

**Content learning in the L1 and in the L2: A comparative study of CLIL
students in grade 9 and grade 10 in Dutch secondary schools**

MA Thesis

**Lisa de Goede
Utrecht University
Linguistics: The Study of the Language Faculty MA
3139069
Supervisor: Rick de Graaff**

Summary

This study has focused on content learning in bilingual education in the Netherlands. Throughout the Netherlands, over 120 schools have adopted bilingual programmes allowing students to both learn a second language (most often English) and study part of their school subjects through this same language, an approach coined Content and Language Integrated Learning (CLIL). Various researchers, teachers as well as parents have expressed their doubts with regards to the effect this has on students' ability to fully understand and express what they are asked to learn in school.

Based on similar research by Gablasova (2012), the goal of this study was to assess and compare the effect of L1- versus L2-mediated learning on the acquisition of content knowledge, and on learning new content words specifically. Furthermore, the study aimed to address the fact that in the Netherlands, students are confronted with a 'language switch' after grade 9: from then on, many of the (content) subjects previously taught in English are instead taught in Dutch, due to the national final examinations also being in Dutch. Students also receive fewer hours of English-taught lessons from grade 10 and onwards.

To assess these matters, 45 students in 9th grade and 51 students in 10th grade were asked to read and listen to an academic text on a historical topic. Half of the students in both grades read an English version of the text, while the other half read the Dutch version. All students were then asked to complete a written test which asked them to recall and define 13 target words included in the text, half of them in English and half of them in Dutch.

The results showed that students performed best in the language in which they also read the text, regardless of whether this was their L1 or L2. L1 readers in grade 9 were found to struggle when translating their knowledge into their L2, possibly due to the fact they are still acquiring this language. Overall, L2 readers performed best in grade 9, whereas in grade 10, L1 readers outperformed the L2 readers. Furthermore, L2-instructed students in grade 9 outperformed the L2-instructed students in grade 10. One of the possible causes of these differences can be sought in the language shift between grade 9 and 10 and the reduced amount of English-taught lessons.

Table of contents

1. Introduction	4
2. Theoretical framework	5
2.1 <i>Bilingual education and Content and Language Integrated Learning (CLIL)</i>	5
2.1.1 <i>CLIL in the Netherlands</i>	6
2.2 <i>Research on CLIL</i>	7
2.2.1 <i>Subject learning through an L2</i>	9
2.2.2 <i>Subject learning through written sources in the L2</i>	10
2.2.3 <i>Learning subject words</i>	11
2.2.4 <i>Expressing content knowledge in written form</i>	14
2.2.5 <i>Role of L2 proficiency in bilingual education</i>	15
2.2.6 <i>L1/L2 transfer</i>	16
2.2.6.1 <i>Transfer of reading and writing strategies between the L1 and L2</i>	17
2.3 <i>Recent work by Dana Gablasova (2012)</i>	18
2.4 <i>Considerations and aims of this study</i>	19
2.5 <i>Research questions</i>	
3. Methodology	24
3.1 <i>Study design</i>	24
3.2 <i>Participants</i>	24
3.3 <i>Materials</i>	26
3.3.1 <i>Instruments used to answer Research Questions 1 and 2</i>	26
3.3.1.1 <i>Academic text and target words</i>	26
3.3.1.2 <i>Post-Test</i>	28
3.3.2 <i>Instruments used to answer Research Question 3</i>	30
3.3.2.1 <i>Proficiency testing</i>	30
3.4 <i>Procedure</i>	31
3.5 <i>Analyses</i>	32
3.5.1 <i>Coding and analysis of Research Question 1</i>	32
3.5.2 <i>Coding and analysis of Research Question 2</i>	36
3.5.2.1 <i>Was there any difference between the students in grade 9 and 10 in terms of total recall of the definitions?</i>	37
3.5.2.2 <i>Was there any difference between students in grade 9 and 10 in terms of the number of meaning components recalled?</i>	37
3.5.2.3 <i>Was there any difference between students in grade 9 and 10 in terms of the quality of the definitions?</i>	38
3.5.3 <i>Coding and analysis of Research Question 3</i>	39
4. Results	40
4.1 <i>Analysis of data for Research Question 1</i>	40
4.2 <i>Analysis of data for Research Question 2</i>	43

4.2.1	<i>Results of the analysis of the definitions provided by grade 10 students</i>	43
4.2.2	<i>Results of the analyses comparing the definitions provided by grade 9 and grade 10 students</i>	46
4.2.2.1	<i>Overall recall of target words</i>	46
4.2.2.2	<i>Mean percentage of meaning components recalled</i>	48
4.2.2.3	<i>Semantic quality of the definitions given</i>	51
4.3	<i>Analysis of data for Research Question 3</i>	52
5.	Discussion	55
5.1	<i>Summary of the results</i>	55
5.2	<i>Discussion of results for Research Question 1</i>	57
5.3	<i>Discussion of results for Research Question 2</i>	59
5.4	<i>Discussion of results for Research Question 3</i>	62
5.5	<i>Limitations</i>	62
5.5	<i>Recommendations for CLIL teaching practice</i>	63
6.	Conclusion	64
7.	References	66
	Appendix 1	71
	Appendix 2	73
	Appendix 3	76
	Appendix 4	79

1. Introduction

Over the last few decades, Content and Language Integrated Learning (CLIL) has been a fast-growing educational approach in secondary schools all over Europe. In the Netherlands alone, this approach to bilingual learning is implemented in over 120 schools. CLIL, which is an approach to bilingual education that allows students to learn content through the second language, thereby teaching both the content as well as the language, has so far been the subject of many studies. Those that have focused on second language learning developments have generally found greater improvement for CLIL students compared to regular education peers. However, less is known about the effects of L2-mediated education on content learning. Those studies that have sought to further examine these effects, have produced varying results. In the Netherlands, specifically, there has not been a strong focus on content learning in research on bilingual education. Since this type of learning, through an additional language that is still being acquired by students, may have an effect on the way they learn and express their knowledge, it is important to assess and compare the effect L1 versus L2-mediated learning has, not only on second language development, but on content learning, too.

This study aims to critically assess content learning in secondary schools with CLIL programmes in the Netherlands. It sets out to analyze both the learning and testing of subject knowledge, and new content words in particular. Furthermore, while previous research carried out in Dutch contexts with a focus on content have tended to look at final examination results during the final year of (CLIL) education, this study will look at the language switch which is currently part of CLIL programmes in the Netherlands. This switch entails a transition from students being taught content subjects such as History and Geography in the L2 English in ninth grade, to these lessons being taught in the L1 Dutch in tenth grade and onwards, since the final examinations for these subjects are to be completed in the L1. By testing students in History classes in both grade 9 and in grade 10, this thesis seeks to examine the effects of this shift in the language in which they are being taught. And finally, as previous studies have found that L2 proficiency plays an important role in students' performance in various modalities, this factor will also be addressed.

Chapter 2 will outline the theoretical framework of this study. The methodology that is employed in terms of the experiments as well as the analyses thereof will be explicated in Chapter 3. The results of the study will be reported in Chapter 4, followed by a discussion of these results in Chapter 5. The conclusion is laid out in Chapter 6.

2. Theoretical Framework

This chapter aims to explore the underpinnings of bilingual education in Europe, and in the Netherlands, specifically. It furthermore sets out to discuss the research that has been conducted exploring bilingual education, with a focus on a few specialized areas of research related to the goal of this thesis.

First, bilingual education will be described from an international perspective, discussing Canadian immersion programmes and Content and Language Integrated Learning (CLIL) education in Europe. Then, the way in which CLIL is currently implemented in the Dutch educational system is outlined. General research trends concerning CLIL will be laid out, followed by a more thorough discussion of research focusing on several aspects of bilingual education that are of particular interest for this thesis: subject learning through an L2; learning knowledge through reading written sources (e.g. textbooks); learning subject words; expressing content knowledge through writing; the role of L2 proficiency in bilingual education and finally, L1/L2 transfer. Following this discussion, the research questions this thesis sets out to answer will be formulated.

2.1 *Bilingual education and Content and Language Integrated Learning (CLIL)*

In 1965, native English-speaking parents living in the predominantly French-speaking environment of Quebec in Canada started a local grass-roots movement that resulted in what is now known as French immersion. They argued that it would be highly beneficial for their monolingually raised children to also become proficient in the language most commonly spoken in their area. Currently, French immersion programmes are implemented as optional educational programmes that, by way of teaching a significant part of the curriculum in French, encourage students whose first language is not French to communicate in this target language, with the goal of attaining functional fluency in it while also fully mastering the English language. Since its start in the 1960's, French immersion has been very successful in Canada, not in the least because of the support from education authorities as well as the parents' enthusiasm, as argued by Shapson (1984).

In Europe, schools offering part of the curriculum in a foreign, regional or minority language have existed for several decades. At first, this concerned mostly schools in linguistically diverse regions or in large cities (Eurydice 2006). Starting in the 1990s, Europe has seen a strong surge of schools offering a bilingual stream, with the Canadian immersion programmes as a source of inspiration. Dalton-Puffer (2008) notes that this rise in bilingual programmes started at two levels: both from local grass-roots movement, reminiscent of the start of French immersion, as well as at the level of European Union policies. Bilingual education in Europe is based on the concept of Content and Language Integrated Learning (CLIL). In bilingual education according to the CLIL approach, a context is created where content learning is achieved through using the target language as the medium of

communication. CLIL can therefore be seen as a dual-focused programme that, besides teaching content, also has several language-specific goals: for example, students will acquire subject-specific terminology in the target language (often English), they will improve their overall competence in this second language (L2), and they will develop oral communication skills in the L2 (Dalton-Puffer 2007).

CLIL differs from Canadian immersion programmes in the sense that in Canada, the language of instruction is the country's (other) official language, with native speakers teaching the subjects that make up the French part of the curriculum. In European countries, the target language of CLIL programmes is often English, which, though it is a prestigious language, is never the countries' official language. Teachers can be native speakers, but are often native speakers of a country's L1 who have completed extra courses in order to be able to teach subjects in their L2.

In many countries, the part of the curriculum of CLIL programmes that is taught in the target language is parallel to that of the regular classes. This means that similar subject material is covered in CLIL and non-CLIL classes; the only difference being the language of instruction and teaching materials. It has to be noted, though, that CLIL is not a programme in and of itself. Gablasova (2012) notes that it can be seen as an umbrella term covering an array of programmes currently running in Europe that include part of the curriculum being taught in an additional language.

2.1.1 *CLIL in the Netherlands*

In the Netherlands, schools started offering CLIL during the 1990s, due to parents' and schools' desire to offer students something extra. The Alberdingk Thijm school in Hilversum was the first, in 1989, to decide to meet parents' wishes and offer a more internationally oriented style of education, formerly only accessible for children of internationally mobile parents. In 1992, more schools followed their example. Currently, as of July 2013, over 120 secondary schools offer CLIL in the Netherlands. Almost all of these offer bilingual education to vwo-students¹, with 49 and 19 schools respectively also offering CLIL as part of the curriculum for havo-students and vmbo-students.² Of these schools, there are two offering bilingual Dutch-German education, as they are situated in the border area of the Netherlands and Germany. The rest use English as the target language (Europees Platform 2013).

Bilingual education is developed collaboratively but implemented individually by Dutch schools, rather than there being a set programme implemented at each school. In order to guarantee quality and consistency between schools, a national network for bilingual education (*landelijk netwerk tto*), coordinated by the European Platform (*Europees Platform*), has developed a bilingual education standard (Europees Platform 2014). This standard stipulates goals in various areas. In terms of language proficiency, for instance, for vwo- and havo-levels it states that students need to reach B2

¹ Vwo is the highest level of general secondary education available in the Netherlands

² Havo: senior general secondary education.

Vmbo: preparatory secondary vocational education

(vwo) or B1 (havo) level according to the Common European Framework of Reference for Languages (CEFR) at the end of grade 9 for speaking, writing, reading and listening proficiency and conversation skills. Furthermore, CLIL students' average grade for the subject Dutch should not deviate negatively from the national average grade for the final Dutch examination, nor should CLIL students' subject knowledge for other subjects taught in English be below average when compared to the national average grade on the final examinations.

In terms of the amount of time Dutch students have to spend in CLIL classes, at least half of the (non-language) subjects have to be taught in English. Dutch students will have to take their final examinations in Dutch, which means that from grade 10 onwards, those subjects that will be part of the central final examination will be taught in Dutch again. Other subjects, such as Social Sciences and P.E., are still taught in English after grade 9. At least two of the L2-taught classes need to be taught by native English speakers; other teachers teaching CLIL classes need to have at least a B2 level of English proficiency (CEFR).

Besides a clear focus on content and language, another goal of bilingual schools in the Netherlands, as set by the standard devised by the European Platform, is to provide students with a European and international perspective by incorporating international topics during lessons. Schools also offer students all sorts of international activities (e.g. exchange trips, participation in Model United Nations).

As there is strong demand for bilingual education in the Netherlands, most schools use selection procedures, which are generally based on the prospective students' performance on a test held at the end of most primary schools, as well as their motivation (Eurydice 2006).

2.2 *Research on CLIL*

The Eurydice Network, which analyzes education policies and organizations throughout Europe, published a report in 2006 which showed that CLIL was "still far from being a consolidated and fully articulated educational model in any of the European countries surveyed", leading Dalton-Puffer to argue that "a great deal more needs to be done, for instance, in order to consolidate theoretical underpinnings of CLIL and to create a conceptual framework both coherent and applicable to different local conditions" (2008: 139). Applied linguistics research was (and is) a necessity in order to accomplish this. After a slow start, from about 2005 onwards, more and more research focusing on CLIL has been carried out in various countries and contexts.

Studies that have looked at language learning outcomes are generally positive about the effects of bilingual education. It is regularly reported that students in CLIL education attain a higher level of L2 proficiency than their peers in regular EFL classes. Haunold (2006), for instance, compared CLIL and non-CLIL students' scores on a standardized placement test, and found that significantly more CLIL students were shown to have achieved B2 level English (CEFR). Verspoor et al. (2010), too, found

that Dutch CLIL students' L2 English proficiency levels improved at a higher rate compared to students in non-CLIL classes, especially during their first year in CLIL education. Following their first year, CLIL students retained this lead.

However, though ample evidence has been documented regarding the improvement in language learning, Dalton-Puffer (2011) noted that relatively few studies have focused on the way that learning in an L2 affects *content* learning and its outcomes. The CLIL approach to content learning has been met with some criticism, for instance from history teachers, as mentioned by Hasberg (2004). Zydatiβ (2012) writes that these teachers regard CLIL as a means of using content subjects to learn a second language, with doubts as to the positive effect this has on the transfer of content knowledge itself. Whether the language of instruction affects subject knowledge is not only a concern of teaching professionals, but of parents too. Hajer (2000) and Dalton-Puffer (2008) further observed that, since the instructional language is a language that students are still in the process of acquiring, there are some concerns as to whether this could lead to less subject competence, either because of students not fully comprehending what is taught to them or because of the possibility of teachers pre-empting this issue by reducing the complexity of content material.

In order to understand the nature of learning subject knowledge in an additional language, and to evaluate what this type of subject learning implicates for students, it is important that more research is carried out in this area. This thesis sets out to do just that, by analyzing the learning and testing of subject knowledge by CLIL students. In doing so, it will contribute to the knowledge on content learning in an additional language, as well further investigating the role of L2 proficiency levels in the expression of academic content knowledge. It also aims to contribute to the practical issue of the language of assessment in bilingual education, by investigating any advantages or disadvantages of using the L1 or the L2 as the language of assessment of content knowledge. Finally, the aspect of written production of students in CLIL programmes will also be further analyzed.

First, however, the following sections will review some of the work that has been done regarding several aspects of bilingual education that are of particular interest considering the goals of this thesis, as mentioned above. Section 2.2.1 will discuss studies that have focused on subject learning through an L2, followed by a summary of research on learning and retaining content knowledge through written sources (e.g. textbooks) in section 2.2.2. Studies analyzing the learning of subject words will be reviewed in section 2.2.3, and those studies that focused on expressing content knowledge through writing in the L2 will be summarized in section 2.2.4. Research that has looked at the role of L2 proficiency in bilingual education will be discussed in section 2.2.5, and finally, in section 2.2.6, previous work on L1/L2 transfer will be summarized. Following this review of relevant literature, the research questions of this thesis will be formulated.

2.2.1 *Subject learning through an L2*

Duff (1997) was one of the first researchers to publish an evaluation of a CLIL programme, describing the performance of students in three high schools in Hungary. Based on their scores on national leaving examinations and university admission rates, Duff concluded that the CLIL students either performed similarly or better compared to their peers in regular education in terms of content mastery.

Similarly, Huibregtse (2001), in her study on the effects of Dutch bilingual education, concluded that students in CLIL education did not perform worse than L1-instructed students. Their performance on content subjects, however, was only measured by looking at their grades for these subjects on the national leaving examinations, as in Duff's (1997) study. It is also important to note that these grades reflect the students' performance on exams conducted in their L1, Dutch. As noted in section 2.1.1., CLIL students in the Netherlands only receive L2 instruction for subjects that involve national examinations up until grade 10, which means that students will have had three years of L1 instruction by the time they are to complete their final examinations. Whether these grades properly reflect students' L2 subject learning performance is debatable.

Other studies have reported that materials of a more complex and abstract nature caused more trouble when studying them in the L2 rather than in the L1. Jäppinen (2005), in a study on students in primary and lower secondary CLIL education in Finland, noted that students' incomplete L2 acquisition led to them having trouble mastering the finer details of more advanced subject knowledge.

In her study on the effect of CLIL education on students in 10th grade in Germany, Coetzee-Lachmann (2007) found that their performance for geography was lacking both in terms of the amount of knowledge gained when compared to grade 10 standards, and in terms of the correctness of this knowledge. L2 instruction, therefore, led to a disadvantage for these students.

Other researchers have found that CLIL students perform similarly to their peers taught in their L1. Stohler (2006), for instance, looked at content learning by L2-taught students in Switzerland. By interviewing students who had taken a class taught in either the L1 or the L2, and comparing their knowledge of the subject that was taught in this class, and repeating this process two months later, Stohler found that there was no significant difference between the performance of the students taught in the L1 and those taught in the L2.

Summarizing, studies that have focused on the effect of L2 instruction on subject learning have produced mixed results, with some finding that L2-taught students perform similarly or even better than L1-taught students, whereas others report a detrimental effect on students' performances when the language of instruction was their L2 rather than their L1. As the way in which CLIL is implemented varies from country to country (and even within countries, for instance in the Netherlands as discussed in section 2.1.1), and given the different populations used for each study, this is perhaps not a surprising result.

2.2.2 *Subject learning through written sources in the L2*

Written texts are one of the main sources of subject learning in secondary school. Consequently, various studies have looked at the effect of reading a text in the L1 versus reading a text in the L2 on the processing of their content. Some of the earliest studies on L2 reading comprehension aimed to find out whether comprehension problems arose because of difficulties with reading in itself or reading in the L2 in particular (e.g. Alderson 1984). Various studies found that a threshold level of L2 proficiency needs to be reached before L1 reading skills can be used during reading in the L2 (e.g. Lee and Schallert 1997, Bernhardt and Kamil 1995). Furthermore, what often tends to be forgotten is that the acquisition of reading skills in the L1 takes years. Grabe and Stoller note that “Seldom are L2 students given as much time to develop stronger reading abilities, despite similarly demanding expectations for success” (2011: xiv), when compared to L1 students.

Nation (2001) also noted that one of the problems that L2-taught students might run into is their limited L2 vocabulary, which can affect their ability to understand a text. According to Grabe (2009), students with a smaller L2 vocabulary can encounter problems with comprehension even with easier texts.

Donin and Silva (1993) analyzed content recall by L2 users. In the first of two studies, a group of 27 Canadian nursing students were asked to read three texts related to their occupation in both their L1 (English) and their L2 (French). Content comprehension was tested in both languages, showing that the participants generally recalled more information from the L2 text in their L1 rather than in the L2. The second study looked at nine nursing students, all proficient L2 French speakers. Students were able to produce more inferences when asked to recall the L2 texts in the L1 than in the L2. Donin and Silva argue that the results of these studies show that using the L2 as language of recall not necessarily accurately reflects L2 comprehension, with L2 proficiency playing a role: a higher L2 proficiency level leads to more similar L2 text comprehension when compared to L1 text comprehension. The role that familiarity with the topics of the texts played in these studies may have influenced the participants’ ability to recall information, however.

In another study, Donin, Graves and Goyette (2004) looked at text comprehension and information recall by sixteen officers that were part of a French language programme in the Canadian Armed Forces. These L1 English speakers were put into two L2 French proficiency groups based on their scores on an initial placement test. They were asked to read texts in both English and French, and overall were found to be able to recall more content from the English texts than the texts in French, with their proficiency levels as well as the language of the text also influencing their reading times.

