Puffer, C., Llinares, A., Lorenzo, F., & Nikula, T. (2014). You can stand under my umbrella. Immersion, CLIL and bilingual education. A response to Cenoz, Genesee & Gorter (2013). *Applied Linguistics* 35(2), 213-218.
- De Bot, K. (2002). CLIL in the European context. In D. Marsh (Ed.), *CLIL/EMILE. The European dimension - actions, trends and foresight potential* (pp. 29-31). Jyväskylä: UniCOM, Continuing Education Centre.
- De Graaff, R., Koopman, G.J., Anikina, Y. & Westhoff, G. (2007). An observation tool for effective L2 pedagogy in Content and Language Integrated Learning (CLIL). *The International Journal of Bilingual Education and Bilingualism* 10 (5). doi:10.2167/beb462.0



- De Graaff, R., & Van Wilgenburg, O. (2015). The Netherlands: Quality control as a driving force in bilingual education. In P. Mehisto & F. Genesee (Eds.), *Building bilingual education systems. Forces, mechanisms and counterweights* (pp 167- 178). U.K.: Cambridge University Press.
- Denman, J., Van Schooten, E., & De Graaff, R. (2018). Attitudinal factors and the intention to learn English in pre-vocational secondary bilingual and mainstream education. *Dutch Journal of Applied Linguistics* 7(2), 203 - 226. doi: 10.1075/dujal.18005.den
- Di Martino, E., & Di Sabato, B. (2012). CLIL implementation in Italian schools: Can long-serving teachers be retrained effectively? The Italian protagonists' voice. *Latin American Journal of Content and Language Integrated Learning*, 5(2), 73-105. doi: 10.5294/laclil.2012.5.2.9 ISSN 2011-6721
- Doiz, A., Lasagabaster, D., & Sierra, J.M. (2014). CLIL and motivation: the effect of individual and contextual variables. *The Language Learning Journal* 42(2), 209-224 doi:10.1080/09571736.2014.889508
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp 589-630). U.K.: Blackwell Publishing Ltd. doi:10.1002/9780470756492.ch18
- Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: a Hungarian perspective*. Clevedon: Multilingual Matters.
- Dörnyei, Z., & Hadfield, J. (2013). *Motivating Learning. Research and Resources in Language Teaching*. Harlow: Pearson Education Limited.
- Ellis, R. (1985). *Understanding second language acquisition*. U.K.: Oxford University Press.
- European Commission. (2006). *Europeans and their languages. Special Eurobarometer 243*. Brussels: European Commission.
- European Commission. (2012). *Europeans and their languages. Special Eurobarometer 386*. Brussels: European Commission.
- European Platform for Education in the Netherlands. (2013). *Content and Language Integrated Learning*. Retrieved from <http://www.europeesplatform.nl/sf.mcgi?2682>
- Eurydice Report. (2006). *Content and Language Integrated Learning (CLIL) at school in Europe*. Brussels: European Commission.
- Eurydice. (2012). *Key data on teaching languages at school in Europe*. Retrieved from [http://eacea.ec.europa.eu/education/eurydice/index\\_en.php](http://eacea.ec.europa.eu/education/eurydice/index_en.php)
- Farkas, J., & Kniesz, V. (2002). Hungarian. In M. Görlach (Ed.), *English in Europe*, (pp. 277 – 290). U.K.: Oxford University Press.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Gardner, R. C. (2010). *Motivation and second language acquisition: The socio-educational model*. New York: Peter Lang.

- Gardner, R. C., & W.E. Lambert. (1972). *Attitudes and motivation in second language learning*. Massachusetts: Rowley, Newbury House.
- Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015). Development of EFL writing over three years in secondary education: CLIL and non-CLIL settings. *Language Learning Journal* 43(3), 1-18. doi:10.1080/09571736.2015.1053278
- Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2016). A methodology for longitudinal research on EFL written production: Capturing writing multidimensionality by combining qualitative and quantitative procedures. *Journal of Research Design and Linguistics and Communication Science* 3(1), 23-48. doi:10.1558/jrds.30221
- Gierlinger, E.M., & T.A. Wagner. (2016). The more the merrier - revisiting CLIL-based vocabulary growth in secondary education. *LACLIL* 9(1), 37-63. doi:10.5294/laclil.2016.9.1.3
- Goris, J., Denessen, E., & Verhoeven, L. (2013). Effects of the Content and Language Integrated Learning approach to EFL teaching. *Written Language & Literacy* 16(2), 186-207.
- Goris, J., Denessen, E., & Verhoeven, L. (2017). The contribution of CLIL to learners' international orientation and EFL confidence. *The Language Learning Journal*. doi:10.1080/09571736.2016.1275034
- Graz Group, Coyle, D., Halbach, A., Meyer, O., Schuck, K., & Ting, T. (2013). Developing literacy through CLIL. *The CLIL Magazine*. European Centre for Modern Languages.
- Grigorenko, E.L., Sternberg, R.J., & Ehrman, M.J. (2000). A theory-based approach to the measurement of foreign language ability: the Canal-F Theory and Test. *The Modern Journal* 3(84), 390-405.
- Hellekjær, G.O. (2006). Screening criteria for English-medium programmes: A case study. In R. Wilkinson, V. Zegers & C. van Leeuwen (Eds.), *Bridging the assessment gap in English-medium higher education* (pp.43-60). Maastricht: Maastricht University.
- Hindmarsh, R. (1982). *Cambridge English lexicon*. U.K.: Cambridge University Press.
- Huibregtse, I. (2001). *Effecten en didactiek van tweetalig voortgezet onderwijs in Nederland [Effects and didactics of bilingual secondary education in the Netherlands]* (Doctoral Dissertation). University of Utrecht, The Netherlands.
- Hummel, K. M. (2013). Target language community involvement: Second language linguistic self-confidence and other perceived benefits. *The Canadian Modern Language Review* 69(1), 65-90. Toronto: University of Toronto Press.
- Ipek, H. (2009). Comparing and contrasting first and second language acquisition: Implications for language teachers. *English Language Teaching* 2(2).
- Kovács, J., & Trentinné Benkő, É. (2014). *The World at Their Feet Children's Early Competence in Two Languages through Education*. Budapest: Eötvös József Könyvkiadó.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. U.K., Oxford: Pergamon.