Zydatiβ (2007) published a large-scale evaluation of bilingual education at two grammar schools in Berlin, including an assessment of students’ subject-matter competence after reading a variety of texts. After comparing CLIL students’ performance, who read texts in their L2 English, to that

of regular students who read the texts in L1 German, Zydatiβ concluded that the CLIL students' competence in processing academic discourse and completing tasks testing the knowledge gained after reading the texts was at a similar level compared to that of the L1-taught students. It was also noted, though, that discourse proficiency was strongly related to language proficiency, with a high level of L2 proficiency necessary for an adequate performance when processing academic discourse. This suggests that limited proficiency in the L2 could result in insufficient success with subject learning in this second language.

In summary, all studies focusing on the acquisition of information through written sources reviewed here found that L2 proficiency influenced their L2 text comprehension, reading times and subject learning from these texts. Participants reading in their L2 were found to be more likely to make errors in their understanding of the texts or recall less information compared to their performance on L1 texts. Readers with a higher L2 proficiency retained information at a level closer to that of native speakers.

2.2.3 *Learning subject words*

Gablasova (2012) argues that to successfully define a technical word requires both lexical and content knowledge, making it a way of demonstrating the extent to which academic learning has been successful. Hulstijn (2003), in an overview chapter on incidental and intentional learning, notes that very few words are acquired intentionally. Vocabulary is mostly acquired through extensive reading. His review mentions work by Nagy, Herman and Anderson (1985) and Nagy and Herman (1987), who estimated American high school students' vocabulary to encompass between 25,000 and 50,000 words, arguing this could not have been the result of intentional vocabulary learning only. Factors affecting learning new words through reading are, amongst others, familiarity with surrounding words, the context that is provided, and the reader's reading ability (Hulstijn 2003).

De Bot et al. (1997) furthermore maintain that the level of L2 proficiency of the reader is critical, as was discussed in the previous subsection, in determining the amount of content that a reader can process, retain and recall from the context. However, other learner factors also play a role: for instance, Paribakht and Wesche (1999) found that those words that are seen as useful by the reader are remembered more easily.

Horst (2009) reported a vocabulary learning study carried out in Canada. He asked 47 adult immigrant L2 English students to read various books at home over the course of six weeks, looking to find out to what degree they could provide definitions of newly learned words from this material. 29 students had read at least one book after the six weeks, with the remaining 18 students who had not read any forming the control group. Before the reading period, the students were pre-tested on their L2 vocabulary. Following the six weeks of reading, they were again tested on their L2 vocabulary, this

time with a list including infrequent words found in the texts they had read. Horst (2009) found that the students that had read at least one book had learned, on average, 42% of these new words, based on a comparison of their results on the L2 vocabulary pre-and post-test. This was significantly more than the words learned by the control group. Through reading, Horst (2009) concluded, students can acquire new vocabulary, and also further improve their L2 proficiency.

Lessard-Clouston (2006) investigated Canadian university students' learning of content words by comparing the breadth of native English speakers' vocabulary to that of non-native speakers. He measured their vocabulary at the beginning and end of a semester, and found that the observed difference at the start of the study, when the non-native speakers had a smaller vocabulary, had disappeared by the end of the semester. There was still a qualitative difference between the two groups, however, for although the depth of knowledge had increased for both groups, the native speakers' increased by twenty percent compared to six percent by the non-native speakers.

In education at a secondary school level, students often have to learn content words from definitions that are provided within texts or in separate lists. This type of lexical familiarization in textbooks has received criticism, particularly when words are presented in isolation in a word list, as, according to Hulstijn (2003), this leads to mostly declarative knowledge of the meaning of words. Whether students will be able to use these new content words during speaking or writing or whether they will be able to interpret them correctly in a text is a matter of concern. Learning new subject words through definitions, however, remains a common practice in schools, and is seen by others as a useful way of familiarizing students with new words in the context of other related information and clues as to their proper use in sentence structures. Snow, Cancino, de Temple and Schley (1991) also argue that definitions link new semantic and lexical information to older information, which makes it easier for learners to incorporate the new subject word into their lexical network. According to Gablasova,

lexical familiarization, one type of which is defining, is a common way of communicating the meaning of new words to students and thus developing both their lexical and subject competence. (...) definitions in this sense are an integral part of the context and thus fulfill both conditions needed for effective learning of the word – providing sufficiently unambiguous information about the word's meaning (Schouten-van Parreren 1989) in addition to all advantages that seeing a word in the context gives (2012:44).

When students read a text with the intention of learning words, the rate of retention of these words is much higher compared to acquiring new form-meaning pairs during incidental reading (Hulstijn 2003).

A study by Mondria and Wit-de Boer (1991) on intentional learning reported Dutch secondary school students' success rate after being asked to learn eight French content words from sentences

that provided context, as well as translated L1 sentences. After ten minutes of studying the words, students' average retention rate was 65%.

In order to compare the effects of incidental and intentional learning conditions, Hulstijn (1992) asked 52 L1 Dutch speakers to read a 900-word English text with 12 pseudo-words in it, with small clues in the L2 English as to their meaning. 24 of the participants were asked to read the text and to expect some comprehension questions afterwards (incidental condition). The other 28 participants knew before reading the text that they would have to complete a vocabulary task after, asking for their definitions of the pseudo-words (intentional condition). The two groups' retention rates were 4% for the incidental learning group, and 53% for the intentional learning group. When the target words were tested during another post-test, including the context that was provided in the text, these rates were 43% and 73% respectively. This study, therefore, points towards the differences in retention rates of words between participants who were explicitly told they would be tested on this knowledge, and those who were only told they would be asked questions regarding general comprehension of the text. However, as the words tested were pseudo-words, whether similar results will be obtained with actual words remains the question.

Another study on learning content vocabulary through reading by Haynes and Baker (1993) found that native speakers of English were better at retaining new words from incidental and intentional learning than Taiwanese university students, who were L2 speakers of English. The native speakers also provided better quality definitions by including more details in their descriptions of words. Here, according to Hayes and Baker, the non-native speakers were at a disadvantage because of their smaller English vocabulary, which caused them to comprehend less information from the text.

A final factor that influences word retention after L2 reading that remains to be discussed, is their reoccurrence during the text. According to Hulstijn (2003), "glossing gives a high return in terms of comprehension but a low return in terms of retention, when glossed words appear only once in a text. Retention of glossed words, however, increases substantially when they reoccur several times." (2003:364)

Summarizing, the studies described in this subsection which looked at learning new words during reading found differences between participants reading and acquiring words in their native language, and participants' retention of words after reading in their L2, with L2-readers being less effective in terms of retention rates and (increased) vocabulary size (e.g. Lessard-Clouston 2006 and Haynes and Baker 1993). Factors that were found to influence retention of content words were the context these words were presented in, L2 proficiency, participants' reading ability, perceived usefulness of a word and the reoccurrence rate of new content words. The research that was reviewed also showed that more words are retained during intentional learning conditions compared to incidental learning conditions.

2.2.4 *Expressing content knowledge in written form*

Although competence in writing is one of the language skills that CLIL aims to improve, Ruiz de Zarobe (2010) and Whittaker, Llinares and McCabe (2011) note that not much work has been published on CLIL students' written production. Lyster's (2007) research on French immersion in Canada showed that, although students' receptive skills (reading, listening) strongly improved over time, this was not the case for their productive skills, among which writing. Dalton-Puffer (2008), in her review of early studies on CLIL, listed a number of language competencies that were either favorably affected, or unaffected or indefinitely affected by L2 instruction. Students' general writing ability was noted as being either unaffected or indefinitely affected. Lasagabaster (2008) and Ruiz de Zarobe (2010) have found that CLIL students surpass students following regular EFL classes when asked to write about a general topic. Verspoor et al. (2010), too, found that students in CLIL classes in the Netherlands produced texts of significantly higher quality compared to those written by their peers, when asked to write a short text given prompts such as "Write a short story about your new school, friends and teachers" and "Write a short story about the most awful (or best) thing that happened to you during summer vacation" (2010:13). CLIL students made fewer errors, used a larger variety of verbs and more low frequency words, and produced sentences of a higher complexity.

However, studies that have looked at discipline-specific writing have found that participants' writing competence was less than stellar. Vollmer et al. (2006) and Llinares and Whittaker (2006) analyzed the writing of students in secondary schools in Germany and Spain respectively, and concluded that their written performance was lacking in various respects, among which adherence to grammar rules, coherence and proper use of discourse style. It has to be noted, however, that the writing ability of both CLIL students as well as their peers in regular L1-taught classes is in need of improvement, as argued by Dalton-Puffer (2008), who notes that similar results were found for writing tasks produced in participants' native language.

Other researchers that explored the expressing of content knowledge through writing also pointed towards some problems with CLIL students' abilities in this area. Zydati (2007) looked at various communicative competences, and found that CLIL students encountered noticeable difficulties in "expository and argumentative writing which was based on subject-matter content materials" (Jexenflicker and Dalton-Puffer 2010: 172). Similar findings were reported by Coetzee-Lachmann (2007), who looked at subject-specific discourse competence in geography of students in grade 10. Deficiencies in CLIL students' academic literacy were found in both their L2 (English) as well as their L1 (German), with insufficiently complex writing products and problems with using subject-specific vocabulary. Interestingly, Kops Hagedoorn (2009) also found a reasonably strong correlation between vocabulary and students' writing abilities.

The studies analyzing CLIL students' writing competence reviewed here have produced varying results, with students out-performing their peers in regular EFL classes when asked to write about

general topics, but running into trouble when told to write texts of a more subject-specific nature.

2.2.5 *Role of L2 proficiency in bilingual education*

One of the goals for this thesis is to address the role of students' L2 proficiency levels in their learning and expression of academic content knowledge. CLIL education aims to improve students' language proficiency, amongst other things, as stipulated in the bilingual education standard in the Netherlands, for instance (see section 2.1.1). Early research on bilingual education in the Netherlands mainly focused on the effects of CLIL on L1 and L2 proficiency. Huibregtse (2001), whose study on the effects of Dutch CLIL education was previously mentioned in section 2.2.1, measured students' L2 proficiency by testing their receptive vocabulary, oral production and reading comprehension. Her data were collected during 1991-1995, making this study one of the earliest ones of CLIL education in the Netherlands. L1 proficiency was also tested, as an important question especially during those early times when bilingual education was being implemented was whether such an emphasis on L2 English would have any detrimental effects on students' L1 Dutch. Huibregtse (2001) found that CLIL had improved their proficiency in all the areas mentioned, compared to L1-taught students.

Admiraal, Westhoff & De Bot (2006) used the same data and found that CLIL students outperformed L1-taught students on L2 oral proficiency and reading comprehension, but found no differences for receptive word knowledge.

In a general sense, previous studies have found that CLIL education has a positive effect on students' L2 proficiency levels compared to students in regular EFL classes (see also Haunold 2006; Verspoor et al. 2010). However, it is important for studies concerning this subject to assess not only whether students' L2 proficiency improves, but also what effects their various levels of proficiency have on their performance in school subjects.

Zydatiβ (2012) remarked that the possible link between L2 proficiency and students' academic performance has been the focus of various studies since the early stages of research into bilingual education. In the previous sections, some of the studies that were reviewed already pointed towards an influential role played by L2 proficiency in various aspects of L2 content learning and the expression of subject knowledge. Recall, for instance, that Zydatiβ (2007) found that a high level of L2 proficiency is essential for processing academic discourse adequately, suggesting that a limited proficiency in a second language could negatively affect L2 subject learning.

In terms of L2 reading and text comprehension, various studies concluded that in order for someone to use their L1 reading skills during L2 reading, a threshold level of L2 proficiency needs to be reached (e.g. Lee and Schallert 1997, Bernhardt and Kamil 1995). Donin, Graves and Goyette (2004) also found that L2 proficiency levels influence reading times, and Donin and Silva (1993) concluded that higher L2 proficiency levels lead to text comprehension at a level similar to that of L1 text comprehension.

In the discussion of research that has investigated the learning of subject words (section 2.2.3), L2 proficiency was argued by De Bot et al. (1997) to be one of the critical factors influencing the amount of information that a reader can process, retain and recall.

The findings discussed in this section all point towards the important role that L2 proficiency plays in academic success in CLIL settings, with higher levels of L2 proficiency leading to results similar to L1 performances in some areas, but lower L2 proficiency levels negatively influencing L2 subject learning, amongst other things.

2.2.6 *L1/L2 transfer*

Many researchers have focused on the way the L1 influences L2 production in people acquiring a second language (e.g. Cenoz 2001; Jarvis 2000; Odlin 2003). Studies on L1 transfer in L2 learners generally conclude that those with lower L2 proficiency levels show more transfer in their L2 production than do more proficient L2 learners (e.g. Poullisse and Bongaerts 1994). Celaya (2008) notes that studies that have looked at transfer from a communicative or compensatory strategy tend to show that L2 learners produce less instances of transfer as they become more proficient (e.g. Granena and Celaya 2001; Olsen 1999).

According to Agustín Llach, “transfer from the mother tongue (be it voluntary or unconscious) has been observed to follow some patterns of behaviour relative to language typology, student age and proficiency in the L2” (2009:113). However, not all facets of a language can be transferred easily. Typological distance between the languages plays a big role: according to Kellerman (1977) and Dewaele (1998, 2001), only those linguistic aspects that are similar between two languages tend to be transferred. This can be seen in the language use of people acquiring a third language, who tend to transfer similar linguistic structures or lexical items from the typologically more similar language rather than their native language by default (Cenoz et al. 2001; Dewaele 1998, 2001).

Navés et al. (2005) analyzed the L2 production of students from grade 5 to grade 12, and found that the number of borrowings (L1 words not adapted to the target language) and lexical inventions (words which have been adapted to the target language morphologically, but do not actually exist (Dewaele 1998)) decreased over time, as students became more proficient in their second language.

Celaya (2008) also looked at students' L2 (English) production, focusing on lexical transfer. L1 Spanish CLIL students in grade 5 and grade 7 were compared to L2 learners in regular EFL classes, in the same grades. All students were asked to respond to the prompt “introduce yourself” in written form, and were given 15 minutes to do so. CLIL students' written production included fewer borrowings compared to the regular EFL group, and both groups showed a decrease in the number of borrowings between grade 5 and 7. Lexical inventions, however, were equally common in the writing of students from both groups, and were actually found to increase from grade 5 to grade 7. The author concludes that this difference can be explained by the different effects these two types of transfer exert

on the L2: "Borrowing results from a type of influence that works independently of the L2 system whereas lexical invention requires the L1 and L2 systems to interact closely and hence a higher degree of mastering of the L2 is expected." (Celaya 2008:45) Celaya also notes, however, that both types of lexical transfer were actually not very common in the students' L2 production, arguing that this may be due to the language mode. Written production may not involve as much lexical transfer as oral production.

Another study looking at CLIL students' transfer during L2 written production was carried out by Agustín Llach (2009). Her focus was on the influence of L2 proficiency on transfer in both quantitative and qualitative terms. 30 CLIL students and 30 regular EFL students (L1 Spanish) were asked to write a letter introducing themselves to a potential host family. All students were in 6th grade (aged 11-12 years). The author distinguished between three types of transfer: borrowings, coinage (similar to what Navés et al. (2005) called lexical inventions – an L1 word adapted to the target language in order to make it look or sound similar to it), and calques (literal translations of L1 words). The regular EFL students were found to transfer more often overall in their written production, compared to the CLIL-students. It is argued that this is due to their lower L2 proficiency compared to their CLIL-instructed peers. Both groups produced calques the most of the three types of transfer distinguished. CLIL-students second most produced transfer type were coinages, with borrowings being the least often used. This order was reversed for EFL students.

CLIL students and EFL students alike, then, appear to revert to literal translations of L1 words when they encounter a lexical gap, although CLIL students do this to a lesser degree. Agustín Llach (2009) argues that, since coinages (or lexical inventions) and calques were the most frequent among CLIL students, as was the case with Celaya's (2008) participants, lexical transfer of these two types occur when learners have a higher proficiency in the target language.

The studies focusing on L1/L2 transfer summarized here again point towards L2 proficiency as an important factor influencing the type and amount of transfers found in L2 production. In general, as students become more proficient in their L2, they tend to transfer less. Lexical inventions (or coinages) were more prevalent in CLIL students' written production compared to that of their EFL peers, which was argued to be due to the fact that lexical inventions require more interaction between the L1 and the L2, compared to other types of transfer. The (perceived) typological distance between languages also influences which linguistic structures or lexemes will be transferred, and finally, it was also shown that the quantity of transfers is affected by the language mode, with less transfer being found in written L2 production compared to oral L2 production.

2.2.6.1 *Transfer of reading and writing strategies between the L1 and L2*

When acquiring a second language, it is also possible that reading and writing strategies are transferred between the native or dominant L1 and the L2.

Hardin (2001) investigated 50 bilingual students in fourth grade, and found that they employed cognitive reading strategies from their native language into their L2 reading. Interestingly, these strategies outweighed the influence of L2 proficiency. Others have also found that L1 reading strategies are transferred to reading processes in a second language. Koda (1990), for instance, found evidence for cognitive strategy transfer when her L1 Japanese participants were reading in their L2 English. Langer et al. (1990) found that her L1 Spanish-speaking participants relied on their L1 reading experience when coming across difficulties during reading in their L2 English. Those that were able to use these L1 strategies were able to comprehend English texts to a much larger extent.

Similarly, various studies have also found evidence for transfer of L1 writing strategies in the L2. Jones and Tetroe (1987), for instance, concluded that people writing in their L2 English were helped by using their L1 when remembering information relevant to the topic of their written work. Cumming (1987), in a study on L1 French writers in L2 English, concludes that participants with little experience in writing in their L2 tend to use their L1 to create content. However, writers with more experience use L2 translation into the L1 when they know this L2 writing experience will benefit their L2 writing content. In her study on L1 Chinese speakers writing in their L2 English, Lay (1982) also found a positive effect of participants' L1 on their L2 writing success. However, Kubota (1998) notes that L1 writing strategies can also have a hindering effect when transferred into L2 writing, as has previously been found in studies focusing on L1 Japanese speakers.

2.3 *Recent work by Dana Gablasova (2012)*

Gablasova (2012) recently published her doctoral thesis on subject learning and the expressing of content words by Slovakian CLIL students, in both their L1 (Slovak) and their L2 (English). Her work brought most of the previously discussed research areas together, with the objective to contribute to the understanding of subject learning in an L2. She did this by investigating 72 L2-taught students who were in their last two years of CLIL education before they went to a university (aged 17-20). She asked 35 of these students to read two texts on the history and geography of New Zealand in their L1, and 37 of them to read the same texts translated in their L2. These texts contained 12 target words each, which students were asked to define in a post-test, both immediately after reading and again one week later. The words and concepts that were selected were new to the students, which allowed for a direct comparison between the students in the L1- and the L2-group in terms of their learning and expressing of these content words. Furthermore, both groups were asked to define half of these target words in the language they read the text in, and the other half in their other language (either their L1 or L2, depending on the language of the texts they had read). This allowed for a closer look at the effects of the language of testing on the way students were able to express their newly acquired knowledge.

Gablasova (2012) also paid attention to the influence of L2 proficiency on the students' performances, noting that previous studies repeatedly linked the differences in performance between students taught in their L2 or L1 to their level of proficiency in the L2 (e.g. Airey 2010, Hincks 2008, 2010). By asking each student to complete two tasks, testing their L2 vocabulary and proficiency, she sought to understand how students' L2 proficiency levels influenced their performance on the post-tests, and ultimately find out whether different levels of L2 proficiency "affect students' ability to benefit academically from L2-delivered education" (Gablasova 2012:66).

Gablasova (2012) found that the L2-instructed participants experienced several disadvantages compared to the students who read the texts in their L1. They recalled fewer of the content words during the immediate post-test, and the content knowledge they had gained was found to be of lower quality compared to that of L1-instructed students. After one week, the students were given the same test, with results showing that the L2-taught participants' remembered even fewer words, whereas the students that read the texts written in their L1 performed similarly compared to the earlier test.

When analyzing the effect of the language of testing, it was found that L2-instructed participants encountered some problems when responding in their L1. Conversely, having read and learned the information through English helped these students perform better in terms of subject-related competence when asked to define target words in the same language (Gablasova 2012: ii).

An examination of students' L2 proficiency revealed a connection to students' ability to learn and recall the definitions of the content words in the texts. This held for both L1-instructed students as well as their L2-instructed peers. Gablasova notes that "In general, the L2-instructed students' performance was connected more strongly to their L2 mastery, both when responding in L2 and in L1" (2012:229). These results only held for the students' recall data in a general sense; when assessing the *quality* of their definitions, L2 proficiency did not predict students' performances.

2.4 *Considerations and aims of this study*

Gablasova's work, which focused on the learning and defining of content words through reading in the L1 and L2, has served as an example for this thesis. Though some of the studies discussed in the previous sections showed mixed results, a large share of them also pointed towards disadvantages encountered by participants using their L2. Gablasova (2012) notes that the earliest studies on bilingual subject learning were often large-scale studies, which could explain why they reported varying results. Though they provided valuable insights into long-term effects of content learning in bilingual education, they often covered heterogeneous populations, and were conducted using a variety of methodological approaches. This, she argues, limits the extent to which they are generalisable to CLIL populations in different countries (Gablasova 2012). Gablasova's (2012) realistic yet controlled study design makes her study replicable in contexts with similar core population

characteristics, which is one of the reasons it was adopted (and adapted, see chapter 3) for this thesis, which focuses on CLIL students in the Netherlands.

Other researchers have addressed some of the gaps that currently exist in the studies on bilingual education in terms of the focus areas discussed in section 2.2. According to Whittaker, Llinares and McCabe (2011), for instance, not a lot of information is available about CLIL students' written production, and "When it comes to discipline-specific writing, there is even less information available (2011:344). Furthermore, recall that although many studies have analyzed students' language learning developments, not many have focused on the effect L2-based learning has on *content* learning (Dalton-Puffer 2011). Those that have, documented varying results, calling for further work to be done in this area.

Although interesting work has been done with regard to bilingual education in the Netherlands, some of which has been discussed in previous sections (e.g. Huibregtse 2001; Admiraal, Westhoff and De Bot 2006, Verspoor et al. 2010; Hulstijn 1992), the aforementioned key areas of interest for this thesis have not really been touched upon. Although Admiraal, Westhoff and De Bot (2006) did look at specific school subjects in their study, specifically Dutch, History and Geography, they only used exam results to determine the effect of L2-mediated education on students' subject knowledge. This is a decidedly course measure, which also does not take into account that after grade 9, these subjects were taught only in Dutch.

In terms of studies focusing on learning subject-related words and their meanings by means of using educational materials, Hulstijn's (1992) study on the retention of definitions after reading a text was aimed more towards exploring the effects of intentional or incidental learning, rather than the quality of the definitions. Furthermore, it did not examine how learning and retaining information through written sources is affected when doing so in an L2 as compared to in the L1, whereas Gablasova (2012) compared students' performances in both languages.

In terms of studies on expressing content knowledge in written form, Huibregtse (2001) and Admiraal, Westhoff and De Bot (2006) only reported oral production results. Verspoor et al. (2010) did use a measure of written production, however, only general L2 written proficiency was tested, as they asked students to respond to a writing prompt asking them to write a short story about everyday life (e.g. about friends, school life or their summer holidays). Therefore, no studies looked at the way students express *content* knowledge in writing.

Furthermore, regarding the results discussed in Admiraal, Westhoff and De Bot (2006) overall, they state that the conclusions of their investigation cannot "simply be generalized to effects of bilingual education in the Netherlands at this moment" (2006:91), as the schools that were part of their study were all pioneer schools, with bilingual secondary education in the Netherlands since having changed as a result of educational reform.

It seems, then, that the research community focusing on bilingual education in general, and in

the Netherlands specifically, could benefit from studies looking at content learning through the L1 as well as the L2. As content learning in an additional language (that is still being acquired) may have an effect on the way bilingually educated students learn and express their knowledge, it is important to gather more data that reveal these effects and also to discuss their possible implications. Gablasova (2012) has made a great step in this direction. However, her results are insufficiently conclusive for bilingual education in the Netherlands, as the way CLIL is implemented differs between Slovakia and the Netherlands.

The current study aims to provide further evidence regarding content learning in CLIL, by looking at students in both grade 9 and grade 10. As discussed in section 2.1.1, some of the content subjects (e.g. history) are taught in L2 English in grade 9, but are taught in L1 Dutch from grade 10 onwards due to the final examinations for these subjects being in Dutch. Research focusing on bilingual education in the Netherlands has tended to look at these final examination results in order to gauge students' performances on content subjects, not taking into account this language transition between grade 9 and grade 10 (e.g. Admiraal, Westhoff and De Bot 2006). According to De Graaff (2013), how Dutch CLIL students perform at the end of grade 9, when most content subjects are still taught in English, is still unclear. This study, then, not only aims to provide an insight into content learning through the L1 and the L2 in general, but also seeks to address this language shift that is currently part of CLIL programmes in the Netherlands, and its consequences with regards to content learning and expression.

In doing so, it will complement the results originating from large-scale studies, and will help identifying any patterns in L2 content learning from different contexts. Furthermore, it will analyze whether any differences between L2- and L1-mediated content learning exist. If they do, these differences and their implications will be described so that educators may benefit from this knowledge. This study also contributes to the practical issue of the language of assessment in bilingual education. Any advantages or disadvantages of using the L1 or the L2 as the language of assessment of content knowledge will be discussed.

Furthermore, the role of L2 proficiency levels in the expression of academic content knowledge will be analyzed, as will the written production of students in terms of expressing content knowledge.

As mentioned at the beginning of this section, this study is based largely on Gablasova's (2012) study (summarized in section 2.3). Various methodological changes were made with practical reasons and time constraints in mind, which will be discussed in the following chapter. The biggest conceptual change has already been mentioned: this study tests students in grade 9 and 10 due to the Dutch educational system and the language switch between these grades for some of the content subjects in CLIL education, whereas Gablasova (2012) tested students in their final two years of CLIL education.

2.5 *Research questions*

The overall goal of this study is to provide further insights into the way L2-taught students acquire and use content knowledge compared to students taught in their L1, in CLIL settings in Dutch secondary schools. In order to achieve this goal, the following research questions have been formulated, adapted from Gablasova (2012:67):

Research question 1: *What is the difference in the level of understanding and the expressing of the meanings of new content words learned through L1 Dutch and through L2 English by grade 9 students in Dutch CLIL schools?*

Research question 2: *What is the difference in the level of understanding and the expressing of the meanings of new content words learned through L1 Dutch and through L2 English by grade 10 students in Dutch CLIL schools, and if any differences are found, how do they relate to possible differences found for grade 9 students?*

Research question 3: *What is the effect of L2 English proficiency on the ability of students in grade 9 and 10 in Dutch CLIL schools to learn and define the meanings of newly acquired content words?"*

The first research question is essentially a combination of Gablasova's first and second research question (2012:66). While she focuses on the level of understanding of new content words in one research question and delves into the expressing of the meanings of these words in a second research question, the choice was made to combine the two in one research question for the current study. In order to answer the first research question here, the methodology employed by Gablasova to answer her first research question will be largely followed (for adaptations, see Chapter 3). This research question was renamed as it is argued that it does in fact also look at the differences in the expressing of the meanings of the content words, in that the semantic quality of the content will be assessed by analyzing the overall content quality of the elicited definitions (see further Chapter 3, section 3.5.1, 1c). Gablasova's (2012) second research question is not reproduced in the current study, in terms of the methodology she used to explore it, due to time constraints which necessitated a focus only on those aspects deemed most relevant and interesting to study in a Dutch context.

Other adaptations to the research questions involve the incorporation of the languages that were relevant for the current study, as well as the grades that it focuses on. A final change is reflected in the inclusion of a comparison of the performance by students in grade 9 and in grade 10 (see section 2.4). Since no previous studies have focused on the performance of students in Dutch CLIL schools at this point in their school career, nor on the language switch between these two grades and its effects on students' performance, no hypotheses on the basis of the literature can be made.

The way the level of understanding of the meanings of new content words by the students is examined, will be explicated in section 3.5.1 of Chapter 3 (1a and 1b). How their expressing of the meanings of the new words is measured can be found under 1c in the same section. Finally, the way students' L2 English proficiency levels were measured can be found in section 3.3.2.1.

3. Methodology

This chapter describes the methodology employed in this study. First, the study design is explained in section 3.1. The students that participated will be described in section 3.2, followed by a discussion of the materials that were used in section 3.3. The testing procedure is laid out in section 3.4. Finally, in section 3.5, the (statistical) analyses used to answer the research questions stated in the previous chapter will be explicated. As discussed in the previous chapter, this study is based on Gablasova's (2012) study on L2 content learning. Due to practical and time-related constraints, various changes were made with regards to the methodology employed by Gablasova (2012). These will be discussed in the sections below.

3.1 *Study design*

In this study, a total of 78 Dutch CLIL students in two 3 vwo³ classes (N = 45) and two 4 vwo⁴ classes (N = 51) read and listened to an academic text. The students in 3 vwo had received 3 years of bilingual education at that point, the students in 4 vwo had received 4 years. The text they were asked to read contained 13 target words that were new to the students. One 3 vwo and one 4 vwo class read and listened to the text in their L2 (English), while the other 3 vwo class and 4 vwo class read and listened to the text in their L1 (Dutch). After this, all students participated in a test that asked them to define the target words in written form. Half of these questions were stated in English, the other half were asked in Dutch. Participants had to reply in the language the question was asked in. At a later date, students in the two 3 vwo classes completed a test measuring their L2 vocabulary, as a measure of L2 proficiency.

3.2 *Participants*

All students that participated in this study went to the same secondary school. Only 3 vwo (14-15 years) and 4 vwo (15-16 years) students in the bilingual stream were tested. The school that participated offers secondary education at the vmbo, havo and vwo level. It has offered bilingual vwo education, besides a regular vwo programme, since 2002, and has been officially recognized as a senior bilingual education school (*tto school*) by the European Platform. This means that the school satisfies all of the quality standards for bilingual education set by the European Platform. The school offers both bilingual athenaeum and bilingual grammar school at the vwo level. The latter incorporates three extra subjects, Latin, Greek and Classical Culture (*Klassiek Culturele Vorming* or *KCV*), as part of the curriculum. In order to be admitted to the bilingual vwo stream at this school, prospective students have to receive a strong bilingual vwo-recommendation from their primary school, and their CITO test

³ grade 9

⁴ grade 10

result needs to be a score of 545 or higher (out of a maximum of 550). This test is administered by most primary schools in the Netherlands, with the results serving as an indicator for the secondary school level most suited for each pupil.

The school also started a bilingual havo stream in the school year 2013-2014. Since this is a recent development, at the time of testing only students in grade 7 were part of this CLIL stream, and were not included in this study.

This study compares both the performance of the students in the two 3 vwo classes, who either read and listened to the academic text in their L2 English (group 1) or in their L1 Dutch (group 2), and students' performances in the two 4 vwo classes, who also read and listened to the text in their L2 (group 3) and L1 (group 4). Furthermore, the performances of 3 vwo students will be compared to those in 4 vwo classes. Table 1 provides an overview of the participant groups.

Table 1. Overview of the participant groups.

Group	Type	Number of students	Text language	Language used during testing
1	3 vwo (grade 9)	18	L2	L1 + L2
2	3 vwo (grade 9)	27	L1	L1 + L2
3	4 vwo (grade 10)	24	L2	L1 + L2
4	4 vwo (grade 10)	27	L1	L1 + L2

CLIL students, rather than students in the regular vwo stream, were recruited to read and listen to the L1 version of the text in order to control for factors like socio-economic status, academic performance and motivation, which are likely to be higher in students in CLIL classes than in regular vwo classes since the school uses selection procedures. Besides this, like in Gablasova's (2012) original study design, students are asked to answer half the questions in the Post-Test in their L1 and the other half in their L2, in order to provide a direct comparison of the effect of the instructional language. For this reason, too, it was essential that the groups reading the L1 text were also CLIL students.

As was mentioned in the previous chapter, Dutch CLIL students are taught certain content subjects, like History and Geography, in English until grade 10. From grade 10 onwards, the subjects are taught in Dutch. As one of the goals of this thesis is to gain some more insights into the effect this language change between grade 9 and 10 has on students' subject learning and ability to retain and recall content words in both languages, the choice to test students in grade 9 and 10 marks a departure from Gablasova's (2012) work. Whereas she tested participants aged 17 to 20 years, this study, because of its focus on grade 9 and grade 10, instead tested students of 14 to 16 years of age.⁵

⁵ The target CEFR level at the end of grade 9 in the Netherlands is B2. Gablasova (2012) does not mention any CEFR levels of students in Slovakian bilingual schools, but does mention that all participants in her study were in

Furthermore, Gablasova (2012) only compared the performance of those participants that had read and listened to the text in their L1 versus those who had read and listened to it in their L2, regardless of their ages, whereas this study explicitly compares performances between the two grades as well.

3.3 *Materials*

3.3.1 *Instruments used to answer Research Questions 1 and 2*

3.3.1.1 *Academic text and target words*

The teaching materials that were used in order to answer the first two research questions were one medium-length textbook text on the history of New Zealand, as well as a recorded version of this text. The knowledge gained from reading and listening to this text will be tested using the Post-Test, to be discussed in section 3.3.1.2.

The English version of the text was constructed by Gablasova (2012), based on an authentic history text⁶, and included six subject-specific words which were defined or explained in-text. As it was important that the subject of the text was unknown to the participants, and to make sure that the academic level of the text was suitable for the younger participants in this study, the text was shown first to (CLIL) History teachers at the school in question. Upon confirmation for both these matters, the text was used for this study.

Gablasova's (2012) actually employed two texts, one on the history of New Zealand (the text adapted for this study) and one on its geography, with six target words included in each text. Due to the smaller scope of this thesis, only one of these texts was used. However, it was important that more target words were tested than the original six target words in the text in order to acquire sufficient data, so with this in mind, seven more target words were added to the History text.

These target words had to satisfy various conditions. First of all, they had to be technical terms that related to the topic of the text. Secondly, all of them had to be nouns, as Gablasova (2012) noted that previous work on eliciting definitions found that using nouns rather than adjectives or verbs resulted in more formal Aristotelian definitions, consisting of a superordinate and a complement (e.g. "A cat (*definiendum*) is an animal (*class word*) that has four legs and a tail (*definitional features*)") (Marinellie and Johnson 2004). Finally, there had to be a cross-linguistic similarity between the English words and their translated Dutch counterparts in terms of form and pronunciation (Gablasova 2012:77), in order to ensure that a comparison could be made between the Dutch and English texts in terms of the acquisition and recall of these words (note that students had to answer questions in the test in their L1 and L2, even though they had read the text in only one of these languages – therefore, the target words needed to be similar in these respects between the two languages). The words that

the upper secondary level of high school, and would all have had a minimum of 3,5 years of bilingual education as well as extensive English lessons in previous years at lower secondary and primary levels of education.

⁶ The primary source was *A Concise New Zealand history* (Wikibooks 2007) – Gablasova (2012:74)

were selected by Gablasova (2012) were either words of Greek or Latin origin, and were therefore cognates between Maori and Dutch too (e.g. *ecocentrism/ecocentrisme*), or were loanwords from Maori that were borrowed to English as well as Slovak, and also to Dutch (e.g. *whanau, moa, pa*).

After consulting various resources⁷ on the history in New Zealand, new target words were selected, and some of the terms originally in the text were used as new target words by adding more context regarding their meaning. To ensure these target words were similar cross-linguistically between English and Dutch, they were either Maori loanwords (e.g. *wero, haka*, see Table 2 below), or the words were cross-referenced to check for similar forms and pronunciation by consulting encyclopedias, dictionaries and other sources⁸.

The inclusion of new words and their meanings meant that Gablasova's (2012) text had to be adapted. The original text can be found in Appendix 1, the adapted version made for this study can be found in Appendix 2.

Table 2 shows the English target words chosen by Gablasova (2012) and the target words that were newly added, as well as their translation into Dutch and the language these words appeared in during the Post-Test.

Table 2. Target words.

Target word (originally in text)	Word in Dutch	Language of testing
whanau	whanau	Dutch
moa	moa	Dutch
pa	pa	Dutch
ecocentrism	ecocentrisme	English
kumara	kumara	English
moko	moko	English
Target word (newly added)	Word in Dutch	Language of testing
wero	wero	English
palisades*	palissades	English
Treaty of Waitangi	Verdrag van Waitangi	English
flax*	vlas	English
muskets*	musketten	Dutch
haka*	haka	Dutch
genealogy	genealogie	Dutch

* words that were already in the text

Following these adaptations, the text was checked by a native English speaker (a Ph.D. student in

⁷ http://www.newzealand.com/travel/en/media/features/maori-culture/maoriculture_powhirimaoriwelcome_feature.cfm and <http://www.newzealand.com/nieuw-zeeland/feature/early-settlement/>; Wikipedia.org (http://en.wikipedia.org/wiki/Treaty_of_Waitangi)

⁸ <http://www.encyclo.nl>; <http://en.wikipedia.org>; <http://nl.wikipedia.org>; Van Dale (Den Boon and Geeraerts 2005).

Linguistics at Utrecht University) for any errors. Then, the text was translated into Dutch by a native speaker (the researcher). The translation was kept as similar to the English text as possible. The Dutch text was checked for errors by four native Dutch speakers. It can be found in full in Appendix 3.

Gablasova (2012) included audio versions of the texts in her study design because she asked her participants to answer the questions in the Post-Test orally (see section 3.4). It was important, therefore, that they were made aware of the pronunciation of the target words. From a methodological standpoint, audio versions were also used because this ensured that students really took in the entirety of the text, in case students were to skip some parts during reading. For this latter reason, audio versions were included in this study, too. Both texts (the English and the Dutch version) were recorded in Utrecht Institute of Linguistics' phonetics lab. The cabins in this lab are sound-treated and contain high quality audio recording equipment. A native English speaker recorded the English text, and the Dutch text was recorded by a native Dutch speaker.

3.3.1.2 *Post-Test*

The Post-Test was a written test that students had to complete after reading and listening to the history text in either their L1 or L2. The test consisted of 31 questions in total. Of these questions, 13 asked students to define the target words presented in the text. The other 18 questions were distractor items, interspersed between the other questions. They were general questions regarding the content presented in the text, included to make sure participants were not aware that the main focus of the test and the research in general was their learning and recalling of the target words. This was important, as the participating classes could not be tested simultaneously. This way, the risk of students informing others that they had to focus on the content words and their meanings was minimized as much as possible. For this reason, more distractor items were included in the test than there were questions asking to define the target words.

Some of the questions in the Post-Test were taken from Gablasova's (2012) Post-Test. New questions were also added. First of all, questions asking students to define the newly added target words were included in the test. Then, in order to ensure that over half of the questions asked were distractor items, new questions concerning the content of the text were also added. These questions were carefully designed to not include any hints about the meanings of the target words, so as not to provide any context that students could base their answers to the target questions on.

As was noted in section 3.1, half of the questions were presented in the L2, while the other half were presented in students' L1. The two languages were not interspersed continuously throughout the test; rather, they were divided into two language blocks. Two versions of the Post-Test were made, with the first starting with the English block of questions and ending with the Dutch block of questions, and the second version containing the language blocks in the opposite order. This way, although all participants answered the same questions, half of the participants were first asked the Dutch questions, while the other half received the English questions first. Table 2 on the previous page

includes the language that each of the target words were asked in. The full Post-Test can be found in Appendix 4. The questions that were added for this study specifically are indicated with an asterisk. Some questions that were originally posed in Slovak when used for the research described in Gablasova (2012) were translated into English by Gablasova in personal communication, and were then translated into Dutch by the researcher. These questions are marked with two asterisks.

As mentioned at the beginning of this section, all participants had to complete the Post-Test in written form. This marks a change from Gablasova's (2012) study, as she asked students to express their acquired content knowledge orally. Various reasons led to this departure, ranging from theoretical and practical to time-related. These will all be discussed here, rather than considering the theoretical reasons in the previous chapter, in order to provide a more coherent reading experience regarding this decision, avoiding the dispersion of this explanation between chapters.

In terms of more theoretically-based reasons for the switch to written responses, oral language has long been of prime interest to research looking into CLIL, as formal education in the form of lessons in school is largely based on oral instruction and discussion (e.g. Dalton-Puffer 2005, 2007; Nikula 2007). Furthermore, the category of oral fluency is the one that is most often observed in empirical studies when aiming to determine the extent of language learning in CLIL-taught students. (Jexenflcker & Dalton-Puffer 2010; Luiz de Zarobe 2010). However, writing, too, is a significant aspect of (second) language competence. The general writing competence of L2-taught students has been tested previously (Haunold 2006, Lasagabaster 2008, Verspoor et al. 2010), however, relatively few studies focus on expressing *content* knowledge in the L2. As was discussed in the previous chapter, studies that did focus on students' competence in expressing content knowledge tend to use prompts asking students to write essays on certain topics discussed in class (Llinares & Whittaker 2010; Whittaker & Llinares 2009). In the current study, the focus is on another aspect of writing that is part of everyday educational practice: the testing of written definitions, acquired during studying a certain topic.

Therefore, because relatively few studies have elicited written content knowledge (certainly in the Netherlands), and since those that did focused on other aspects of written L2 competence, the choice was made to focus on writing as the aspect of language competence that will be elicited.