- Krashen, S.D. (2002). The comprehension hypothesis and its rivals. In: *Selected papers from the eleventh international symposium on English teaching / Fourth Pan-Asian conference* (pp. 395-404). English Teachers Association/ROC.
- Küppers, A., & Trautmann, M. (2013). It is not CLIL that is a success – CLIL students are! Some critical remarks on the current CLIL boom. In S. Breidbach & B. Viebrock(Eds.), *Content and Language Integrated Learning (CLIL) in Europe. Research perspectives on policy and practice* (pp. 285-296). Frankfurt am Main: Peter Lang.
- Langé, G. (2014, April). *CLIL/Emile in Italy: a real driver for lifelong learning*. Paper presented at the ALTE 5<sup>th</sup> International Conference: Language Assessment for Multilingualism, Paris. Abstract retrieved from <http://events.cambridgeenglish.org/alte-2014/docs/presentations/alte2014-gisella-lange.pdf>
- Lasagabaster, D. (2008). Foreign language competence in Content and Language Integrated courses. *The Open Applied Linguistics Journal* 1, 31-42.
- Lasagabaster, D., & Sierra, J. M. (2010). Immersion and CLIL in English: more differences than similarities. *ELT Journal* 64(4), 367-75. doi:10.1093/elt/ccp082
- Lasagabaster, D., & Doiz, A. (2015). A longitudinal study on the impact of CLIL on affective factors. *Applied Linguistics* 38(5): 688–712. doi:10.1093/applin/amv059
- Lasagabaster, D., & Ruiz de Zarobe, Y. (2010). Ways forward in CLIL: Provision issues and future planning. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. 278-295). U.K., Newcastle uponTyne: Cambridge Scholars Publishing.
- Lauridsen, K. (2007). Professional development of international classroom lecturers. In J.Valcke & R. Wilkinson (Eds.) *Integrating Content and Language in Higher Education. Perspectives on professional practice* (pp.25-37). Frankfurt am Main:Peter Lang.
- Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of content and language integrated learning in European education: Key findings from the Andalusian sections evaluation project. *Applied Linguistics* 31(3), 418-442. doi:10.1093/applin/amp041
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K.A. (1998). Conceptualizing willingness to communicate in a L2. A situational model of L2 confidence and affiliation. *Modern Language Journal* 82, 545-562. doi:10.1111/j.1540-4781.1998.tb05543.x
- MacIntyre, P. D., Baker, S.C., Clément, R., & Conrod, S. (2001). Willingness to communicate, social support, and language orientations of immersion students. *Studies in Second Language Acquisition* 23, 369-388.
- Mannion, G., Biesta, G., Priestley, M., & Ross, H. (2011). The global dimension in education and education for global citizenship: Genealogy and critique. *Globalisation, Societies and Education* 9 (3-4), 443-456. doi:10.1080/14767724.2011.605327
- Marsh, D. (2002). (Ed.) *CLIL/EMILE. The European Dimension - Actions, Trends and Foresight Potential*. Jyväskylä: UNICOM, Continuing Education Centre.
- Marsh, D. (2013). *The CLIL Trajectory: Educational Innovation for the 21<sup>st</sup> Century iGeneration*. Córdoba: Academic Press.

- Marsh, D., Maljers, A., & Hartiala, A.K. (2001). *Profiling European CLIL Classrooms -Languages Open Doors*. Finland: University of Jyväskylä and The Netherlands: European Platform for Dutch Education.
- Meara, P. M. (1992). *EFL Vocabulary Test*. U.K., Swansea: Centre for Applied Language Studies.
- Mearns, T., & De Graaff, R. (2018). Bilingual education and CLIL in the Netherlands. The paradigm and the pedagogy. *Dutch Journal of Applied Linguistics* 7(2), 122–128. doi:10.1075/dujal.00002.int
- Mehisto, P. (2007). What a school needs to consider before launching a CLIL programme: The Estonian experience. In D. Marsh & D. Wolff (Eds.), *Diverse Contexts –Converging Goals: CLIL in Europe* (pp. 61-77). Frankfurt: Peter Lang.
- Merino, J.A., & Lasagabaster, D. (2017). The effect of content and language integrated learning programmes' intensity on English proficiency: A longitudinal study. *International Journal of Applied Linguistics* 1(13). doi:10.1111/ijal.12177
- Nation, P. (1986). *Word Lists*. Wellington: Victoria University English Language Centre.
- Navarro Pablo, M., & García Jiménez, E. (2018). Are CLIL students more motivated? An analysis of affective factors and their relation to language attainment. *PortaLinguarum* 29.
- Nieto Moreno de Diezmas, E. (2016). The impact of CLIL on the acquisition of L2 competences and skills in primary education. *International Journal of English Studies*. doi:10.6018/ijes/2016/2/239611
- Nikula, T. (2017). CLIL: A European Approach to Bilingual Education. In N. van Deusen-Scholl & S. May (Eds.) *Second and Foreign Language Education: Encyclopedia of Language and Education* (pp. 1-14). doi:10.1007/978-3-319-02246-8\_10
- Noels, K. A., & Clément, R. (1996). Communicating across cultures: Social determinants and acculturative consequences. *Canadian Journal of Behavioural Science*, 28(3), 214-228.
- Nuffic. (2018). *TTO 2.0*. Retrieved from <https://www.nuffic.nl/nieuws/tweetalig-onderwijs-20-staat-voor-de-deur/>
- Oonk, G.H. (2004). *European integration as a source of innovation in education*. (Doctoral Dissertation). Alkmaar/Den Haag: Europees Platform.
- Otwinowska, A. (2013). CLIL lessons in the upper-primary: The interplay of affective factors and CALP. In D. Gabryś-Barker & J. Bielska (Eds.), *Affective Dimension in Second Language Acquisition* (pp. 211-225). Bristol: Multilingual Matters.
- Otwinowska, A., & Foryś, M. (2015). They learn the CLIL way, but do they like it? Affectivity and cognition in upper-primary CLIL classes. *International Journal of Bilingual Education and Bilingualism*. doi:10.1080/13670050.2015.1051944
- Papaja, K. (2013). The role of a teacher in a CLIL classroom. *Glottodidactica XL* (1) Poznań, Adam Mickiewicz University Press.
- Paran, A. (2013). Content and Language Integrated Learning: Panacea or policy borrowing myth? *Applied Linguistics Review* 4(2), 317-342. doi:10.1515/applirev-2013-0014

- Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism* 15 (3), 315-341. doi:10.1080/13670050.2011.630064
- Pérez-Cañado, M. L. (2016). Stopping the 'pendulum effect' in CLIL research: Finding the balance between Pollyanna and Scrooge. *Applied Linguistics Review*. doi:10.1515/applirev-2016-2001
- Pérez-Cañado, M.L. (2018). CLIL and educational level: A longitudinal study on the impact of CLIL on language outcomes. *Porta Linguarum* 29.
- Pérez-Cañado, M.L., & Lancaster, N.K. (2017). The effects of CLIL on oral comprehension and production: A longitudinal case study. *Language, Culture, and Curriculum* 30(3),300-316. doi:10.1080/07908318.2017.1338717
- Pérez-Vidal, C., & Roquet, H. (2015). CLIL in context: Profiling language abilities. In M.Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based Language Learning in Multilingual Educational Environments* (pp.237-255). New York: Springer-Verlag
- Pétery, D. (2011). English in Hungarian advertising. *World Englishes* 30(1), 21-40. Blackwell Publishing Ltd.
- Piesche, N., Jonkmann, K., Fiege, C., & Kebler, J. (2016). CLIL for all? A randomised controlled field experiment with sixth-grade students on the effects of content and language integrated science learning. *Learning and Instruction* 44, 108-116.
- Pladevall-Ballester, E., & Vallbona, A. (2016). CLIL in minimal input contexts: A longitudinal study of primary school learners' receptive skills. *System* 58, 37-48.
- Prüfer, K. (2013). CLIL modules in the mathematics classroom. Reasons for their implementation and first empirical results. In S. Breidbach & B. Viebrock (Eds.), *Content and Language Integrated Learning (CLIL) in Europe, Research perspectives on Policy and Practice* (pp. 251-266). Frankfurt: Peter Lang.
- Roquet, H., & Pérez-Vidal, C. (2015). Do productive skills improve in Content and Language Integrated Learning contexts? The case of writing. *Applied Linguistics* 1,24. Oxford: Oxford University Press.
- Ruiz de Zarobe, Y. (2008). CLIL and foreign language learning: A longitudinal study in the Basque country. *International CLIL Research Journal* 1(1), 5. Retrieved from <http://www.icrj.eu/11/article5.html>
- Ruiz de Zarobe, Y. (2011). Which language competencies benefit from CLIL? An insight into applied linguistics research. In Y. Ruiz de Zarobe, J.M. Sierra & F. Gallardo delPuerto (Eds.), *Content and foreign language integrated learning. Contributions to multilingualism in European contexts* (pp. 129-153). Bern: Peter Lang.
- Ruiz-Garrido, M. F., Palmer-Silveira, J. C., & Fortanet-Gómez, I. (Eds.). (2010). *English for Professional and Academic Purposes*. Amsterdam - New York: Rodopi.
- Rumlich, D. (2013). Students' general English proficiency prior to CLIL: Empirical evidence for substantial differences between prospective CLIL and non-CLIL students in Germany. In S. Breidbach & B. Viebrock (Eds.), *Content and Language Integrated Learning (CLIL) in Europe. Research perspectives on Policy and Practice*,(pp.181- 201). Frankfurt am Main: Peter Lang.