This choice was also made with practical and time-related reasons in mind, however, as mentioned above. Because the scope of this thesis is much smaller than Gablasova's (2012) study, certain aspects needed to be adapted in order to make the current study feasible. As oral elicitation requires the researcher to test each participant individually, and necessitates the time-consuming task of transcribing all the oral data acquired, written elicitation provides a more practical and achievable method of collecting the necessary data. Furthermore, it is argued that written testing of definitions is in fact more similar to common educational practice rather than asking participants to express their

acquired content knowledge orally.

The Post-Test was considered by Gablasova to be the core measure of her study, and was used as the primary source of data in answering her research questions (2012:85). However, Gablasova (2012) also used four other data sets. One of these comprises the proficiency tests, which will be discussed in the next section. The other three were an extended definition task, a word associations task, and a delayed post-test. These four data sets were regarded by Gablasova (2012) as complementary, only to be used when necessary (e.g. in order to pursue an issue that could become apparent from results on the Post-Test). Therefore, these latter three tests were not used here, keeping in mind the limited amount of time that was available to conduct the current study and the extensive analysis and testing time required to complete and use these complementary tests.

Out of these three tests, the delayed post-test would have been interesting to use in order to test the amount of content information students could still recall at a later date. Hulstijn noted that studies that report experiments incorporating vocabulary learning that have only used immediate tests but no delayed tests are “often met with skepticism from teachers as well as researchers” (2003:371) regarding the validity of such experiments for bilingual education. This skepticism, though, mainly relates to direct reproductive tasks rather than the task in Gablasova’s (2012) study and in the current study, which involves participants’ self-formulated definitions based on the knowledge that was available in the text. Furthermore, according to Hulstijn, this skepticism of the use of immediate testing only rather than the incorporation of delayed post-testing is not justified. He argues that long-term retention of information, such as newly learned vocabulary, generally requires more repetition or exposure. When new terminology is only presented during a testing session, only immediate post-testing is necessary, since delayed post-tests would not be able to determine what cognitive processing effects have affected a participant’s performance (2003: 372).

3.3.2 *Instruments used to answer Research Question 3*

3.3.2.1 *Proficiency testing*

In order to answer the third research question, which focuses on the role students’ L2 proficiency plays in learning and recalling subject information, Meara’s (2005, 2006) X_Lex and Y_Lex tests were used to test participants’ receptive vocabulary size. These tests only take between five and ten minutes to administer and are very easy to use, making them particularly useful for quickly estimating students’ L2 skills. Previous work on CLIL has often seen the use of these tests (e.g. Admiraal, Westhoff and De Bot 2006; Goris, Denessen and Verhoeven 2013).

Both X_Lex and Y_Lex tests present words (one at a time) to the participant, without any context, asking them to indicate whether they know the meaning of the word or not (yes/no). For each test, the words that are presented are selected from five different frequency bands. The Y_Lex is a more

advanced version of the X_Lex, designed for more advanced L2 speakers. X_Lex tests vocabulary in the 0-5000 word range, whereas the words that are part of Y_Lex are in the 6000-10000 word range.

Each test contains pseudo words besides real words. These imaginary words are similar to English words in their form, but do not exist, and are used to measure the reliability of participants' answers. When participants claim to know the meaning of imaginary words, their score on the tests is adjusted downwards, relative to the number of times this happens.

Gablasova (2012) also used a C-test as another measure of L2 proficiency, besides Meara's (2005, 2006) X_Lex and Y_Lex. Although this test is a well-known and respected measure of general proficiency, it was not part of the current study due to the time it takes to administer and score.

3.4 Procedure

In order to determine the amount of time the participants should be allotted for reading the text and answering the questions in the Post-Test, 3 vwo students in the bilingual stream at another school were asked to read the text carefully and answer the questions. By pre-testing these students, a general indication of the amount of time the testing session would take was obtained.

Contrary to Gablasova (2012), who tested each participant individually, the participants in this study were tested in groups. This was done primarily to ensure that teachers would not 'lose' too much time for their regular lessons, as testing was done during regular school hours.

Each group (see Table 1 for the division of participants into groups) was first given an instruction by the researcher of the testing session that was about to follow. They were given a global idea of what the study was about, but were not told that the focus was on learning content words specifically. All students gave consent for participation in this project, and all of their data were processed anonymously, as they were also told. Then, the participants were asked to read the text carefully, keeping in mind that they had to answer some questions about its contents later. All participants were given enough time to finish reading the text. Then, the audio version of the text was played to the whole group. Following this, each group was given a short break. The researcher indicated that the students could relax for a few minutes, and asked them questions about their current reading materials and projects they were completing for their English classes. After this break, all students had to complete the Post-Test. This test was handed out to each participant on paper, after they had handed in their copy of the text. Students sitting next to each other were given different versions of the Post-Test. One of them would receive the test starting with the English language block, while the other received the test that started off with the Dutch language block. This way, the risk of students cheating by checking their classmate's answers was minimized. The participants were explicitly asked to answer questions in the same language that they were asked in.

The proficiency testing was done at another time. Each group was tested in the computer

classroom at the participating school. After receiving an oral instruction by the researcher, participants had to complete the X_Lex and Y_Lex tests individually on a computer.

Unfortunately, only the two 3 vwo groups (group 1 and group 2) were able to participate in the proficiency testing session. This will still allow for an analysis of the role of L2 proficiency on the learning and recalling of content words both when instructed in the L1 and the L2, but only for students in 3 vwo.

In Table 3 below, the order in which the testing was carried out as well as the duration of each component is laid out.

Table 3. Testing procedure.

Time	Testing session		Duration	Language of testing session	
				L1 groups	L2 groups
Testing day 1	1.	Reading text	~15 minutes	Dutch	English
	2.	Listening to audio version	~7 minutes	Dutch	English
	3.	<i>Break</i>	5 minutes		
	4.	Post-Test	~20 minutes	Dutch/English	Dutch/English
Testing day 2	5.	Proficiency testing	5-10 minutes	English	English

3.5 Analyses

The three sections below discuss the way the data were coded and analyzed in order to answer the three research questions stated in chapter 2. For ease of reading, each of these are repeated in the relevant sections. Besides the statistical analyses discussed in each section, it should be noted that effect sizes were also calculated for some of the results, but only when the difference between the participant groups were found to be statistically significant.

3.5.1 Coding and analysis of Research Question 1

This section goes into the way the data were coded and analyzed to answer the first research question: *What is the difference in the level of understanding and expressing of the meanings of new content words learned through L1 Dutch and through L2 English by grade 9 students in Dutch CLIL schools?*

In order to answer this question, three aspects of the elicited definitions were analyzed. These, like in Gablasova (2012:92), are stated here as sub questions (though rephrased). All methods of analysis that are discussed for each sub question were based on Gablasova's (2012) work, though not all measures and analyses performed by her were used.

1a *Were the students able to recall the definitions of the target words?*

This is a simple yes/no question. First, all the relevant data were typed out for each student in group 1 and 2 (so only the questions in the Post-Test asking for definitions of target words in the text). The answers given were then divided into two categories: they either showed at least some learning of the meaning of the target word, or they did not show any learning. Definitions that included mostly correct information but also some false information were still grouped in the first category (e.g. they included information that was not mentioned in the text and did not in fact relate to the meaning of the target word, see Example 1). Of course, fully correct definitions were also part of this category (Example 2). The second category contained completely incorrect definitions (Example 3) or those instances when no answer was given.

Example 1) *“Een soort dorp om geweld in groepen tegen te gaan”* (“A sort of village to combat violence in groups”) – definition of *pa* by participant 3.21

Example 2) *“A tatoo that defines what your rank is in a tribe. Men wore it on the face, buttocks and thighs, woman on lips and cheeks.”* – definition of *moko* by participant 1.3

Example 3) *“Ecocentrism is when other people came to New Zealand and changed the society. Eg- Christianity”* – definition of *ecocentrism* by participant 1.14

Participants’ answers that fit the first category were awarded 1 point, the answers in the second category were awarded 0 points. The total ‘successful learning’ score for each group was then obtained by adding up the scores of all participants, and calculating the mean score. A second coder coded all answers to examine the reliability of the scores that were awarded to the participants’ answers. Cronbach’s kappa was used to measure the inter-rater reliability, and produced a kappa of .826, which indicates a high level of agreement between the two raters.

Following Gablasova (2012:97), an independent samples t-test was used for an overall comparison of the two grade 9 groups, who differed in terms of the language that they had read the text in. A further analysis was carried out comparing the total scores of the students in group 1 and group 2 in terms of the language they were asked to answer questions in. The results of these analyses can be found in chapter 4 (section 4.1, 1a).

1b *How many meaning components were recalled?*

Following Gablasova (2012: 105), the answers that were included in category 1 for the first sub question were further analyzed in terms of the number of correct meaning components that were recalled, or in other words, how much information was (correctly) recalled from what was presented in the text. The total number of meaning components expressed by the two participant groups is compared in order to answer this sub question.

First, a list was made that contained all possible meaning components for each target word, based on the answers that were given by the participants, the information that was available in the text and the core meaning components listed by Gablasova (2012:250). Synonyms and paraphrases that were used by students were considered equal to the original meaning components in the text, and therefore were also seen as possible meaning components. Following Gablasova (2012: 99), for each target word, a distinction was made between core meaning components and minor meaning components. Core meaning components are the definitional components that are “critical for understanding the word and for providing sufficiently limiting information about its meaning” (Gablasova 2012: 99), whereas minor meaning components are definitional features that make a definition of a word more precise, but are not essential in conveying the meaning of the word (Gablasova (2012). This distinction was crucial for a fair assessment of the definitions provided by the students, as some may have given short and succinct answers, whereas others may have provided more comprehensive definitions. Taking into account all possible meaning components in the appraisal of the definitions would mean that the students that employed the latter strategy would score much higher overall, whereas the other students, while providing correct definitions, would receive lower scores since they only used core meaning components. Since a good definition, as argued by Gablasova (2012: 100), does not need to include all possible components associated with a word, the distinction between core and minor meaning components was also made in the current study.

While Gablasova (2012:250) has provided a list of core meaning components for the target words used in her study, the division between core and minor meaning components had to be made for the target words that were added for the current study. Following her example, this division was made based on information available in the Oxford English Dictionary (Simpson 1989) and Te Ara: The Encyclopedia of New Zealand (Phillips 2005). A maximum of 11 meaning components were identified per target word, with 2 to 4 of these being core meaning components.

Then, for each of the definitions that were (partly) recalled by the participants, the number of

core meaning components used was noted, including the total number of core meaning components that could have been used based on the aforementioned list.

A mean number of core meaning components that were recalled for each language (Dutch and English, as all participants had to answer questions in both languages) for each group (group 1 and 2) was calculated by first calculating, for each participant and each target word, the percentage of core meaning components recalled. For instance, if for a definition of a given target word four possible core meaning components could be used and two were actually used, 50% of possible core meaning components were used in this definition. A second coder again also coded all answers in order to examine the inter-rater reliability. Cronbach's kappa was calculated and produced a kappa of .912, indicating an excellent level of agreement between the two raters. Then, an overall mean standardized score was calculated by taking the average percentage of meaning components used for each language and participant group. An independent-samples t-test was then used to compare the two groups. The results of this analysis can be found in chapter 4 (section 4.1, 1b).

1c *What was the semantic quality of the definitions given?*

For this sub question, again, only the answers that correctly included at least some of the meaning of the target words were further analyzed. This analysis focuses on the overall content quality of the definitions.

All of the definitions that were provided by the participants were coded on a three-point scale. Table 4 below was adapted from Gablasova's table 26 (2012:110), and shows the way the answers were coded specifically.

Table 4. Coding scale for the semantic content quality of the definitions.

Score	Category	Description of the category	Example
3 points	Adequate definition	Sufficiently precise definition – more than 50% of core meaning components included ⁹	Moko: <i>“A tatoo that defines what your rank is in a tribe. Men wore it on the face, buttocks and thighs, woman on lips and cheeks.”</i>

⁹ In the case of target words for which only two core meaning components were identified, at least 1 of these had to be used for a definition to be classified as an adequate definition (i.e. target words *whanau* and *haka*).

2 points	Partially adequate definition	<p>a. Omission of an important meaning component, making the definition too narrow or too broad</p> <p>b. Incorrect information included</p>	<p>a. Kumara: <i>“plant from polynesia.”</i></p> <p>b. Pa: <i>“Een soort dorp om geweld in groepen tegen te gaan”</i> (“A sort of village to combat violence in groups”)</p>
1 point	Insufficiently adequate definition	<p>a. Includes a correct keyword/phrase from the original definition, but otherwise incorrect</p> <p>b. Vague</p>	<p>a. <i>“Ecocentrism is when other people came to New Zealand and changed the society. Eg- Christianity”</i></p> <p>b. <i>“Wooden things”</i></p>

The overall quality of the target words for both the L1 and the L2 (language of test questions) between the two groups was compared using a chi-square test, as the data here were at a nominal level. The results of these analyses can be found in chapter 4 (section 4.1, 1c).

The first research question will be answered based on all the analyses carried out for each of the sub questions discussed above.

3.5.2 Coding and analysis of Research Question 2

This section goes into the way the data was coded and analyzed to answer the second research question: *What is the difference in the level of understanding and the expressing of the meanings of new content words learned through L1 Dutch and through L2 English by grade 10 students in Dutch CLIL schools, and how do these differences relate to the differences found in grade 9 students?*

The first part of this research question is nearly identical to the first research question, only here it concerns the responses provided by the participants in groups 3 and 4: students in grade 10, rather than the participants in grade 9. Therefore, the same analyses that were discussed in section 3.5.1 will be carried out, this time using the data that was gathered in groups 3 and 4. These analyses can be found in chapter 4, section 4.2.1.

The second part of the question posed here requires a comparison of the data gathered and analyzed for the grade 9 students and the grade 10 students.

3.5.2.1 *Was there any difference between the students in grade 9 and 10 in terms of total recall of the definitions?*

First, the results of the grade 9 groups in terms of the total number of (partly) correct definitions that were recalled were compared to those of grade 10 (see section 3.5.1 1a). This was done with a focus on the language of instruction (i.e. the language of the text participants read). To do this, independent samples t-tests were used to compare the results between group 1 and group 3, and the results between group 2 and group 4. This way, it was possible to find out whether there was any significant difference between students in grade 9 and 10, when looking at the language of instruction, in terms of the number of target words that were recalled and defined correctly (or partially correctly). A more in-depth comparison between groups 1 and 3 and groups 2 and 4 in terms of the language of response (L1 or L2) was also carried out. The results of these statistical analyses can be found in chapter 4, section 4.2.2.1.

3.5.2.2 *Was there any difference between students in grade 9 and 10 in terms of the number of meaning components recalled?*

Several comparisons between the groups were made in order to answer this question. First off, as a general measure, an independent samples t-test was used to compare the mean overall number of core meaning components recalled between students in grade 9 and grade 10, taking together groups 1 and 2, and groups 3 and 4, regardless of language of instruction and language of response. This was done to get an overall feel of the data: was there any difference to begin with between the students in the two grades in terms of the amount of information that was recalled?

Then, four different independent samples t-test were carried out, comparing the following results:

- Mean number of core meaning components recalled between group 1 and group 3 for the questions answered in the L1 Dutch
- Mean number of core meaning components recalled between group 1 and group 3 for the questions answered in the L2 English
- Mean number of core meaning components recalled between group 2 and group 4 for the questions answered in the L1 Dutch
- Mean number of core meaning components recalled between group 2 and group 4 for the questions answered in the L2 English

This way, the amount of information recalled is compared between grade 9 and grade 10 in terms of the language of instruction and the language the questions were posed and answered in.

Following this, two independent samples t-tests were used to compare the effect of language of instruction overall between students in grade 9 and grade 10. In order to do this, the mean number of core meaning components were added together for groups 1 and 3, and for groups 2 and 4 (regardless of the language the questions were asked in).

Finally, the effect of the language the questions were posed and answered in on the amount of information that was recalled was compared between students in grade 9 and grade 10. To do this, the mean number of core meaning components recalled by students in grade 9 were added together and averaged in terms of the language the question was asked in, regardless of the language of instruction. The same was done for students in grade 10. Two independent samples t-tests were used: one to compare the amount of information recalled between students in grade 9 and 10 for the questions asked and answered in L1 Dutch, and one to compare the amount of information recalled between the two grades when the questions were asked and answered in L2 English.

All results can be found in chapter 4, section 4.2.2.2.

3.5.2.3 *Was there any difference between students in grade 9 and 10 in terms of the quality of the definitions?*

In order to analyze the performance of grade 9 and grade 10 students, the overall quality of the definitions provided by students in both grades were compared in four different ways, using independent samples t-tests:

- Overall mean quality score between group 1 and group 3 for the questions asked in the L1 Dutch
- Overall mean quality score between group 1 and group 3 for the questions asked in the L2 English
- Overall mean quality score between group 2 and group 4 for the questions asked in the L1 Dutch
- Overall mean quality score between group 2 and group 4 for the questions asked in the L2 English

This way, the overall quality of the definitions is compared between grade 9 and grade 10 in terms of the language of instruction and the language the questions were posed and answered in. The results can be found in chapter 4, section 4.2.2.3.

3.5.3 Coding and analysis of Research Question 3

This section describes the way the obtained data was analyzed in order to answer the third research question: *What is the effect of L2 English proficiency on the ability of students in grade 9 and 10 in Dutch CLIL schools to learn and define the meanings of newly acquired content words?*

As was discussed in section 3.4, unfortunately, only proficiency data for students in grade 9 (groups 1 and 2) could be collected. Therefore, only the relevant data analyzed with regards to the first research question will be discussed here, with respect to the L2 proficiency results gathered for these students.

First, all participants' scores on the X_Lex and Y_Lex were added up. Both tests provided a raw score, the number of words that a participant indicated they knew the meaning of, and the corrected score, which adjusts the raw score depending on the number of pseudo words the participant claimed to know. The corrected score for both tests was used here. A combined proficiency score was calculated by adding the corrected score on the X_Lex to the corrected score on the Y_Lex, as proposed by Meara and Miralpeix (2006).

Following Gablasova (2012:131), various correlations were carried out to find out if any relationship existed between the students' recall of the target words and their L2 proficiency. The first of these was an analysis of the relationship between L2 proficiency and the number of target words that were recalled and defined by the participants, with respect to both the language they recalled the definition in and the language of instruction (i.e. the group they were in).

Then, students' L2 proficiency was correlated to the mean number of core meaning components recalled in both the L1 and the L2, by each group (see section 3.5.1, sub question 1b).

Finally, the relationship between L2 proficiency and overall semantic content quality was also examined, again both for recall in the L1 and the L2 and for each group (see section 3.5.1, sub question 1c).

All correlational analyses can be found in chapter 4, section 4.3.

4. Results

4.1 Analysis of data for Research Question 1

The results for the first research question, which focuses on the difference between the level of understanding and expressing the meaning of the target words by the two grade 9 groups that were tested, are discussed below. Each aspect that was analyzed (see section 3.5.1) will be addressed separately. For group sizes smaller than 25 participants, the data were checked before carrying out each analysis to ascertain there was a normal distribution (Baarda, de Goede and van Dijkum 2007:147).

1a *Were the students able to recall the definitions of the target words?*

An independent samples t-test was carried out to compare the two groups in terms of the total number of definitions of target words that were (correctly) recalled overall, so in both the L1 and the L2. The results of this test can be found in Table 5.

Table 5. *Between-group comparison for grade 9 students of the mean number of definitions recalled overall.*

Group	Mean (max = 13)	SD	t	df	Sig.	d
L2-instructed (group 1) (N=18)	8.39	2.09	2.50	43	0.016	0.76
L1-instructed (group 2) (N=27)	6.56	2.59				

As can be seen from this table, the L2 Group recalled significantly more correct definitions when compared to the students in the L1 Group.

Besides this overall comparison between the two grade 9 groups, another independent samples t-test was carried out to analyze the results in more detail. In Table 6, the mean number of target words that were recalled and defined can be found broken down into recall in the L1 and recall in the L2.