- Rumlich, D. (2017). CLIL theory and empirical reality: Two sides of the same coin? *Journal of Immersion and Content-Based Language Education* 5(1), 110-134. John Benjamins Publishing Company.
- Rumlich, D. (2018). Current research on CLIL and bilingual education in the Netherlands: A discussion. *Dutch Journal of Applied Linguistics* 7(2), 264-273. doi: 10.1075/dujal.00003.rum
- San Isidro, X., & Lasagabaster, D. (2018). The impact of CLIL on pluriliteracy development and content learning in a rural multilingual setting: A longitudinal study. *Language Teaching Research* 1–19. doi:10.1177/1362168817754103
- Seikkula-Leino, J. (2007). CLIL Learning: Achievement levels and affective factors. *Language and Education* 21(4), 328-341. doi:10.2167/le635.0
- Serragiotto, G. (2007). Assessment and evaluation in CLIL. In D. Marsh & D. Wolff (Eds.), *Diverse Contexts – Converging Goals: CLIL in Europe* (pp.271-283). Frankfurt am Main: Peter Lang.
- Shehadeh, A. (2005). Task-based language learning and teaching: Theories and applications. In C. Edwards & J. Willis (Eds.), *Teachers Exploring Tasks in English Language Teaching* (pp. 13-30). London: Palgrave Macmillan.
- Skehan, P. (1989). *Individual Differences in Second-Language Learning*. London: Routledge.
- Spratt, M. (2017). CLIL teachers and their language. *Research Papers in Language Teaching and Learning* 8 (1), 44-61. Greece: Hellenic Open University.
- Sylvén, L. K. (2010). *Teaching in English or English Teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*. (Doctoral Thesis). Gothenburg: University of Gothenburg.
- Van der Leij, A., Bekebrede, J., & Kotterink, M. (2010). Acquiring reading and vocabulary in Dutch and English: The effect of concurrent instruction. *Reading & Writing* 23(3-4), 415-434. doi:10.1007/s11145-009-9207-5
- Van Ek, J.A. (1976). *The Threshold Level for Modern Language Learning in Schools*. Groningen: Wolters-Noordhoff-Longman.
- Van Kampen, E., Mearns, T., Meirink, J., Admiraal, W., & Berry, A. (2018). How do we measure up? A review of Dutch CLIL subject pedagogies against an international backdrop. *Dutch Journal of Applied Linguistics* 7(2), 129-155. doi: 10.1075/dujal.18004.kam
- Verspoor, M., De Bot, K., & Xu, X. (2015). The effects of English bilingual education in the Netherlands. *Journal of Immersion and Content-Based Language Education* 3(1), 4-27. doi:10.1075/jicb.3.1.01ver
- Wächter, B., & Maiworm, F. (2008). *English-Taught Programmes in European Higher Education*. Bonn: Lemmens Medien GmbH.
- Wächter, B., & Maiworm, F. (2014). *English-Taught Programmes in European Higher Education. The State of Play in 2014*. Bonn: Lemmens Medien GmbH.
- Wannagat, U. (2007). Learning through L2-Content and Language Integrated Learning (CLIL) and English as Medium of Instruction (EMI). *The International Journal of Bilingual Education and Bilingualism* 10 (5), 663-682. doi:10.2167/beb465.0

- Warschauer, M. (2000). The changing global economy and the future of English teaching. *TESOL Quarterly* 34, 511–535.
- Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*. doi:10.1017/S0261444816000276
- Weenink, D. (2005). *Upper middle-class resources of power in the education arena. Dutch elite schools in an age of globalization*. Amsterdam: University of Amsterdam.
- Westhoff, G. J. (1994). *Tweetalig Onderwijs in de praktijk. Verslag van een studiereis. [Bilingual Education in practice. Report of a study tour]*. Utrecht: WCC.
- Wilkins, D. A. (1976). *Notional Syllabuses*. Oxford, Oxford University Press.
- Wolff, D. (1997). Content-based bilingual education or using foreign languages as working languages in the classroom. In D. Marsh, B. Marsland & T. Nikula (Eds.), *Aspects of Implementing Plurilingual Education* (pp.51-64). Jyväskylä: University of Jyväskylä.
- Wolff, D. (2002). On the importance of CLIL in the context of the debate on plurilingual education in the European Union. In D. Marsh (Ed.), *CLIL/EMILE. The European Dimension - Actions, Trends and Foresight Potential* (pp.47-48). Jyväskylä: UNICOM, Continuing Education Centre.
- Wolff, D. (2007). CLIL: Bridging the gap between school and working life. In D. Marsh & D. Wolff (Eds.), *Diverse Contexts – Converging Goals: CLIL in Europe* (pp.15-25). Frankfurt am Main: Peter Lang.
- Zydatib, W. (2012). Linguistic thresholds in the CLIL classroom? The threshold hypothesis revisited. *International CLIL Research Journal* 1(4), 2. Retrieved from <http://www.icrj.eu/14/article2.html>