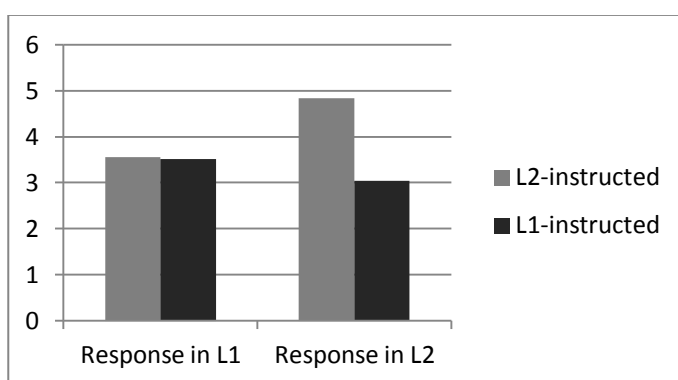
Table 6. *Between-group comparison for grade 9 students of the mean number of definitions recalled in L1 and L2.*

	L2-instructed (group 1) (N=18)		L1-instructed (group 2) (N=27)		t	df	Sig.	d
	Mean	SD	Mean	SD				
Response in L1 (Dutch)	3.56	1.50	3.52	1.37	0.09	43	0.93	-
Response in L2 (English)	4.83	1.43	3.04	1.63	3.81	43	0.000	1.16

The results in Table 6 show that while there was no significant difference between the responses in the L1 for both groups, the mean number of definitions correctly recalled in the L2 by group 1, who were also instructed in the L2, was significantly higher than that of the L1 Group. So, the overall greater performance by students in the L2 Group in terms of total recall that could be seen in Table 5 was due to their superior performance on the English part of the Post-Test.

The graph in Figure 1 below shows a representation of the difference found between the two grade 9 groups.

Figure 1. Graph showing the mean number of definitions recalled for L1- and L2 responses of students in grade 9.



1b *How many meaning components were recalled?*

An independent samples t-test comparing the mean percentage of core meaning components recalled for each group showed that on average, students in the L2 Group used more meaning components in their correct definitions than students in the L1 Group. This difference, however, was not significant, as can be seen in Table 7. Students who read the English text used on average 52% of the core meaning components that could be used based on the information in the text. Students in the L1 Group used 51% of these meaning components.

Table 7. Within-group comparison for grade 9 students of the mean percentage of core meaning components recalled overall.

Group	Mean	SD	t	df	Sig.
L2-instructed (group 1) (N=18)	0.52	0.11	0.30	43	0.766
L1-instructed (group 2) (N=27)	0.51	0.10			

1c *What was the semantic quality of the definitions given?*

A chi-square test was used to analyze the overall content quality of the correct definitions provided by students in the L1 and L2 group. The descriptive statistics can be found in Table 8, the results of the

chi-square test are reported in Table 9 below. In order to provide a more visual illustration of the content scores awarded to the responses provided by the grade 9 students in the L1 and the L2 Group, Figure 2, showing the percentages of each score given to the definitions, is also included.

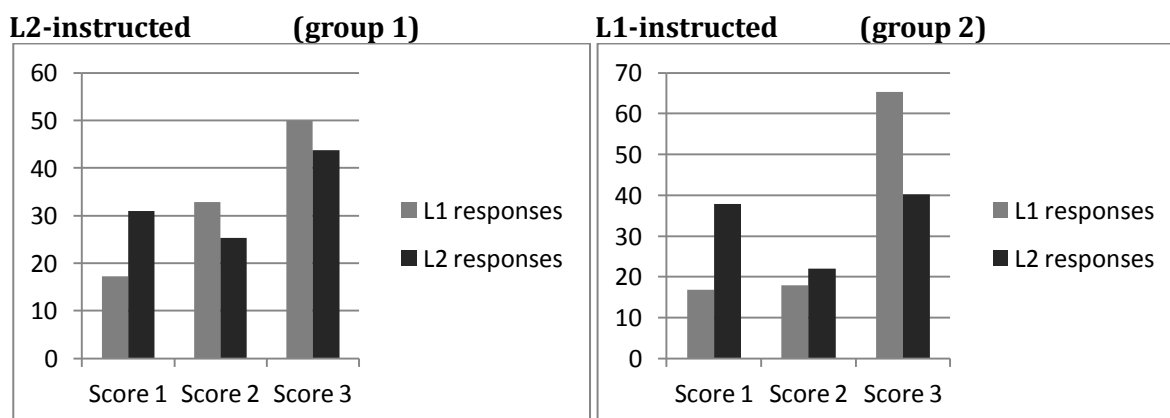
Table 8. Between-group comparison for grade 9 students of the semantic content quality of the definitions in L1 and L2 (descriptive statistics).

	L2-instructed (group 1) Score			L1-instructed (group 2) Score		
	1	2	3	1	2	3
Response in L1 (Dutch)	11 (17.2%)	21 (32.8%)	32 (50%)	16 (16.8%)	17 (17.9%)	62 (65.3%)
Response in L2 (English)	27 (31.0%)	22 (25.3%)	38 (43.7%)	31 (37.8%)	18 (22.0%)	33 (40.2%)

Table 9. Within-group comparison for grade 9 students of the semantic content quality of the definitions in L1 and L2 (chi-square analysis).

Group	Chi-square	Df	Sig.	Cramer's V
L2-instructed (group 1) (N=17)	3.861	2	0.145	-
L1-instructed (group 2) (N=25)	12.783	2	0.002	0.27

Figure 2. Graph showing the semantic content quality of provided L1 and L2 definitions for L2- and L1-instructed students.



For the students who had read the English text (group 1), no significant difference was found for the content quality of their answers when taking into account the language they had provided these answers in. However, the answers of the students who had read the Dutch text (group 2) were shown to differ significantly in terms of semantic quality content between the two languages used (see Table 9). Their Dutch definitions were scored significantly higher on content quality compared to their English definitions.

4.2 Analysis of data for Research Question 2

This research question first focuses on the definitions provided by students in grade 10 (groups 3 and 4), and secondly compares their results to the results of the participants in grade 9 (groups 1 and 2). First, each aspect of the provided definitions by grade 10 students that was analyzed will be discussed in section 4.2.1. The comparison between the definitions given by students in both grades follows in section 4.2.2.

4.2.1 Results of the analysis of the definitions provided by grade 10 students

1a Were the students able to recall the definitions of the target words?

The results of the independent samples t-test comparing the mean number of definitions recalled correctly by the participants in the two groups can be found in Table 10.

Table 10. Between-group comparison for grade 10 students of the mean number of definitions recalled.

Group	Mean	SD	t	df	Sig.
L2-instructed (group 3) (N=24)	6.00	2.90	-1.39	49	0.17
L1-instructed (group 4) (N=27)	7.11	2.81			

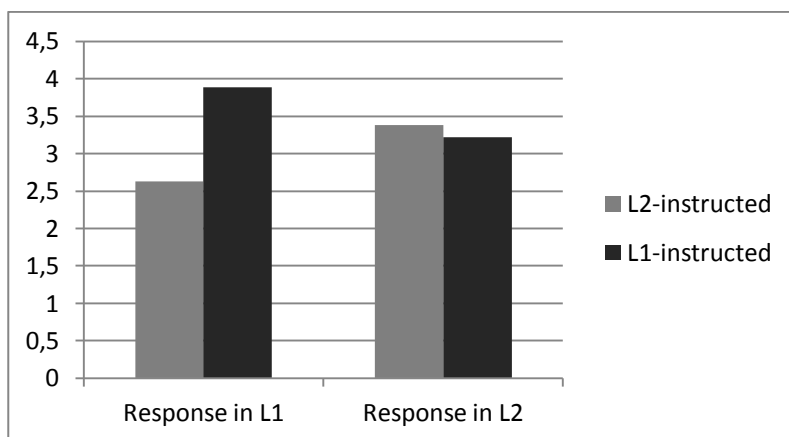
On average, the students in the L2 Group (grade 10), provided less correct definitions than the students in the L1 Group (grade 10). As can be seen in Table 10, however, this difference was not significant.

A more detailed analysis of the mean number of target words correctly recalled, however, shows that the students in the L1 Group (grade 10) actually provided significantly more L1 definitions than students in the L2 Group (grade 10) (see Table 11 below). No significant difference was found in terms of both groups' L2 responses.

Table 11. Within-group comparison for grade 10 students of the mean number of definitions recalled in L1 and L2.

	L2-instructed (group 3) (N=24)		L1-instructed (group 4) (N=27)		t	df	Sig.	d
	Mean	SD	Mean	SD				
Response in L1 (Dutch)	2.63	1.41	3.89	1.55	-3.03	49	0.004	-0.87
Response in L2 (English)	3.38	1.97	3.22	1.63	0.30	49	0.763	-

Figure 3. Graph showing the mean number of definitions recalled for L1- and L2 responses of students in grade 10.



1b *How many meaning components were recalled?*

The mean percentage of core meaning components recalled by the participants for each group was compared using an independent samples t-test. The results can be found in Table 12 below.

Table 12. *Between-group comparison for grade 10 students of the mean percentage of meaning components recalled.*

Group	Mean	SD	t	df	Sig.
L2-instructed (group 3) (N=21)	0.48	0.11	0.04	48	0.97
L1-instructed (group 4) (N=26)	0.47	0.11			

Table 12 shows that there was no significant difference between the average percentage of core meaning components used overall in the definitions provided by students who read the English text (group 3) and that of the students who read the Dutch text (group 4).

1c *What was the semantic quality of the definitions given?*

In order to analyze the overall content quality of the definitions provided by the participants in the L1 Group (grade 10) and the L2 Group (grade 10), a chi-square test was used. The descriptive statistics are reported in Table 13. The results of the chi-square test can be found in Table 14. The percentage that each quality score was awarded to the definitions provided by the students in both groups is presented in the graphs in Figure 4.

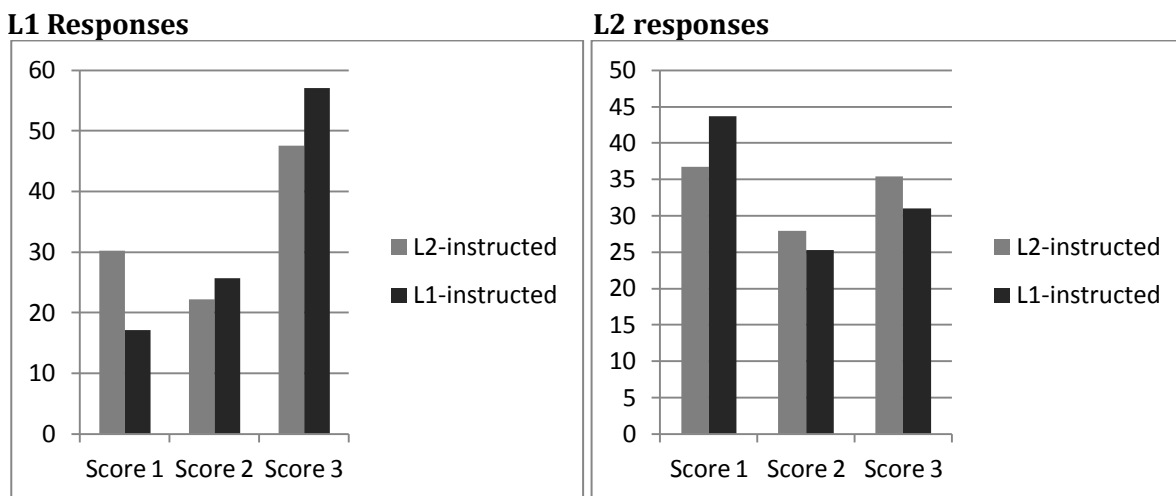
Table 13. Between-group comparison for grade 10 students of the semantic content quality of the definitions in L1 and L2 (descriptive statistics).

	L2-instructed (group 3) Score			L1-instructed (group 4) Score		
	1	2	3	1	2	3
Response in L1 (Dutch)	19 (30.2%)	14 (22.2%)	30 (47.6%)	18 (17.1%)	27 (25.7%)	60 (57.1%)
Response in L2 (English)	29 (36.7%)	22 (27.9%)	28 (35.4%)	38 (43.7%)	22 (25.3%)	27 (31.0%)

Table 14. Within-group comparison for grade 10 students of the semantic content quality of the definitions in L1 and L2 (chi-square analysis).

Group	Chi-square	Df	Sig.	Cramer's V
L2-instructed (group 3) (N=21)	2.155	2	0.341	-
L1-instructed (group 4) (N=26)	18.647	2	0.000**	0.31

Figure 4. Graph showing the semantic content quality of provided definitions for L1- and L2 responses.



The participants in the L2 Group (grade 10) showed no significant difference between the content quality of their responses in the L1 and the L2. The grade 10 students in the L1-instructed group, however, provided definitions that differed significantly in terms of their content quality. Their L1 answers were of a significantly higher quality than their L2 answers. Overall, Figure 4 also shows that the L1-instructed students scored higher in terms of content quality of their L1 definitions than the L2-instructed students. The opposite is the case for the L2 responses: here, the L2-instructed students score higher on average.

4.2.2 Results of the analyses comparing the definitions provided by grade 9 and grade 10 students

This subsection relates the results of the analyses that compared the performance of the students in grade 9 (groups 1 and 2) and the students in grade 10 (groups 3 and 4).

First, in section 4.2.2.1, the results of the comparisons between these groups in terms of the overall mean number of target words that were recalled and (correctly) defined are reported. Following this, the results of the analyses focusing on the core meaning components used in the students' answers are given in section 4.2.2.2. Finally, the analyses comparing the quality of the definitions given by students in grade 9 and grade 10 will be reported in section 4.2.2.3.

4.2.2.1 Overall recall of target words

First, grade 9 and grade 10 students that had read the English version of the text were compared in terms of the mean number of target words they correctly recalled. These results can be found in Table 15.

Table 15. Between-group comparison for L2-instructed grade 9 and grade 10 students of the mean number of definitions recalled.

Group	Mean	SD	t	df	Sig.	d
L2-instructed (group 1) (N=18)	8.39	2.09	2.96	40	0.005	0.94
L2-instructed (group 3) (N=24)	6.00	2.90				

As can be seen from the table above, on average, the grade 9 students in the L2 Group (group 1) provided significantly more correct definitions than the students in the L2 Group in grade 10 (group 3).

Following this analysis, the mean number of recalled target words by students in grade 9 and grade 10 that had read the Dutch text were also compared. The results of this independent samples t-test are reported in Table 16.

Table 16. Between-group comparison for L1-instructed grade 9 and grade 10 students of the mean number of definitions recalled.

Group	Mean	SD	t	df	Sig.
L1-instructed (group 2) (N=27)	6.56	2.59	-0.76	52	0.453
L1-instructed (group 4) (N=27)	7.11	2.81			

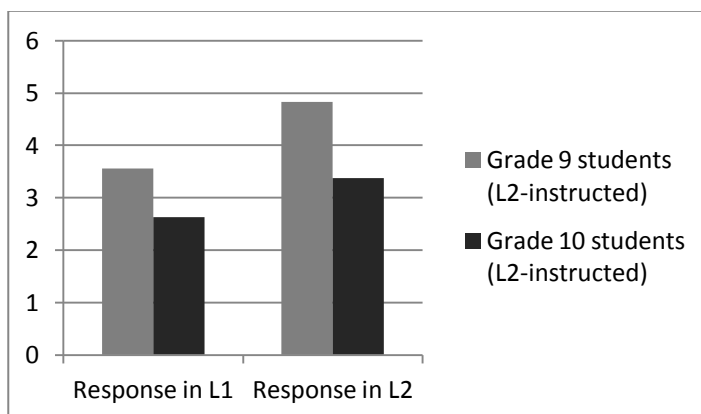
On average, L1-instructed grade 9 students were able to recall less target words correctly compared to grade 10 students. This difference was not significant, however.

A more in-depth analysis of the total recall by grade 9 and grade 10 students that had read the English text showed that the grade 9 students recalled significantly more target words both when asked to respond in their L1 and when answering in their L2 (see Table 17 and Figure 5). Furthermore, both grade 9 and grade 10 students recalled more definitions in their L2 than they did in their L1.

Table 17. Within-group comparison for L2-instructed grade 9 and grade 10 students of the mean number of definitions recalled in L1 and L2.

	Grade 9 students (group 1, L2-instructed) (N=18)		Grade 10 students (group 3, L2-instructed) (N=24)		t	df	Sig.	d
	Mean	SD	Mean	SD				
Response in L1 (Dutch)	3.56	1.50	2.63	1.41	2.06	40	0.046	0.65
Response in L2 (English)	4.83	1.43	3.38	1.98	2.66	40	0.011	0.84

Figure 5. Graph showing the mean number of definitions recalled in the L1 and L2 by L2-instructed grade 9 and grade 10 students.



A different result was obtained when comparing the L1-instructed grade 9 and grade 10 students. Although grade 10 students did recall more target words in both Dutch and English when compared to grade 9 students, these differences were not found to be significant, as can be seen in Table 18 on the next page.

Table 18. Within-group comparison for L1-instructed grade 9 and grade 10 students of the mean number of definitions recalled in L1 and L2.

	Grade 9 students (group 2, L1-instructed) (N=27)		Grade 10 students (group 4, L1-instructed) (N=27)		t	df	Sig.
	Mean	SD	Mean	SD			
Response in L1 (Dutch)	3.52	1.37	3.89	1.55	-0.93	52	0.357
Response in L2 (English)	3.04	1.63	3.22	1.63	-0.42	52	0.677

4.2.2.2 Mean percentage of meaning components recalled

To start off, the results of an independent samples t-test comparing the mean overall percentage of core meaning components recalled by students in grade 9 and grade 10 are reported in Table 19.

Table 19. Between-group comparison of grade 9 and grade 10 students in terms of the mean percentage of core meaning components recalled.

Group	Mean	SD	t	df	Sig.	d
Grade 9 (groups 1 and 2) (N=42)	0.52	0.11	1.90	93	0.060	0.39
Grade 10 (groups 3 and 4) (N=47)	0.48	0.11				

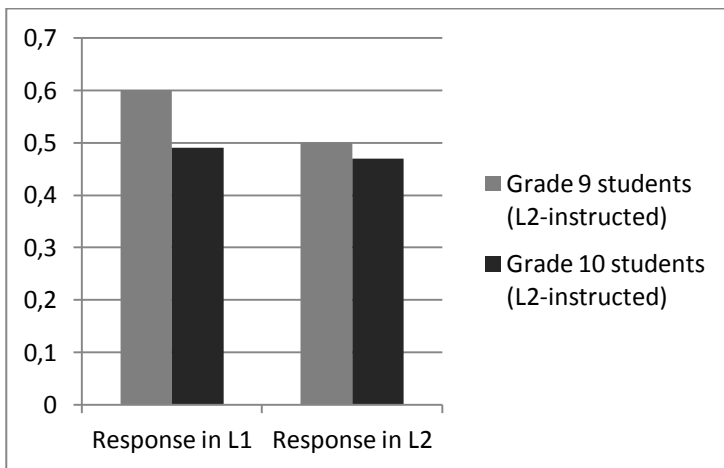
Grade 9 students recalled more core meaning components in their answers than students in grade 10. This difference was not significant, although it does approach significance.

Following this broad analysis, the students in grade 9 and grade 10 that were asked to read the English text were compared in terms of their use of meaning components in both their L1 and L2 answers. The results of this independent samples t-test can be found in Table 20 below. They are also illustrated in the graph in Figure 5.

Table 20. Within-group comparison of L2-instructed grade 9 and grade 10 students in terms of the mean percentage of core meaning components recalled in the L1 and L2.

	Grade 9 students (group 1, L2-instructed)		Grade 10 students (group 3, L2-instructed)		t	df	Sig.	d
	Mean	SD	Mean	SD				
Response in L1 (Dutch)	0.60	0.16	0.49	0.16	2.07	38	0.045	0.67
Response in L2 (English)	0.50	0.14	0.47	0.12	0.76	37	0.45	-

Figure 5. Graph showing the mean percentage of core meaning components recalled in the L1 and L2 by L2-instructed grade 9 and grade 10 students.



On average, students in grade 9 on average used more core meaning components in their Dutch answers than grade 10 students did. This difference was significant. Grade 9 students also used more core meaning components in their English answers, but not significantly more than the meaning components in grade 10 students' answers.

The definitions provided by the students in grade 9 and 10 who had read the Dutch text were also compared in terms of their recall of meaning components. The results of this analysis are reported in Table 21.

Table 21. Within-group comparison of L1-instructed grade 9 and grade 10 students in terms of the mean percentage of core meaning components recalled in the L1 and L2.

	Grade 9 students (group 2, L1-instructed)		Grade 10 students (group 4, L1-instructed)		t	df	Sig.	d
	Mean	SD	Mean	SD				
Response in L1 (Dutch)	0.60	0.11	0.55	0.11	1.75	51	0.085	(0.49)
Response in L2 (English)	0.43	0.15	0.39	0.15	0.94	50	0.35	-

Here, no significant difference was found between students in grade 9 and 10 for both L1 and L2 answers.

Then, to assess the effect of the language of instruction overall between students in grade 9 and 10, the mean percentage of core meaning components that were recalled by students who had read the English text was compared to that of students in grade 9 and 10 who had read the Dutch text. The results of this analysis can be found in Table 22 below.