## Samenvatting

Dit proefschrift is geschreven met als doel inzicht te verschaffen in de opbrengsten van *Content and Language Integrated Learning* (CLIL) op het gebied van het leren van Engels als vreemde taal in verschillende Europese landen. CLIL is een onderwijskundige vernieuwing die haar intrede heeft gedaan in de jaren negentig van de vorige eeuw. De aanleiding voor deze vernieuwing was het feit dat er ontevredenheid bestond bij docenten, studenten, ouders en andere maatschappelijk betrokkenen over de tekorten van het gangbare vreemdetalenonderwijs met betrekking tot spreekvaardigheid en communicatieve vaardigheden. Het doel van lessen in vreemde talen zou te zeer zijn gericht op kennis van grammatica en het vertalen van teksten, doelen die te weinig aansloten bij internationale ontwikkelingen in Europa en de rest van de wereld en de eisen aan taalvaardigheid die daarmee gepaard gingen. De innovatieve benadering van het leren van een vreemde taal door middel van CLIL - naast de gebruikelijke lessen Engels door de taaldocent - had tot doel de leerling in staat te stellen om de taal samen met de inhoud van andere vakken te leren, waarbij het intensieve gebruik van de vreemde taal een belangrijke rol zou spelen bij het ontwikkelen van betere spreekvaardigheid en communicatieve vaardigheden dan voorheen het geval was. Deze vaardigheden werden als onontbeerlijk gezien voor de toekomstige generatie, die moest worden voorbereid op toegenomen internationalisatie van de economie en het studieaanbod in het hoger onderwijs.

### CLIL op scholen in verschillende Europese landen

Globaal gezien zijn initiatieven met CLIL gestart in de jaren negentig van de vorige eeuw. Veelal gebeurde dit *bottom-up*, dat wil zeggen vanuit de praktijk van het onderwijs, niet *top-down*, van bovenaf opgelegd door onderwijskundige autoriteiten. CLIL is doorgaans selectief: er gelden toelatingscriteria die gerelateerd zijn aan academische vaardigheden, aanleg om taal te leren en motivatie voor onderwijs door middel van Engels. De eerste CLIL implementaties vonden plaats in de eerste leerjaren van het voorbereidend wetenschappelijk onderwijs, in Europese landen het nationale equivalent van het gymnasium. Tussen landen bestonden aanvankelijk verschillen in de taal die als doeltaal voor CLIL fungeerde - in Duitsland was dit bijvoorbeeld de Franse taal - maar door de jaren heen werd steeds meer de Engelse taal als instructietaal voor CLIL ingevoerd. Kennis van de lingua franca werd als essentieel gezien voor leven en werken in Europa en de rest van de wereld.

Het proefschrift beschrijft een onderzoek naar de leeropbrengsten van CLIL onderwijs en attitudes van leerlingen ten aanzien van de Engelse taal in de eerste twee leerjaren van scholen voor voorbereidend wetenschappelijk onderwijs in vier Europese landen: Nederland, Duitsland, Italië en Hongarije. Om een zo compleet mogelijk beeld



te geven van de effecten van CLIL op Engelse taalvaardigheid is daarnaast een review uitgevoerd van vergelijkbaar onderzoek dat in de afgelopen twintig jaar door heel Europa is uitgevoerd op scholen voor primair en secundair onderwijs.

De bevindingen van het onderzoek zijn dat er tussen landen grote verschillen zijn wat betreft de aandacht voor de Engelse taal in het onderwijs en de mate waarin Engels is ingeburgerd in het dagelijkse leven. Dit bleek van invloed te zijn op de implementatie van CLIL. In **Nederland** wordt de Engelse taal op alle scholen onderwezen, is Engels aanwezig in de media en wordt het, mede door de gelijkenissen met Nederlands, door bijna de gehele bevolking in enigerlei mate gesproken. De CLIL werkwijze is omvangrijk: op middelbare scholen die CLIL aanbieden wordt vanaf de eerste dag - naast de gebruikelijke lessen Engels - 50 á 60 % van het lesrooster onderwezen door middel van Engels. De vakken worden gegeven door docenten met een lesbevoegdheid voor hun vak, maar niet voor de Engelse taal. Ze hebben bijscholing in de Engelse taal gevolgd om deze als instructietaal te kunnen gebruiken.