Table 22. *Between-group comparison of grade 9 and grade 10 students in terms of the mean percentage of core meaning components recalled (divided by language of instruction).*

	Grade 9 students		Grade 10 students		t	df	Sig.
	Mean	SD	Mean	SD			
L2-instructed (groups 1 and 3)	0.52	0.11	0.48	0.11	1.33	39	0.189
L1-instructed (groups 2 and 4)	0.51	0.10	0.47	0.11	1.34	52	0.185

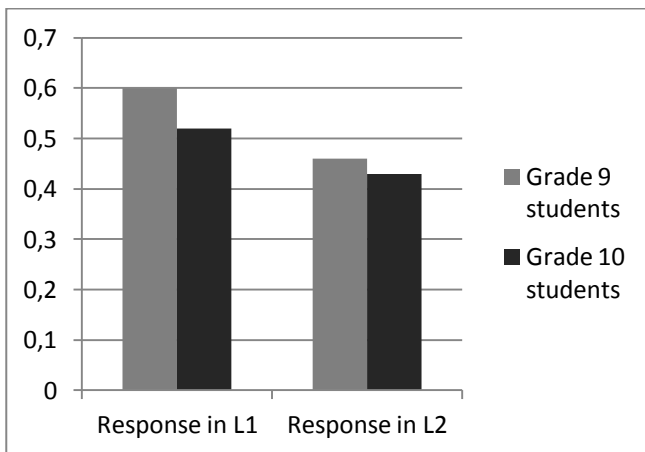
The results of this analysis show that, for the students who had read the English text, those in 9th grade on average used more core meaning components in their answers compared to students in 10th grade, although this difference was not significant. A similar picture emerged when looking at students who had read the Dutch text. Here, students in grade 9 again used more core meaning components in their definitions. However, this difference was also not found to be significant.

Finally, the overall effect of the language the questions were posed and answered in on the number of core meaning components that were recalled was assessed by two independent samples t-tests reported in Table 23, grouping together both grade 9 groups and both grade 10 groups, focusing only on the language the students had responded in. A graph illustrating the results can be found in Figure 6.

Table 23. *Between-group comparison of grade 9 and grade 10 students in terms of the mean percentage of core meaning components recalled (divided by language of response).*

	Grade 9 students		Grade 10 students		t	df	Sig.	d
	Mean	SD	Mean	SD				
L1 responses	0.60	0.13	0.52	0.14	2.81	91	0.006	0.59
L2 responses	0.46	0.15	0.43	0.14	1.13	89	0.26	-

Figure 6. Graph showing the mean percentage of core meaning components recalled in the L1 and L2 by grade 9 and grade 10 students.



Students in 9th grade used significantly more core meaning components in their Dutch answers compared to grade 10 students. 9th grade students also used more meaning components on average in their English answers, but in this case, the difference between grade 9 and 10 was not significant. Finally, the analyses comparing the quality of the definitions given by students in grade 9 and grade 10 will be reported in section 4.2.2.3.

4.2.2.3 Semantic quality of the definitions given

First, the quality of the definitions provided by the students in grade 9 and 10 that read the English text (group 1 and 3) was compared in terms of the language of response (see Table 24).

Table 24. Within-group comparison of L2-instructed grade 9 and grade 10 students in terms of the quality of the definitions given in the L1 and L2.

	Grade 9 students (group 1, L2-instructed)		Grade 10 students (group 3, L2-instructed)		t	df	Sig.
	Mean	SD	Mean	SD			
Response in L1 (Dutch)	2.33	0.76	2.17	0.87	1.06	125	0.291
Response in L2 (English)	2.13	0.86	2.00	0.86	0.95	165	0.126

Although the definitions provided by grade 9 students were scored higher on average than those given by grade 10 students, for both the L1 and the L2 responses, no significant differences were found. The same held true for the definitions provided by grade 9 and grade 10 students who read the Dutch text (group 2 and 4), as reported in Table 25 below.

Table 25. Within-group comparison of L1-instructed grade 9 and grade 10 students in terms of the quality of the definitions given in the L1 and L2.

	Grade 9 students (group 2, L1-instructed)		Grade 10 students (group 4, L1-instructed)		t	df	Sig.
	Mean	SD	Mean	SD			
Response in L1 (Dutch)	2.48	0.77	2.40	0.77	0.78	198	0.440
Response in L2 (English)	2.02	0.89	1.87	0.86	1.12	167	0.264

4.3 Analysis of data for Research Question 3

The third and final research question focuses on the possible correlation between L2 proficiency and the ability of students in CLIL schools to learn and define new content words. This section relates the correlational analyses that could be carried out. In order to be able to use the correlation coefficient r , the sample needs to have over 30 participants (Baarda, De Goede and Van Dijkum 2007: 186). As only proficiency data for the participants in grade 9 were collected, the two grade 9 groups needed to be grouped together into one big group to achieve a usable group size (N=45). As a result, it was not possible to analyze the possible correlation between L2 proficiency and the language of instruction. However, the descriptive statistics for grade 9 students' L2 proficiency, as measured by Meara's (2005) X_Lex and Y_Lex tests, are reported for each group below in Table 26.

Table 26. Descriptive L2 proficiency statistics for grade 9 students.

Group	N	Minimum score	Maximum score	Mean	SD
L1 Group	27	2500	8000	6051.85	1113.68
L2 Group	18	3900	8200	6433.33	1108.13

The correlational analyses that were possible to complete are reported below.

First, the possible relationship between the mean number of definitions that were (correctly) recalled overall for students in grade 9 was considered. There was only a very weak, non-significant positive correlation between these two variables ($r = 0.163$, $N = 45$).

Then, another analysis was carried out that correlated L2 proficiency with the number of target words that were recalled in terms of the language of recall (see Table 27).

Table 27. Correlation between mean number of definitions recalled in the L1 and L2, and L2 proficiency.

Language of response		General proficiency
Response in L1 (Dutch)	Pearson Corr.	0.079
	N	45
Response in L2 (English)	Pearson Corr.	0.171
	N	45

Again, only a very weak non-significant relationship was found between L2 proficiency as measured in this study and the definitions that were provided by the 9th grade students, both in the L1 and in the L2.

Following these analyses, the possible relationship between L2 proficiency and the percentage of core meaning components used on average by students in grade 9 was examined. The results of this analysis are reported in Table 28. A weak positive non-significant relationship between these two variables was found.

Table 28. Correlation between mean percentage of core meaning components overall and L2 proficiency.

Group	General proficiency	
Group 1 and 2	Pearson Corr.	0.291
	N	45

A second analysis further explored the possible correlation between the language of recall and L2 proficiency (see Table 29).

Table 29. Correlation between mean percentage of core meaning components recalled in the L1 and L2, and L2 proficiency.

Language of response		General proficiency
Response in L1 (Dutch)	Pearson Corr.	0.143
	N	44
Response in L2 (English)	Pearson Corr.	0.313*
	N	43

* Correlation is significant at the 0,05 level (2-tailed).

A weak correlation was found between the mean percentage of core meaning components used in English definitions by grade 9 students and their L2 proficiency. This was the only correlation that was found to be significant. Only a very weak non-significant relationship was found between L2 proficiency and the average number of core meaning components used in students' L1 responses.

Finally, the relationship between the quality scores awarded to the definitions provided and students' L2 proficiency scores was investigated. The results in Table 30 show that only a weak non-significant

relationship was found between the overall mean quality score achieved by students in grade 9 and their L2 proficiency.

Table 30. Correlation between overall mean quality score and L2 proficiency.

Group	General proficiency	
Group 1 and 2	Pearson Corr.	0.211
	N	42

A further analysis of the possible correlation between the mean quality scores of the provided definitions in the two languages and L2 proficiency found very weak (L1 responses) and weak non-significant relationships (L2 responses) (see Table 31).

Table 31. Correlation between mean quality score of definitions recalled in the L1 and L2, and L2 proficiency.

Language of response	General proficiency	
Response in L1 (Dutch)	Pearson Corr.	0.023
	N	44
Response in L2 (English)	Pearson Corr.	0.246
	N	43

5. Discussion

This chapter will first summarize the main findings that were reported in the previous chapter. Section 5.2, 5.3 and 5.4 will delve into these findings for each research question respectively, linking them to previous work discussed in chapter two, and to Gablasova's (2012) work in particular. The limitations of this study will be discussed in section 5.5.

5.1 *Summary of the results*

The first research question focused on students in Dutch grade 9 CLIL classes, and asked whether there are any differences in their level of understanding and expressing of the meanings of new content words learned through the L1 (Dutch) and L2 (English). In doing so, the language of recall was also a point of interest, with both groups of students (L1 and L2 readers) recalling half of the content words in their L1, and the other half in their L2.

The results showed that all grade 9 students were able to acquire meanings of new content words after reading the text one time, and listening to an audio version of the text while reading along. In terms of the recall of the target words overall, the L2-instructed students were shown to have an advantage, recalling significantly more definitions than their L1-instructed counterparts. When looking at these results in more detail, it was found that there was no significant difference between the L1 responses for both groups, but when asked to recall target words in the L2, the L2-instructed group recalled significantly more definitions.

The definitions that were correctly recalled by the students were further analyzed. Here, no advantage of either the L1- or L2-instructed group was found when focusing on the number of core meaning components that were recalled. There was no significant difference between the two groups. However, in terms of the quality of the answers that were given, a difference was found between the two groups. While the students that read the L2 text showed no difference in quality between their L1 and L2 answers, the L1-instructed group's L1 answers were of a significantly higher quality than their L2 answers. Interestingly, the L2 responses provided by both groups were scored lower overall than their L1 responses.

The aim of the second research question was firstly to assess whether any differences occurred in the understanding and expressing of newly acquired content words by grade 10 students in Dutch CLIL classes.

Both the L1-instructed and L2-instructed students were able to acquire and recall meanings of

new content words. The students in grade 10 who had read the L1 text were shown to be advantageous in this respect, with significantly more correct L1 definitions being provided by them when compared to the L2-instructed students. No significant difference between the two groups was found in terms of the content words recalled and defined in the L2. Both these results mark a change from the pattern that was visible for grade 9 students.

Similarly to grade 9 students, though, after analyzing the correctly provided definitions no significant difference was found between the number of core meaning components recalled by the L1- and L2-instructed participants. The results of the semantic quality analysis of the students' answers were also similar to those found for grade 9 students. No difference was found between the L1 and L2 answers provided by the group that read the L2 text, but, also similarly, a significant difference was found between the answers provided by students who were L1-instructed: they performed better in their L1, Dutch than they did in their L2.

Secondly, this research question sought to compare the results found for the students in grade 9 and 10 more closely. Some quick comparisons of the results of the students in the two grades have already been mentioned above, but further analyses provided a clearer picture of the differences and similarities between the performance of these students.

In terms of the total number of content words that were recalled by the students in the two grades, when focusing on L2-instructed students it was found that grade 9 students performed better than their grade 10 counterparts, providing significantly more correct definitions in both their L1 and L2. When looking at the results of the students that read the L1 Dutch text, the students in grade 10 tended to recall more content words overall, but not significantly so. Another interesting finding was the fact students in both grades that had read the English text recalled more definitions in their L2 than in their L1.

A closer look at the provided definitions that were correct shows that overall, grade 9 students tended to use more core meaning components in their answers than students in grade 10, but this difference was not significant. Looking at the language the definitions were provided in, it can be seen that grade 9 students that read the English text did recall significantly more core meaning components in their L1 answers than did grade 10 students. The students in grade 9 and 10 that read the Dutch text did not differ significantly in their ability to recall core meaning components for either language of response.

Both grade 9 and grade 10 students were able to recall more core meaning components in their L1 answers than in their L2 answers, regardless of the language of the text they had read.

A comparison of the semantic quality of the definitions provided by grade 9 and grade 10

students showed that grade 9 students tended to score better overall. Their L1 and L2 responses were scored higher than those by grade 10 students, for both the L1- and L2-instructed groups. These differences were not significant, however.

The aim of the third research question was to find out whether the students' L2 proficiency played a role in their ability to learn and define the meanings of new content words. Only the grade 9 students were tested. Their results in terms of the total number of content words that were recalled, the average number of core meaning components they used in their correct answers and the content quality scores that were awarded to these answers were all compared to their L2 proficiency testing results. Only one of these correlations proved to be significant: the relationship between the percentage of core meaning components used by grade 9 students in their L2 answers significantly correlated with their L2 proficiency.

5.2 *Discussion of results for Research Question 1*

The fact that the L2-instructed students in grade 9 were able to recall significantly more of the newly acquired content words in their L2 responses compared to the students who read the text in their L1, seems to signify that the students are better able to recall content knowledge in the same language they acquired this knowledge in. This is also corroborated by the fact that the L2-instructed group on average recalled more definitions in their L2 than in their L1, while the L1-instructed students provided more correct definitions in their L1 rather than in their L2. In a more recent paper Gablasova mentions the work of (amongst others) Abedi, Hofstetter and Lord (2004) and Airey (2010), who have also found that "students had problems with retrieving knowledge in their non-instructional language" (2014: 152). As discussed by Gablasova (2012), previous research on the effects of certain types of cognitive processing has found that "Transferring knowledge acquired in one context (in this case in a particular language) into a different context (into another language) creates greater demands on cognitive processing and it requires a good mastery of both linguistic codes (Baddeley 1997; Hulstijn 2003; Morris, Bransford and Franks 1977" (2012:154).

In this case, the results show that for those students who read the L1 Dutch text, it was more difficult to translate their newly acquired knowledge into their L2 English, arguably due to the fact that these students are still in the process of acquiring this language and due to greater processing demands, while the L2 text readers did not have to translate the content they learned when answering in the L2, resulting in a higher score in terms of the number of content words correctly retained and recalled.

Translating knowledge that was acquired in the L2 into the students' dominant L1 proved to be less problematic, considering the fact that no significant difference was found between the L1 responses for both groups.

While Gablasova (2012) also found that her L1- and L2-instructed students did not differ from each other with regards to the number of target words they were able to recall in their L1, she found very different results for the L2 responses. She found that the L1-instructed students recalled significantly more content words than the L2-instructed students did in their L2. While Gablasova (2012) argues that the poorer performance by L2-instructed students could be due to either difficulties during reading in the L2 (e.g. “insufficient vocabulary size or insufficiently automatic word recognition” (2012:142), or an increase in cognitive processing demands resulting in interference in the working memory capabilities with regards to processing and retaining new information, these proposed causes do not hold for the opposite result discussed here. The results that were found for this study in fact seem to contradict her findings, which seem to point towards an advantage for the L2-instructed students instead. Differences between Gablasova’s (2012) findings and the findings of this thesis could follow from the different student populations that were tested, the adaptations that were made to the testing methods (e.g. written rather than oral elicitation, and the addition of new target words), as well as the fact that CLIL education is implemented differently in Slovakia and in the Netherlands. For instance, while CLIL education in the Netherlands benefits from a national network for bilingual education and the bilingual education standard devised by them (Europees Platform 2014), nothing similar exists in Slovakia. Gondová, for instance, notes that content subject teachers in Slovakia who provide CLIL education are not trained in basic principles and teaching techniques of foreign language education (Gondová 2012), while teachers in the Netherlands that teach in a CLIL curriculum have to fulfill an extensive competency profile constructed by the Europees Platform. Furthermore, no significant language switch between grades, as is apparent in Dutch CLIL education programmes, can be found in Slovakia, even though Gablasova does mention that while most tests are conducted in the L2, content knowledge can also be assessed in the L1, “such as the secondary education leaving examination” (Gablasova 2012: 5).

A more in-depth analysis of the content words that were correctly recalled showed that there were no differences between the L1- and L2-instructed students in terms of the mean number of core meaning components that were used. The quality of their answers, however, did differ. The L2 responses provided by both groups were scored lower compared to their L1 definitions. Furthermore, while the L2-instructed group showed no significant difference in quality between their L1 and L2 responses, the L1-instructed group’s L2 answers were significantly lower in quality than their L1 responses. Again, while both groups performed worse overall in terms of the quality of their responses in the L2, it seems that it was particularly difficult for the participants to translate content knowledge well into the L2, when this knowledge was learned through the L1 (note that only correct responses were analyzed in terms of their quality – the fact that a similar ‘disadvantage’ for L1-instructed students is found when recalling knowledge in their L2 as was found when taking into account the total number of target words recalled in the L1 and in the L2 adds further weight to this finding). Here, it appears that using a

language that students are still in the process of acquiring (English) has an effect on the level of quality of the content knowledge they are able to recall and provide. The group that read the L1 text provided definitions of a significantly better quality in the L1, when they did not have to translate the knowledge they acquired into their L2. It seems that these students were more adept at providing adequate answers when using their dominant language. Furthermore, the fact that there was no significant difference between the quality of the L2-instructed students' answers seems to indicate that it was feasible for them to provide adequate answers when translating to their dominant language, just as they are capable of providing adequate definitions in the same language as the one they were instructed in.

Similarly to the results discussed in the current study, Gablasova (2012) also found no significant difference between the L1 and L2 answers given by the L2-instructed students in terms of their content quality, but she did find that the content quality of the definitions provided by the L1-readers differed significantly, also noting that their L1 definitions were of significantly higher quality compared to their L2 responses. So, there was a difference between Gablasova's (2012) study and the current one in that Gablasova found that L2-instructed students recalled significantly less content words in the L2 than L1-instructed students did, while the current study found the opposite. However, when it came to the analyses of the definitions that were scored as correct, the findings between Gablasova (2012) and this thesis are more similar.

5.3 *Discussion of results for Research Question 2*

The results for grade 10 students showed that the L1-instructed students performed best during the test, in particular on their L1 responses. This paints a completely opposite picture of what was found in grade 9. These results are more similar to those Gablasova (2012) discussed in that she also found that L1-instructed students performed better. However, in her case, there was a significant difference between the groups' L2 answers, rather than the L1 responses. Still, it could be argued that her explanation for the superior performance of L1-instructed students also holds for the results discussed here, in that the L2-instructed students' could have had difficulties with reading in the L2, or increased processing demands due to reading in the L2 may have hindered their acquisition of new content knowledge. However, the opposite result found here compared to what was found for the students in grade 9 is interesting. As the students in both grades received the same instruction for this experiment and performed the same test, the cause of this difference in overall recall of the new content words can also be sought in the language shift that occurs in Dutch CLIL schools between grade 9 and 10, particularly for History lessons and other content subjects for which the final examinations are given in Dutch, as these are taught in the L2 English in grade 9, but taught in L1

Dutch in grade 10 (and onwards).¹⁰ Another source of the differing results found between grade 9 and grade 10 students could be their individual L2 proficiency levels. Recall, for instance, that Zydatiβ (2007) argued that a limited proficiency in a second language could negatively affect L2 subject learning, and the finding reported by De Bot et al. (1997) that L2 proficiency is one of the critical factors influencing the amount of information readers can process, retain and recall. However, due to the fact that no L2 proficiency data could be collected for the students in grade 10, unfortunately this potential influencing factor cannot be further examined. Additionally, the different results could also be due to motivational differences between the students in the two grades, or due to differences in testing conditions (e.g. testing taking place at different times of the day, or on different days).

When the results of the students in grade 9 and 10 were compared directly with regards to the total number of content words that were recalled correctly, it was found that the L2-instructed students in grade 9 outperformed the L2-instructed students in grade 10. The L1-instructed students in grade 9 also performed better, but not significantly so. The fact that the students in grade 9 only significantly outperformed their grade 10 counterparts when the language of instruction was English points in the direction of a considerable advantage that students may have when more lessons are taught in English (see footnote 1), and specifically when it comes to the types of lessons taught in English. Students in grade 9 are familiar with reading and studying the types of texts that they were asked to read for this experiment in their L2 English, whereas students in grade 10 were nearly done with a year of History instruction in their L1 Dutch, as well as receiving Dutch instruction for other content subjects.

Interestingly, it should be noted that students in both grades performed the best in the language in which they had read the text: the L1-instructed students provided more correct L1 responses than L2 responses, while the L2-instructed students gave more correct answers in their L2 rather than in their L1 (see Figures 1 and 3). This, again, seems to indicate that students are best at retaining and recalling newly learned content knowledge when asked to do so in the same language in which it was required, rather than in another language. The fact that this result was found in both grades further solidifies this suggestion.

This finding goes against what was argued by some of the studies discussed in Chapter 2, which reported that participants reading a text in their L2 were found to be more likely to make errors in their understanding of the texts, or recall less information compared to their performance on texts in their L1. Donin and Silva (1993), for instance, noted that participants in their content recall study generally recalled more information from the L2 text in their L1 rather than in the L2. These results are the opposite of what was found in the current study, as discussed above. This is a very interesting

¹⁰ Note, too, that quantitatively, at least 50% of lessons in grade 9 need to be offered in English, while only 1150 out of the 4800 hours of lessons taught in grade 10-12 need to be taught in English according to the bilingual education standard (Europees Platform). This is less than 24%; a big change from the 50% of lessons taught in English in grade 9.

result, because instead of the studies that argue that there are some important drawbacks of teaching and testing knowledge in the L2, this study instead points towards a more balanced effect of language of instruction versus language of testing: the participants in the current study simply performed best when these two languages were one and the same, when we look at the total amount of knowledge (content words) they were able to recall.