In **Duitsland** is de Engelse taal minder populair en kennis ervan wordt niet als vanzelfsprekend gezien. Engels is niet sterk aanwezig in de media, maar daarentegen wel in een vroeg stadium op het lesrooster van de basisschool. Leerlingen die naar een CLIL klas willen krijgen extra lessen. De CLIL vakdocenten hebben een dubbele lesbevoegdheid: voor hun vak en tevens voor de Engelse taal. In **Italië** spreekt het overgrote deel van de bevolking helemaal geen Engels en in de media is de taal nauwelijks te horen. Wel zijn Italianen enthousiast over de onderwijsvernieuwing die CLIL heeft gebracht en op grote schaal is inmiddels begonnen met de introductie van CLIL. De CLIL scholen in dit onderzoek werkten met modules: een deel van een vak werd een aantal maanden per schooljaar door middel van Engels onderwezen. Omdat Engelse taalvaardigheid van leraren beperkt was, hadden de vakdocenten te weinig talenkennis en werden in de les geassisteerd door de docent Engels, een benadering die bekend staat onder de naam *team-teaching*. **Hongarije** heeft een afwijkende geschiedenis. Hier was ten tijde van het communisme het leren van Russisch verplicht. Andere vreemde talen waren verboden op het lesrooster. Toen dit veranderde, wilden de jongeren massaal andere talen leren, met name Duits en Engels. Er ontstond meteen een groot lerarentekort. Om de introductie van CLIL mogelijk te maken zijn veel *native speaker teachers* uit Engeland en Amerika aangetrokken. Op het CLIL gymnasium werd een talenjaar ingevoerd, waarin de doeltaal eerst grondig werd bestudeerd alvorens in het tweede jaar aan vakonderwijs door middel van Engels - of ook wel Duits - te beginnen.

### **De effecten van CLIL op Engelse taalvaardigheid**

De effecten van CLIL op Engelse taalvaardigheid zijn onderzocht voor de onderdelen woordenschat, grammatica, idioom en tekstbegrip gemeten met een pre-

test en een post-test: de eerste Engelse test aan het begin van CLIL en een tweede, meer gevorderde test na twee jaar. Er is gebruik gemaakt van controlegroepen die niet het CLIL programma volgden, leerlingen uit klassen die alleen de gebruikelijke Engelse lessen kregen. Voor dit deel van het onderzoek was Hongarije uitgesloten, omdat er op de onderzoeksscholen geen geschikte controlegroepen aanwezig waren. De bevindingen laten zien dat leerlingen in CLIL klassen hogere scores halen, maar niet in elk van de drie landen, niet voor elk onderdeel en niet in dezelfde mate. Significante vooruitgang was te vinden in Nederland, waar de leerlingen het beste presteerden op de onderdelen woordenschat en idioom en ook de grootste vooruitgang vertoonden na twee jaar. Meer significante effecten waren te vinden in Italië. De Italiaanse CLIL leerlingen gingen het best vooruit op het onderdeel tekstbegrip, hoewel de beginscore niet de hoogste was: die was te vinden in Duitsland, maar daar liepen de prestaties terug, voor tekstbegrip en ook voor grammatica. CLIL klassen lieten in vergelijking met de controlegroepen over het algemeen de beste vooruitgang zien met betrekking tot kennis van idioom. Er scheen geen verband te zijn tussen het aantal vakken dat in CLIL klassen door middel van Engels was onderwezen, noch tussen het Engelse taal niveau van de vakdocenten en de ontwikkeling van de leerlingenresultaten voor het vak Engels.

Om na te gaan hoe de conclusies waren van andere onderzoeken naar de effecten van CLIL op scores voor Engels zijn publicaties in de afgelopen twintig jaar van andere onderzoekers geanalyseerd. Het werd duidelijk dat er lang niet in alle Europese landen - die bijna allemaal CLIL hebben geïntroduceerd - onderzoek is gedaan die de vooruitgang van CLIL leerlingen meten en vergelijken met die van leerlingen in een controlegroep. Er is veel onderzoek gedaan in Nederland en Duitsland, een enkele studie in Zweden en Oostenrijk, maar het grootste deel bleek te zijn gedaan in allerlei delen van Spanje, waar ook de meeste significante resultaten voor CLIL zijn geconstateerd. Hier is CLIL ingevoerd met steun van educatieve autoriteiten, en op veel plaatsen vanaf het eerste leerjaar op de basisschool.

### **De effecten van CLIL op attitudes van leerlingen**

Naast de twee Engelse testen zijn op ongeveer dezelfde testmomenten de houding en gerichtheid van leerlingen ten opzichte van het leren van de Engelse taal gemeten, alsmede hun aanleg daarvoor. Daarvoor zijn vragenlijsten gebruikt - ook volgens het pre-test / post-test principe. De vragen hadden betrekking op a) de mate waarin de leerlingen 'internationaal georiënteerd' zijn, d.w.z. belangstelling hebben voor andere culturen, graag vreemde talen leren, zich identificeren met sprekers van de Engelse taal en die voor internationale doeleinden willen leren; b) de mate waarin ze vertrouwen hadden in hun eigen taalvaardigheden en c) in welke mate ze aanleg hadden voor het leren van Engels. Daarnaast is gekeken naar de invloed van de Engelse

taal zoals aanwezig in de media in hun land, een factor die het leren van die taal zou vergemakkelijken, een aanwezigheid die verschilde in de landen.

De resultaten gaven aan dat de CLIL leerlingen in de drie landen met controlegroepen niet alleen hogere beginscores voor Engels hadden dan de leerlingen in de controlegroepen, maar ook bovengemiddeld vertrouwen in hun taalvaardigheid - de Nederlandse en ook de Duitse leerlingen meer dan de Italiaanse. Ook hadden CLIL leerlingen meer belangstelling voor de internationale wereld, het meest in Duitsland. Wat de toename van de gemeten constructen betreft was er geen significant effect van CLIL onderwijs aanwezig: CLIL leerlingen ontwikkelden in positieve zin, maar de controlegroepen ook en ongeveer in dezelfde mate.