A closer look at the provided definitions that were correctly recalled shows more similarities between the performance of the students in both grades. A similar pattern became obvious, as the definitions of grade 9 and grade 10 students alike did not differ with respect to the number of core meaning components recalled by L1- and L2-instructed participants. Furthermore, just as was found for grade 9 students, the only significant difference in terms of the quality of the provided answers was found between L1-instructed students' answers, where their L1 answers were of significantly better quality than their L2 answers. So, while the L2-instructed students in grade 10 did recall significantly less content words overall, the quality of the correct definitions provided by grade 10 students was similar to that of those given by students in grade 9. Also, besides the fact that the patterns of the content analyses correspond between the two grades, no significant difference was found between the students in grade 9 and 10 in terms of the semantic quality of the definitions, for either language of response.

The one difference that should be noted is that the L2-instructed students in grade 9 recalled significantly more meaning components in their L1 responses compared to the L2-instructed students in grade 10. This result was only found for the participants that read the English text. It could be that, due to the language switch between grade 9 and 10 and the lesser L2 exposure experienced by students in grade 10 in general and for this subject in particular, that their L2 reading comprehension skills deteriorated slightly compared to those of students in grade 9 (who at the time of testing, were more familiar with reading in their L2), causing the grade 10 students to be retain less L2 meaning components. Again, though, the differences could also be due to motivational differences or differences in testing conditions.

The differing findings between grade 9 and grade 10 and the possible causes listed above call for more work to be done in this area, and in the Netherlands specifically. For instance, it would be very interesting to compare grade 10 and grade 11 students in a similar manner, in order to find out whether they perform alike or whether grade 11 students instead perform more similarly to grade 9 students, which would call into question the proposed effects of the difference in L2 instruction between the grades. Furthermore, similar studies comparing grade 9 and grade 10 students in different CLIL schools in the Netherlands would also very insightful in this respect.

5.4 *Discussion of results for Research Question 3*

Despite much of the work discussed in Chapter 2 noting the influence of L2 proficiency in all manner of aspects related to the topic at hand, the proficiency data that were collected for this study only resulted in one significant correlation. This positive correlation is one that could be expected, as it states that students' L2 proficiency is related to their L2 responses, in terms of the percentage of core meaning components they used on average in these responses. Interestingly, no significant correlations were found for their L2 responses when focusing on the mean number of definitions recalled or their content quality.

Gablasova (2012) found different results, in that she not only did not find a significant correlation between the meaning components recalled by her participants and their L2 proficiency data; she also *did*, unlike the current study, find significant correlations between students' L2 proficiency and the number of definitions they correctly recalled, as well as for the semantic content quality of L1-instructed students in their L2 responses, and the semantic content quality of L2-instructed students' L1 and L2 responses.

It is important to note that the L2 proficiency data gathered by Gablasova (2012) differed from the data gathered for the current study, as discussed in chapter 3. Gablasova (2012) not only used her participants' results on Meara's (2005) X_Lex and Y_Lex vocabulary tests to make up their L2 proficiency score; she also asked them to complete a C-test made up of six texts in total, with these texts differing in terms of their difficulty level. The addition of this productive general proficiency test along with the more receptive vocabulary task developed by Meara (2005) makes for a more well-rounded and complete assessment of her participants' L2 proficiency levels. It is, in my estimation, possible that the differing results between her work and the work discussed here in terms of the influence of the L2 proficiency factor is affected by this difference. It is therefore recommended that future work on this topic includes a more elaborate L2 proficiency measure, so that its effects can be more properly studied.

5.5 *Limitations*

It is important to identify and acknowledge the limitations that exist in the current study. Some of these have already been mentioned in previous sections, such as the limited amount of L2 proficiency data that could be gathered for the participants due to time constraints. Secondly, because of the replication aspect of this study, the same text as was used by Gablasova (2012) was used. Since only one text could be used within the timeframe available for this study, more target words needed to be added. Since the text and the topic of this text were fixed, there was a limited choice of new target words, and some of them might not be typical historical terminology. However, they were chosen in a

similar fashion to Gablasova's (2012) target words, and fit the topic and story.

It is possible that the switch to using written data rather than oral elicitation data could have resulted in students being less elaborate in their answers. Oral elicitation allows the researcher to speak to each student one on one, and ask whether a given response is the full extent of what the student remembers, or whether perhaps there are more aspects of a definition that they can recall. Still, there was a good number of students who were really quite comprehensive in their written responses, and those that were not often *did* incorporate core meaning components in their answers. Given that only core meaning components were taken into account for the meaning component analyses, it is argued that any unfavorable effect of written elicitation on students' definitions being a proper reflection of the information that they recalled was avoided.

It is recommended that future work in this area further examines and compares the performance of students in grades 9 and 10 in Dutch CLIL schools to not only achieve a larger sample size that allows for a further assessment of the effect CLIL education has on content learning, but also to assess the effect that the language switch from English to Dutch for certain subjects has on students' performance more thoroughly, given the differing findings between grade 9 and 10 students in the current study and the tentative conclusions that could be drawn from them.

Of course, it would also be beneficial to the study content learning by students in CLIL streams to carry out similar studies in other countries, to supplement the data now available from Slovakian and Dutch CLIL students. Testing other populations would provide more insight into L1- and L2 learning and assessment, while also highlighting the different ways CLIL is implemented in other countries, and the effects of these different implementations.

Finally, it is reiterated that, given the importance placed on the effect of L2 proficiency on CLIL students' performance by previous studies in many aspects, future work would benefit from a more elaborate L2 proficiency measure.

5.6 *Recommendations for CLIL teaching practice*

The current study has shown that students in both grades performed best when the language of testing was the same as the language of instruction. This shows that it is very important to keep these two languages the same: when students are asked to learn content knowledge in one language, they should also be assessed in that same language in order to get the most accurate reflection of their knowledge. This may seem like an obvious statement, but as was discussed by Gablasova (2014), the language of important exams is decided by authorities regardless of students' proficiency levels or preferences. In fact, Ullman (1999) has discussed a situation in Great Britain where examination boards did not allow students, who had studied content subjects in French, to take their GCSE's in French – instead, they had to take it in English. Gablasova (2014) notes that a similar case was reported by Duff (1997). It is very

important to stress that studies, among which the current one, have found that knowledge learned in one language is not that easily transferred into the other language.

Furthermore, with regards to CLIL teaching practice in the Netherlands specifically, this study found that students in grade 9 were able to outperform students in grade 10 when the language of instruction was English. This could reflect the decreased portion of instruction received in English by students in grade 10 versus the students in grade 9. This points toward the importance of qualitatively and quantitatively excellent English input provided by teachers, especially when less input is received overall, from grade 10 and onwards.

6. Conclusion

This thesis set out to analyze the effects of bilingual education on content learning, and on the learning of new content words in particular. While previous research on CLIL education has found a positive effect on (second) language development compared to that of students following regular education programmes, few studies have looked at the effect of learning in an L2 that students are still in the process of acquiring on their ability to learn and correctly recall content knowledge. Gablasova (2012) has been one of the few researchers who has looked at this topic more thoroughly. However, since her results are not specific for an assessment of bilingual education in the Netherlands, the current study sought to replicate and adapt part of her experiment so as to assess and compare L1- and L2-mediated content learning by Dutch CLIL students.

Furthermore, while previous research carried out in Dutch contexts with a focus on subject learning has generally looked at final examination results during students' final year of CLIL education (e.g. Admiraal, Westhoff and De Bot 2006), this thesis focused on students in grade 9 and 10. This was done in order to analyze whether the language switch that occurs between these two grades with regards to certain content subjects (taught in the L2 in grade 9 and previous years, but in the L1 in grade 10 and onwards) has any discernible effect on the students' ability to learn and recall content terminology.

CLIL students were asked to read a text on a historical topic in either their L1 or their L2. Their recall of some of the content words in this text was tested in both the L1 and the L2, regardless of the language in which they were asked to read the text.

Results showed that in grade 9, students who read the L2 text were able to recall more content words than L1 readers were able to, when asked to define them in their L2. Furthermore, L2 readers did not underperform in their L1 answers when compared to the L1-instructed students, and their L1 and L2 answers were very similar in terms of quality. However, the L1 readers did perform better in their L1 answers than L2 readers did when looking at their quality. It is argued that students performed especially well when recalling content information in the same language that the

information was acquired in, but that the L1 readers particularly struggled when translating their knowledge into their L2, possibly due to the fact that they are still in the process of acquiring this language.

Contrary to the results found for students in grade 9, the L1 readers in grade 10 performed better than the L2 readers. They were able to recall significantly more words in their L1 when compared to the L2-instructed students in this grade. As students in both grades received the same test and instructions, the cause of this difference can be sought in the language shift between grade 9 and grade 10, though they could also be caused by motivational differences. Students who at time of testing were taught History in English performed significantly better in English when instructed in the L2, compared to L1-instructed students, whereas students who were taught History in Dutch outperformed L2 readers when asked to recall the content information in Dutch. When compared directly, the L2-instructed students in grade 9 were also shown to outperform the L2-instructed students in grade 10. This also points in the direction of a considerable advantage that students may have when more lessons are taught in English, and specifically content subjects such as History. Previous studies have pointed towards disadvantages encountered by students using their L2 during content learning, among which Donin and Silva (1993) and Gablasova (2012). However, it is argued here that it depends on the language in which students are instructed: when the participants of this study were instructed in the L2, they actually outperformed those who were instructed in the L1, when it came to their L2 answers. When looking at the quality of the definitions provided though, overall, the L2 answers of students in both grade 9 and grade 10 were scored lower than their L1 answers, perhaps showing the effect of using a language that they had not yet fully acquired.

7. References

- Abedi, J., Hofstetter, C.H., Lord, C. (2004). Assessment accommodations for English language learners: implications for policy-based empirical research. *Review of Educational Research*, 74 (1), 1-28.
- Admiraal, W., Westhoff, G., De Bot, C.J.L. (2006). Evaluation of bilingual secondary education in The Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, 12(1), 75-93.
- Agustín Llach, M.P. (2009). The role of Spanish L1 in the vocabulary use of content and non-content EFL learners. In: Y. Ruiz de Zarobe & R.M. Jiménez Catalán (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 112-129). Bristol: Multilingual Matters.
- Airey, J. (2010). The ability of students to explain science concepts in two languages. *Hermes-Journal of Language and Communication Studies*, 45, 35-49.
- Alderson, J.C. (1984). Reading in a foreign language: A reading problem or a language problem? In: J.C. Alderson & A.H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1-24). London: Longman.
- Baarda, D.B., De Goede, M.P.M., & Van Dijkum, C.J. (2007). *Basisboek Statistiek met SPSS: Handleiding voor het verwerken en analyseren van en rapporteren over (onderzoeks)gegevens*. Groningen: Wolters-Noordhoff.
- Bernhardt, E.B., & Kamil, M.L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, 16(1), 15-34.
- Celaya, M.L. (2008). "I study *natus* in English": lexical transfer in CLIL and regular learners". In: R. Monroy & A. Sánchez (Eds.), *25 Years of Applied Linguistics in Spain: Milestones and Challenges*. Murcia: Servicio de Publicaciones de la Universidad de Murcia.
- Cenoz, J. (2001). The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. *Cross-linguistic influences in third language acquisition: Psycholinguistic perspectives*, 111(45), 8-20.
- Coetzee-Lachmann, D. (2007). *Assessment of subject-specific task performance of bilingual geography learners: Analysing aspects of subject-specific written discourse*. Unpublished PhD thesis. Osnabrück : Universität Osnabrück.
- Cumming, A. (1987). Decision making and text representation in ESL writing performance. Paper presented at the 21st Annual TESOL Convention, Miami, April.
- Dalton-Puffer, C. (2005). Negotiating interpersonal meanings in naturalistic classroom discourse: directives in Content and Language Integrated classrooms. *Journal of Pragmatics*, 37, 1275-1293.
- Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated Learning (CLIL) classrooms*. Amsterdam: John Benjamins.
- Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe. In: W. Delanoy & L. Volkman (Eds.), *Future Perspectives for English Language Teaching* (pp. 139-157). Heidelberg: Carl Winter.
- Dalton-Puffer, C. (2011). Content and Language Integrated Learning: From practice to principles? *Annual Review of Applied Linguistics*, 31(1), 182-204.
- De Bot, K., Paribakht, T.S., & Wesche, M.B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition. *Studies in Second Language Acquisition*, 19, 309-329.
- De Graaff, R. (2013). Taal om te leren: iedere docent een CLIL-docent. *Levende Talen Magazine*, 100(7), 1-13.
- Den Boon, T. & Geeraerts, D. (2006). *Van Dale: Groot Woordenboek der Nederlandse Taal: 3 Dl.*

Utrecht: Van Dale Lexicografie.

- Dewaele, J.M. (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics*, 19(4), 471-490.
- Dewaele, J.M. (2001). Activation or inhibition? The interaction of L1, L2 and L3 on the language mode continuum. In: J. Cenoz, B. Hufeisen & U. Jessner (Eds.), *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives* (pp. 69-80). Clevedon: Multilingual Matters.
- Donin, J., Graves, B., & Goyette, E. (2004). Second language text comprehension: Processing within a multilayered system. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 61(1), 53-77.
- Donin, J., & Silva, M. (1993). The relationship between first and second language reading comprehension of occupation specific texts. *Language Learning*, 43(3), 373-401.
- Duff, P. (Ed.). (1997). *Immersion in Hungary: an EFL experiment*. Cambridge: Cambridge University Press.
- Europees Platform. (2013). Tweetalig Onderwijs. Retrieved from <http://www.europeesplatform.nl/tto/scholen/>.
- Europees Platform. (2014). Tto-standaarden. Retrieved from <http://www.europeesplatform.nl/tto/het-netwerk-tto/tto-standaarden/>.
- Eurydice. (2006). Content and Language Integrated Learning (CLIL) at school in Europe. Retrieved from <http://bookshop.europa.eu/en/content-and-language-integrated-learning-clil-at-school-in-europe-pbNCX106001/>.
- Gablasova, D. (2012). *Learning and expressing technical vocabulary through the medium of L1 and L2 by Slovak-English bilingual high-school students*. Auckland: The University of Auckland.
- Gablasova, D. (2014). Issues in the assessment of bilingually educated students: expressing subject knowledge through L1 and L2. *The Language Learning Journal*, 42 (2), 151-164.
- Gondová, D. (2012). Identifying the needs of CLIL teachers at Lower and Upper-Secondary Schools. *Journal of Interdisciplinary Philology*, 3 (3), 5-22.
- Goris, J., Denessen, E., & Verhoeven, L. (2013). Effects of the Content and Language Integrated Learning approach to EFL teaching: A comparative study. *Written Language & Literacy*, 16(2).
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Grabe, W., & Stoller, F.L. (2013). *Teaching and researching: Reading*. Routledge: Abingdon.
- Grañena, G., & Celaya, M.L. (2001). Compensatory strategies in the written production of young EFL learners. *Anuari de Filologia XXIII, A*, 10, 21-32.
- Hajer, M. (2000). Creating a language-promoting classroom: Content-area teachers at work. In: J.K. Hall & L.S. Verplaetse (Eds.), *Second and Foreign Language Learning through Classroom Interaction*, (pp. 265-285). Mahwah, NJ: Lawrence Erlbaum.
- Hardin, V.B. (2001). Transfer and variation in cognitive reading strategies of Latino fourth-grade students in a late-exit bilingual program. *Bilingual Research Journal*, 25(4), 539-561.
- Hasberg, W. (2004). Historisches Lernen im bilingualen Geschichtsunterricht. In: A. Bonnet & S. Breidbach (Eds.), *Didaktiken im Dialog. Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht* (pp. 221-236). Frankfurt/M.: Lang.
- Haunold, C. (2006). *English as a Medium of Instruction in Austrian Secondary Education*. MA thesis. Vienna: University of Vienna.
- Horst, M. (2009). Developing definitional vocabulary knowledge and lexical access speed through extensive reading. In: Z. Han & N. Anderson (Eds.), *Second language reading: Research and instruction*. Ann Arbor, MI: University of Michigan Press.

- Haynes, M., & Baker, I. (1993). American and Chinese readers learning from lexical familiarization in English texts. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 130-152). Norwood, NJ: Ablex.
- Huibregtse, I. (2001). *Effecten en didactiek van tweetalig voortgezet onderwijs in Nederland*. PhD thesis. Utrecht: Universiteit Utrecht.
- Hulstijn, J.H. (2003). Incidental and intentional learning. In C.J. Doughty & M.H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Oxford: Blackwell.
- Jäppinen, A. K. (2005). Thinking and content learning of mathematics and science as cognitional development in content and language integrated learning (CLIL): Teaching through a foreign language in Finland. *Language and Education*, 19(2), 148-169.
- Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning*, 50(2), 245-309.
- Jexenflcker, S. & Dalton-Puffer, C. (2010). The CLIL differential: Comparing the writing of CLIL and non-CLIL students in higher colleges of technology. In: C. Dalton-Puffer, T. Nikula & U. Smit (Eds.), *Language Use and Language Learning in CLIL Classrooms* (pp. 169-190). Amsterdam: John Benjamins.
- Jones, S., & Tetroe, J. (1987). Composing in a second language. In: A. Matsushashi (Ed.), *Writing in real time* (pp. 34-57). Norwood: Ablex.
- Kellerman, E. (1977). Toward a characterisation of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin*, 2, 58-145.
- Koda, K. (1990). The use of L1 reading strategies in L2 reading: Effects of L1 orthographic structures on L2 phonological recoding strategies. *Studies in Second Language Acquisition*, 12(4), 393-410.
- Kops-Hagedoorn, S. (2009). *Yes or no: Do you know? Paul Meara's English as a foreign language vocabulary test as a predictor of general foreign language proficiency* (MA thesis). Groningen: University of Groningen.
- Kubota, R. (1998). An investigation of L1-L2 transfer in writing among Japanese university students: Implications for contrastive rhetoric. *Journal of Second Language Writing*, 7, 69-100.
- Langer, J., Bartolome, L., Vasquez, O., & Lucas, T. (1990). Meaning Construction in school literacy tasks: A study of bilingual students. *American Educational Research Journal*, 27(3), 427-471.
- Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal* 1, 31-42.
- Lay, N. (1982). Composing process of adult ESL learners. *TESOL quarterly*, 16, 406.
- Lee, J. & Schallert, D.L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *Tesol Quarterly*, 31(4), 713-739.
- Lessard-Clouston, M. (2006). Breadth and depth specialized vocabulary learning in theology among native and non-native English speakers. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63(2), 175-198.
- Llinares, A. & Whittaker, R. (2006). Oral and written production in social science. *Current Research on CLIL. VIEWS*, 15(3), 28-32.
- Lyster, R. (2007). *Learning and Teaching Languages Through Content: A counterbalanced approach*. Amsterdam: John Benjamins.
- Marinellie, S. A., & Johnson, C. J. (2004). Nouns and verbs: A comparison of definitional style. *Journal of Psycholinguistic Research*, 33(3), 217-235.
- Meara, P.M. (2005). *X_Lex: the Swansea Vocabulary Levels Test. v2.05*. Swansea: Lognostics.
- Meara, P.M., & Miralpeix, I. (2006). *Y_Lex: the Swansea Advanced Vocabulary Levels Test. v2.05*. Swansea: Lognostics.