Toelating tot een CLIL klas is aan voorwaarden gebonden, waarvan aanleg om Engels te leren er één is. Deze aanleg en vertrouwen in de eigen taalvaardigheid bleken significante invloed te hebben op de scores voor het vak Engels van beginnende leerlingen in de drie eerder besproken landen en ook in Hongarije. Deze twee factoren, maar vooral de factor vertrouwen in eigen taalvaardigheid, bleken een grotere invloed te hebben op de leerresultaten dan de aanwezigheid van de Engelse taal in de media of de internationale georiënteerdheid van de leerlingen. De conclusie kan worden getrokken dat met name leerlingen die vertrouwen hebben in hun taalvaardigheid het meeste baat hebben bij leren in een CLIL klas: hun zelfvertrouwen was na twee jaar nog sterker geworden.

### **De toekomst van CLIL**

De CLIL innovatie leidt tot positieve leerresultaten wat betreft de Engelse taal. Ook ontwikkelen de leerlingen zelfvertrouwen en kunnen zij hun taalaanleg gebruiken. Gebaseerd op positieve bevindingen uit het onderwijs- en onderzoeksveld wordt de CLIL werkwijze op steeds meer schooltypes ingevoerd, maar is nog steeds selectief en slechts in enkele gevallen een verplicht onderdeel van een leertraject. Ontwikkelingen in de toekomst zullen gericht zijn op het toegankelijk maken van CLIL voor een grotere groep leerlingen, aangezien CLIL voor alle leerlingen kan bijdragen aan hun toekomst in de internationale wereld. Verder moet gekeken worden in hoeverre de voorbereiding van studenten op Engelstalige onderwijsprogramma's in het hoger onderwijs gefaciliteerd wordt door CLIL. Daarbij is ook de vraag van belang in hoeverre deze programma's zelf in voldoende mate aansluiten bij de leermogelijkheden en -behoeften van de internationale studenten die zich aanmelden voor Engelstalige programma's.

## Curriculum Vitae

José Goris was born in Bergen op Zoom, the Netherlands, on January 7, 1951. She attended lower secondary education in her home town and continued studying later in life, when she first completed upper secondary education at evening school. After that she trained to be a teacher of English as a Foreign Language and worked in secondary and adult education. After completing the European postgraduate programme “Educating the Gifted” she became a specialist in developing enrichment materials for learners of high ability. Inspired by a passion for the English language, she obtained a Master’s Degree in English Language and Literature at Radboud University Nijmegen in 2001. In the past few decades she travelled vastly throughout Europe and developed an interest in internationalisation in education. This led her to conduct research into Content and Language Integrated Learning in secondary pre-university education in various European countries, the results of which are reported in the present dissertation. At present she is retired but still active in the field of Content and Language Integrated education.

## Dankwoord

Zo'n 15 jaar geleden werd mijn belangstelling gewekt voor een opmerkelijke onderwijsvernieuwing: TTO, tweetalig onderwijs met voornamelijk Engels als instructietaal. Ik ben me er steeds meer in gaan verdiepen, en geleidelijk groeide het idee om zelf onderzoek te gaan doen. Ziehier het resultaat: mijn proefschrift! Hier is het kleine steentje dat ik heb kunnen bijdragen aan de wetenschap en met name aan mijn vakgebied, onderwijs in de Engelse taal.

In de eerste plaats bedank ik de Radboud Universiteit voor alle hulp die me is geboden bij het realiseren van mijn onderzoeksproject en het publiceren van mijn dissertatie. Om te beginnen denk ik daarbij aan Janet van Hell, die me - voordat ze naar de V.S. emigreerde - deskundig en met betrokkenheid heeft geholpen bij het ontwerpen van de tests en vragenlijsten en het in gang zetten van het onderzoek. In het jarenlange vervolg daarop was Ludo Verhoeven een rots in de branding. Met de overzichtelijke schema's die hij keer op keer op papier toverde kon ik maandenlang verder, zeker toen Eddie Denessen erbij kwam met zijn deskundige blik op statistiek en SPSS. Steeds weer zaten we met z'n drieën rond de tekentafel om te brainstormen over het vinden van een opzet om de interculturele cognitieve en affectieve uitkomsten van mijn onderzoek te verwerken tot ideeën voor interessante analyses en publicaties, gevolgd door het perfectioneren van de manuscripten. De koffer vol met data die ik binnenbracht bij de RU had ik zonder hun hulp nooit kunnen uitwerken tot dit veelzijdige resultaat. Ludo en Eddie, ontzettend bedankt voor jullie inzet. Ik dank uiteraard ook de manuscriptcommissie voor de bereidheid tot het lezen en beoordelen van het eindresultaat.

Een bedankje is ook zeker op zijn plaats voor de docenten Lex Hendriks en Chris Visscher. Ik heb veel plezier gehad van jullie statistieklessen, nieuwe materie voor iemand die van de Faculteit Letteren komt. En zeker niet op de laatste plaats bedank ik Marleen Hofman, die mijn zelfvertrouwen bij het presenteren van mijn onderzoek een enorme *boost* heeft gegeven met haar door een aangename sfeer gekenmerkte training.

Vervolgens wil ik mijn waardering uitspreken voor de welwillendheid van mijn internationale onderzoekspopulatie. Ik bedank alle docenten voor de welkome ontvangst op hun school, en voor hun bereidheid om mee te werken aan mijn onderzoek. Ik heb ook ontzettend genoten van de inzet waarmee klassen vol met leerlingen zich over mijn tests en vragenlijsten bogen. Jongens en meisjes, ik wil jullie allen heel hartelijk danken. Zonder jullie zou mijn onderzoek niet mogelijk zijn geweest, en zou dit proefschrift er niet zijn.

De niet-Nederlandse deelnemers hoop ik een plezier te doen met een bedankje in hun eigen taal. Ik ben niet voor niks polyglot en tenslotte hoeft niet altijd alles in het Engels:

### **Liebe** Lehrer und Lehrerinnen, liebe Schüler und Schülerinnen!

An diesem schönen Tag, da ich meine Dissertation vollendet habe, möchte ich mich nochmals herzlich für Ihre große Bereitschaft an der Mitwirkung an meiner wissenschaftlichen Arbeit bedanken. Ich habe mich in Ihren Schulen sehr willkommen gefühlt und es hat mich gefreut, dass so viele junge Leute bereit waren, an meiner Forschungsarbeit teilzunehmen. Ohne Ihre Hilfe wäre es für mich unmöglich gewesen, diese Untersuchung durchzuführen. Vielen Dank dafür!

### **Gentile** professori e professoressa, gentile scolari e scolare!

In questa lieta giornata in cui ho finito la mia tesi di dottorato, vorrei nuovamente ringraziarVi calorosamente per la Vostra buona volontà e per la collaborazione al mio lavoro scientifico. Mi sono sentita ben voluta nelle Vostre scuole e mi ha fatto piacere che tanti giovani fossero disposti a partecipare alla mia ricerca. Effettuarla sarebbe stato impossibile senza il Vostro aiuto. Mille grazie per tutto!

In het Hongaars kom ik helaas niet verder dan: *Köszönöm szépen!* Hoewel dit de essentie van mijn boodschap weergeeft wil ik er toch iets meer van maken. Ik heb daarom de hulp van een collega ingeschakeld en we lezen het volgende:

### **Kedves** tanárok és diákok!

A mai szép napon, amikor befejeztem a doktori disszertációmat, újra meg szeretném köszönni önöknek/nektek az együttműködést a tudományos munkámban. Nagyon jól éreztem magam az osztályokban és nagyon örültem, hogy olyan sok fiatal vállalta, hogy részt vesz a kutatási projektemben. A rengeteg segítség nélkül nem sikerült volna befejeznem a vizsgálatot. Nagyon szépen köszönök mindenkinek mindent!

Bedankt voor deze vertaling, Krisztina, *köszönöm szépen!*

Het motto van mijn proefschrift, *“The future belongs to those who believe in the beauty of their dreams”*, is afkomstig van Eleanor Roosevelt, voormalig First Lady van de V.S., een sterke, wilskrachtige vrouw. Ik wil haar woorden meegeven als persoonlijke boodschap aan mijn kinderen - Hans en Marcha, Karin en Serge - en aan mijn kleinkinderen - Michael, Katy, Rick, Ferry, Britt en Nina. Laat je bij het vinden van je weg door het leven niet ontmoedigen.

Mijn eigen dromen ontlene hun schoonheid vooral aan Chris, mijn maatje door dik en dun. Zijn rotsvaste geloof in mij is een van de pilaren onder mijn onderzoek en dit proefschrift. Vanaf het prille begin leefde hij mee met mijn zoektocht naar onderzoeksscholen, mijn hoofdbrekens over testen en vragenlijsten, het monnikenwerk van invoeren van data in SPSS, het bestuderen van statistiek en de eindeloos lijkende herzieningen van manuscripten. Zonder hem had ik ook nooit met de auto of de boot naar mijn onderzoeksscholen door heel Europa kunnen reizen. Iedere keer weer zorgde Chris ervoor dat ik stipt op tijd in de klaslokalen stond met mijn stapels papierwerk. Chris, bedankt voor de gedenkwaardige tijd die we samen met al die aardige en bijzondere onderwijzers hebben beleefd. Bedankt dat je alles zoveel leuker maakte en nog steeds maakt!

Tenslotte komt er nog een groot boekje van bedankjes aan kinderen, kleinkinderen, broers, zus, verdere familie en schoonfamilie, vrienden, oud-collega's, RU medewerkers en aio's voor alle bemoedigende woorden, enthousiaste reacties, belangstellende vragen en welgemeende complimenten die me zoveel goed hebben gedaan. Jullie maakten vaak het verschil tussen een gewone weekdag en een dag met een gouden randje. En natuurlijk bedank ik mijn zoon Hans en mijn dochter Karin omdat ze mijn paraniften willen zijn. Jullie enthousiaste reacties maakten me blij.

Ik voel me een bevoorrecht mens. Ik prijs me gelukkig omdat het me gegeven is om mijn naam op dit proefschrift te kunnen zetten. Maar nog meer omdat ik mijn naam onder dit dankwoord mag zetten.

*José Goris*