- Mondria, J.A. & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics*, 12, 249-267.
- Nagy, W.E., Anderson, R.C., & Herman, P.A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 232-270.
- Nagy, W.E., Herman, P.A., & Anderson, R.A. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233-253.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.
- Navés, T., Miralpeix, I., & Celaya, M.L. (2005). Who transfer more... and what? Cross-linguistic influence in relation to school grade and language dominance in EFL. *International Journal of Multilingualism*, 2(2), 113-134.
- Nikula, T. (2007). The IRF pattern and space for interaction: Observations on EFL and CLIL classrooms. In: C. Dalton-Puffer & U. Smit (Eds.), *Empirical Perspectives on CLIL Classroom Discourse* (pp. 179-204). Frankfurt: Peter Lang.
- Odlin, T. (2003). Cross-linguistic influence. In: C. Doughty & M. Long (Eds.), *Handbook on Second Language Acquisition* (pp 436-486). Oxford: Blackwell.
- Olsen, S. (1999). Errors and compensatory strategies: A study of grammar and vocabulary in texts written by Norwegian learners of English. *System*, 27, 191-205.
- Paribakht, T.S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition. *Studies in Second Language Acquisition*, 21(02), 195-224.
- Phillips, J. (Ed.). (2005). *Te Ara: The Encyclopedia of New Zealand*: <http://www.teara.govt.nz>.
- Poullisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15(1): 36-57.
- Ruiz de Zarobe, Y. (2010). Written production and CLIL: An empirical study. In: C. Dalton-Puffer, T. Nikula & U. Smit (Eds.), *Language Use and Language Learning in CLIL Classrooms* (pp 191-210). Amsterdam: John Benjamins.
- Schouten-van Parreren, C. (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts. *AILA Review*, 6, 75-89.
- Shapson, S.M. (1984). Bilingual and multicultural education in Canada. In: S.M. Shapson & V. D'Oyley (Eds.), *Bilingual and Multicultural Education: Canadian Perspectives*, (pp 1-16). Clevedon: Multilingual Matters.
- Simpson, J. (Ed.). (1989). *The Oxford English dictionary*. Oxford: Oxford University Press.
- Snow, C.E., Cancino, H., de Temple, J., & Schley, S. (1991). Giving formal definitions: a linguistic or a metalinguistic skill? In: E. Bialystok (Ed.), *Language processing in bilingual children*. Cambridge: Cambridge University Press.
- Stohler, U. (2006). The acquisition of knowledge in bilingual learning: an empirical study on the role of language in content learning. *Views*, 15(3), 41-46.
- Ullman, M. (1999). Communicative aspects of definitions in classroom interaction: learning to define in class for first and second language learners. *Linguistics and Education*, 20, 162-144.
- Verspoor, M.H., Schuitemaker-King, J., E.M.J. Van Rein, K. de Bot, & Edelenbos, P. (2010). *Tweetalig onderwijs: vormgeving en prestaties*. Groningen.
- Vollmer, H.J., Heine, L., Troschke, R., Coetzee, D. & Küttel, V. (2006). Subject-specific competence and language use of CLIL learners: The case of geography in grade 10 of secondary schools in Germany. Presentation at ESSE8 Conference in London 2006.
- Whittaker, R., Llinares, A. & McCabe, A. (2011). Written discourse development in CLIL at secondary school. In: R. Lyster & S. Ballinger (Eds.), *Content-Based Language Teaching: Convergent Concerns Across Divergent Contexts*. Special issue of *Language Teaching Research* (pp. 343-62).
- Zydatiβ, W. (2007). *Deutsch-Englische Züge in Berlin: Eine Evaluation des bilingualen Sachfachunterrichts*

an Gymnasium. Kontext, Kompetenzen, Konsequenzen. Frankfurt: Peter Lang.
Zydatiβ, W. (2012). Linguistic thresholds in the CLIL Classroom? The threshold hypothesis revisited.
International CLIL Research Journal, 1(4), 36-46.

New Zealand History and Society Maori Lifestyle

New Zealand has a shorter human history than any other country. The precise date of settlement is a matter of debate, but the first people arrived probably around the year 1300 AD from Polynesia. They spread throughout the country and developed a distinctive culture and lifestyle.

Once Polynesians arrived in New Zealand, they had to dramatically change their lifestyle to suit the new environment. One of the biggest changes the Polynesians had to adapt to was that New Zealand was much larger and had a more temperate climate than the tropical islands they had migrated from. This meant they had to build houses in the ground instead of on wooden pillars to make them warmer, and they also had to develop much warmer clothing.

The first settlers introduced dogs and rats to the country along with many tropical plants. Due to the climate, only one of the Polynesian crops could be grown with success – “kumara”, a sweet potato of tropical origin which became the major cultivated food-crop. It was a quick maturing crop (requiring five to six months to grow) which grew particularly well in the warmer northern regions of New Zealand. Eventually, the Polynesians became a new group of people who called themselves Maori.

The first Maori settlements were mostly located around river mouths where fish and sea birds lived abundantly. New Zealand, unlike their original islands, was abundant in wild game, so the Maori switched from agriculture to hunting. One of their biggest sources of food was the moa, a large flightless bird. Moas were herbivorous, feeding mostly on leaves, grasses and berries. The birds had large skeletons, with powerful legs, feathers like emus and small heads with short beaks. The moas varied in size from the height of a turkey, to 3.7 metres high. Unfortunately this made them easy targets, and they became extinct due to over-hunting by about 1500. As a result of this, the Maori switched back to agriculture.

The social and political organisation of Maori society was based on blood ties and particularly on descent from a common ancestor. The basic unit was whanau – family group or extended family. The family group usually consisted not only of parents and children but also of grandparents and other relations, and generally numbered between 20 and 30 people. The whanau made all decisions that affected it as a group, including the choice of husbands and wives for the young people. The whanau provided the base for the sub-tribe and the tribe.

Gradually, the Maori dispersed themselves over New Zealand in different tribes, with different chiefs as leaders. The different tribes became more aggressive however, and inter-tribal warfare became much more frequent over time. This led to the introduction of the pa (a fortified village). An average pa was placed near the top of a hill or cliff and it included ditches and palisades as protection.

New Zealand eventually became divided up by tribal territories which were recognised by other tribes by predominant land features (rivers, mountains, lakes). This culture remained up until the 18th century, when Europeans came to New Zealand.

Europeans came to New Zealand in increasing numbers from the late 18th century. The technology and

diseases they brought with them destabilised Maori society. The first Europeans that came to New Zealand were whalers and traders. One of the most popular commodities the Maori were interested in trading for were muskets. As the Maori had no long-range weapons, muskets were a valuable asset to tribes. The introduction of muskets made inter-tribal wars far more dangerous, especially if it was a tribe with muskets against a tribe without.

Before the arrival of missionaries, Maori culture involved pagan customs and ecocentric views of the world. Ecocentrism is a nature-centered worldview based on the belief that all living organisms are equally important. When Europeans arrived, that all changed, and the Maori were gradually converted to Christianity. A written form of the Maori language was also created for the Maori by the missionaries, and gradually the Maori culture and lifestyle became something completely different than before.

After 1840, the Maori lost much of their land, but their population began to increase again from the late 19th century. Today, despite the attempts of the New Zealand government to create a multicultural environment that would integrate the Maori as equal members of the society, 50% of the Maori live in poor socio-economic conditions, compared to 24% of the rest of the population. The Maori people's main problems are in the following areas:

- crime (in particular robbery, violence and car thefts) – although the Maori make up to 14% of the population, they make up almost 50% of the total prison population
- lack of suitable education
- high unemployment rates
- high suicide rates
- health problems – especially related to alcoholism and drugs
- homelessness – especially in summer

Despite the pressure of the Western culture, the Maori were able to preserve their cultural identity. Some of the traditional cultural features that are still practiced by the Maori as a sign of identity today are:

1. haka – a traditional war dance performed by a group of dancers.
2. moko – this is a type of tattoo on body and face signifying the person's rank in society. Men generally received moko on their faces, buttocks and thighs. Women usually wore moko on their lips and chins.

New Zealand History and Society
Maori Lifestyle

New Zealand has a shorter human history than any other country. The precise date of settlement is a matter of debate, but the first people arrived probably around the year 1300 AD from Polynesia. They spread throughout the country and developed a distinctive culture and lifestyle.

Once Polynesians arrived in New Zealand, they had to dramatically change their lifestyle to suit the new environment. One of the biggest changes the Polynesians had to adapt to was that New Zealand was much larger and had a more temperate climate than the tropical islands they had migrated from. This meant they had to build houses in the ground instead of on wooden pillars to make them warmer, and they also had to develop much warmer clothing.

The first settlers introduced dogs and rats to the country along with many tropical plants. Due to the climate, only one of the Polynesian crops could be grown with success – “kumara”, a sweet potato of tropical origin which became the major cultivated food-crop. It was a quick maturing crop (requiring five to six months to grow) which grew particularly well in the warmer northern regions of New Zealand. Eventually, the Polynesians became a new group of people who called themselves Maori ("the local people", or "the original people").

The first Maori settlements were mostly located around river mouths where fish and sea birds lived abundantly. New Zealand, unlike their original islands, was abundant in wild game, so the Maori switched from agriculture to hunting. Maori gradually became expert hunters and fishermen. They carved fishhooks from bone and stone, and wove fishing nets from flax – a fiber found in the similarly named plant that can be used for making linen. One of their biggest sources of food was the moa, a large flightless bird. Moas were herbivorous, feeding mostly on leaves, grasses and berries. The birds had large skeletons, with powerful legs, feathers like emus and small heads with short beaks. The moas varied in size from the height of a turkey, to 3.7 metres high. Unfortunately this made them easy targets, and they became extinct due to over-hunting by about 1500. As a result of this, the Maori switched back to agriculture.

The social and political organisation of Maori society was based on blood ties and particularly on descent from a common ancestor. The basic unit was whanau – family group or extended family. The family group usually consisted not only of parents and children but also of grandparents and other relations, and generally numbered between 20 and 30 people. The whanau made all decisions that affected it as a group, including the choice of husbands and wives for the young people. The whanau provided the base for the sub-tribe and the tribe.

Gradually, the Maori dispersed themselves over New Zealand in different tribes, with different chiefs as leaders. The different tribes became more aggressive however, and inter-tribal warfare became much more frequent over time. This led to the introduction of the pa (a fortified village). An average pa was placed near the

top of a hill or cliff and it included ditches and palisades (strong fences made from wooden stakes) as protection.

New Zealand eventually became divided up by tribal territories which were recognised by other tribes by predominant land features (rivers, mountains, lakes). This culture remained up until the 18th century, when Europeans came to New Zealand. However, even today, people belonging to certain tribes are able to trace back which of the Polynesians that came to New Zealand are part of their genealogy (the ancestry and history of a person, family or group).

Europeans came to New Zealand in increasing numbers from the late 18th century. The technology and diseases they brought with them destabilised Maori society. The first Europeans that came to New Zealand were whalers and traders. One of the most popular commodities the Maori were interested in trading for were muskets – heavy, long-barreled shoulder guns, used between the 16th and 18th century by infantry soldiers. As the Maori had no long-range weapons, muskets were a valuable asset to tribes. The introduction of muskets made inter-tribal wars far more dangerous, especially if it was a tribe with muskets against a tribe without.

Before the arrival of missionaries, Maori culture involved pagan customs and ecocentric views of the world. Ecocentrism is a nature-centered worldview based on the belief that all living organisms are equally important. When Europeans arrived, that all changed, and the Maori were gradually converted to Christianity. A written form of the Maori language was also created for the Maori by the missionaries, and gradually the Maori culture and lifestyle became something completely different than before.

In 1840, the Treaty of Waitangi was signed by representatives of the British Crown and various Maori chiefs. In the Treaty, a British Governor of New Zealand was appointed, the Maori were granted ownership of their lands, and they were given the rights of British subjects. Nowadays, it is generally considered the founding document of New Zealand as a nation. However, the English and Maori versions of the Treaty differed significantly, and as there is no consensus as to what exactly was agreed to, it is often the subject of much disagreement by both Maori and non-Maori New Zealanders.

Today, despite the attempts of the New Zealand government to create a multicultural environment that would integrate the Maori as equal members of the society, 50% of the Maori live in poor socio-economic conditions, compared to 24% of the rest of the population. The Maori people's main problems are in the following areas:

- crime (in particular robbery, violence and car thefts) – although the Maori make up
- 14% of the population, they make up almost 50% of the total prison population
- lack of suitable education
- high unemployment rates
- high suicide rates
- health problems – especially related to alcoholism and drugs
- homelessness – especially in summer

Despite the pressure of the Western culture, the Maori were able to preserve their cultural identity. Some of the traditional cultural features that are still practiced by the Maori as a sign of identity today are:

1. haka – a traditional war dance performed by a group of dancers.
2. moko – this is a type of tattoo on body and face signifying the person's rank in society. Men generally received moko on their faces, buttocks and thighs. Women usually wore moko on their lips and chins.

The Maori are also known for their ceremonies. The first and most notable ceremony that still exists within Maori culture is powhiri, a welcoming ceremony performed to welcome visitors to their land. The most spectacular part of this ceremony is the wero, an aggressive challenge of the visitor. Wero means "cast a spear". During the wero, three Maori warriors advance cautiously towards the guests with traditional ceremonial weapons, performing threatening gestures and grimaces while calling out battle screams, seemingly ready to explode into violence at any moment.

De geschiedenis en maatschappij van Nieuw-Zeeland De Maori

Nieuw-Zeeland heeft een kortere menselijke geschiedenis dan welk land dan ook. Het precieze moment waarop mensen Nieuw-Zeeland ontdekten is onduidelijk, maar de eerste mensen arriveerden waarschijnlijk rond het jaar 1300 n.Chr. vanuit Polynesië. Ze verspreidden zich door het land en ontwikkelden een kenmerkende cultuur en levensstijl.

Toen de Polynesiërs aankwamen in Nieuw-Zeeland moesten ze hun levensstijl drastisch aanpassen aan de nieuwe omgeving. Eén van de grootste veranderingen voor de Polynesiërs was dat Nieuw-Zeeland veel groter was en een meer gematigd klimaat had dan de tropische eilanden waar ze vandaan waren geëmigreerd. Dit betekende dat ze huizen in de grond moesten bouwen om ze warmer te maken, in plaats van op houten palen, en ook moesten ze warmere kleding gaan maken.

De eerste migranten introduceerden honden en ratten in het land, en voerden ook vele tropische planten in. Door het klimaat kon maar één Polynesisch gewas met succes worden geteeld – de “kumara”, een zoete aardappel van tropische origine, die vervolgens het belangrijkste gekweekte voedselgewas werd. Het was een snelgroeiend gewas (het had vijf tot zes maanden nodig om te groeien), dat vooral goed groeide in de warmere noordelijke regionen van Nieuw-Zeeland. De Polynesiërs werden uiteindelijk een nieuw volk, en noemden zichzelf Maori (“de lokale bevolking”, of “de originele bevolking”).

De eerste nederzettingen van de Maori waren vooral gelegen rond riviermondingen waar vissen en zeevogels in overvloed leefden. In Nieuw-Zeeland leefden veel wilde dieren, vergeleken met hun oorspronkelijke eilanden, waardoor de Maori van landbouw op jagen overstapten. Ze werden geleidelijk steeds betere jagers en vissers. Ze kerfden vishaken uit bot en steen, en weefden visnetten van vlas – een vezel afkomstig van de gelijknamige plant, die gebruikt kan worden om linnen van te maken. Een van de grootste voedselbronnen van de Maori was de moa, een grote loopvogel. Moa's waren herbivoren die vooral bladeren, grassen en bessen aten. De vogels hadden grote skeletten, met sterke benen, veren als die van emoes en kleine koppen met korte snavels. De moa's kwamen in verschillende groottes voor, van zo groot als een kalkoen tot 3,7 meter hoog. Helaas maakte dit ze ook makkelijke doelwitten, en raakten ze rond 1500 uitgestorven door de jacht. Het gevolg hiervan was dat de Maori weer overgingen op de landbouw.

De sociale en politieke organisatie van de Maori-maatschappij was gebaseerd op bloedverwantschap en met name op afstamming van een gemeenschappelijke voorouder. De basiseenheid was de whanau – familiegroep of uitgebreide familie. De familiegroep bestond vaak niet alleen uit ouders en hun kinderen, maar ook uit grootouders en andere familieleden, en bevatte over het algemeen tussen de 20 en 30 mensen. De whanau maakte alle beslissingen die de groep aangingen, waaronder de keuze van de toekomstige man of vrouw voor de jongere mensen. De whanau was de basis voor de substam en de stam.

Langzamerhand verspreidden de Maori zich in verschillende stammen over Nieuw-Zeeland, met verschillende stamhoofden als leiders. De verschillende stammen werden echter meer agressief, en oorlogvoering tussen de stammen kwam steeds vaker voor. Dit leidde tot de introductie van de pa (een versterkt dorp). Een pa werd over het algemeen gevestigd in de buurt van de top van een heuvel of een klif, en bevatte dijken en palissades (sterke omheiningen gemaakt van houten palen) als bescherming.

Nieuw-Zeeland werd uiteindelijk opgedeeld in stamgrondgebieden door middel van belangrijke landschapkenmerken (rivieren, bergen, meren). Deze stamgrondgebieden werden erkend door de andere stammen. Deze cultuur bleef bestaan tot de Europeanen naar Nieuw-Zeeland kwamen, in de 18de eeuw. Tegenwoordig kunnen mensen die tot een bepaalde stam behoren echter nog steeds achterhalen welke van de oorspronkelijke Polynesiërs die naar Nieuw-Zeeland kwamen deel uitmaken van hun genealogie (de afkomst en geschiedenis van een persoon, familie of groep).

Vanaf het einde van de 18de eeuw kwamen steeds meer Europeanen naar Nieuw-Zeeland. De technologie en ziekten die ze met zich meebrachten destabiliseerden de Maori-samenleving. De eerste Europeanen die naar Nieuw-Zeeland kwamen waren walvisjagers en handelaren. Eén van de meest populaire goederen waar de Maori geïnteresseerd in waren, waren musketten – zware schoudergeweren met een dubbele loop, die tussen de 16de en 18de eeuw gebruikt werden door de infanterie. Aangezien de Maori geen wapens hadden voor de lange afstand waren musketten een waardevolle aanwinst voor stammen. De introductie van musketten had als gevolg dat oorlogen tussen stammen veel gevaarlijker werden, vooral wanneer het om een stam ging die over musketten beschikte tegenover een stam zonder.

Voor de komst van de missionarissen maakten heidense gewoontes en een ecocentrische visie op de aarde deel uit van de cultuur van de Maori. Ecocentrisme is een natuurgericht wereldbeeld dat gebaseerd is op het geloof dat alle levende organismen even belangrijk zijn. Toen de Europeanen arriveerden veranderde dit allemaal, en werden de Maori geleidelijk aan bekeerd tot het christendom. De missionarissen ontwikkelden ook een geschreven vorm van hun taal, het Maori, en langzamerhand werden de cultuur en levensstijl van de Maori compleet anders dan voorheen.

In 1840 werd het Verdrag van Waitangi getekend door afgevaardigden van het Britse Rijk en diverse Maori-leiders. In het verdrag werd een Britse Gouverneur van Nieuw-Zeeland benoemd, kregen de Maori het eigendom van hun grondgebieden toegekend, en kregen ze dezelfde rechten als Britse onderdanen. Tegenwoordig wordt het Verdrag van Waitangi gezien als het oprichtingsdocument van Nieuw-Zeeland als een natie. De Engelse en de Maori-versie van het verdrag verschillen echter behoorlijk, en omdat er geen consensus bestaat over wat precies is afgesproken, is het vaak het onderwerp van veel onenigheid onder zowel Maori als niet-Maori Nieuw-Zeelanders.

Ondanks de pogingen van de regering van Nieuw-Zeeland om een multiculturele samenleving te creëren die de Maori integreren als gelijkwaardige leden van de maatschappij, leven vandaag de dag 50% van de Maori in slechte sociaal-economische omstandigheden, vergeleken met 24% van de rest van de bevolking. De voornaamste problemen van de Maori worden gevonden op de volgende gebieden:

- criminaliteit (voornamelijk berovingen, geweld en autodiefstal) – hoewel de Maori 14% van de bevolking vormen, vormen ze bijna 50% van de gehele gevangenisbevolking
- gebrek aan passend onderwijs
- hoge werkloosheid
- hoge zelfmoordcijfers
- gezondheidsproblemen – vooral gerelateerd aan alcoholisme en drugs
- veel dakloosheid – vooral in de zomer

Ondanks de druk van de Westerse cultuur hebben de Maori hun culturele identiteit weten te behouden. Een paar van de traditionele culturele gewoontes die nog steeds worden uitgevoerd door de Maori als een teken van identiteit zijn:

1. de haka – een traditionele oorlogsdans uitgevoerd door een groep dansers
2. de moko – dit is een soort tattoo op het lichaam en gezicht die de rang in de maatschappij van een persoon aanduidt. Mannen hadden gewoonlijk moko's op hun gezichten, billen en dijen. Vrouwen droegen een moko doorgaans op hun lippen en kin.

De Maori staan ook bekend om hun ceremonies. De belangrijkste ceremonie die nog wordt uitgevoerd in de Maori-cultuur is de powhiri, een welkomstceremonie om bezoekers in hun land te verwelkomen. Het meest spectaculaire onderdeel van deze ceremonie is de wero, een agressieve uitdaging van de bezoeker. Wero betekent “werp een speer”. Tijdens de wero bewegen drie Maori-krijgers met traditionele ceremoniële wapens zich behoedzaam richting de gasten en maken ze bedreigende gebaren met een grimas op hun gezicht, terwijl ze strijdkreten slaken en eruitzien alsof ze elk moment in geweld uit kunnen barsten.

Appendix 4

Post-Test

1. When did the first people come to New Zealand?
2. Why is it surprising that the Maori make up 50% of the prison population in New Zealand?
3. What is a kumara?
4. Who were the first people that came to New Zealand?
5. What is a moko?
6. What did the missionaries do for the Maori?
7. How many percent of the Maori live in bad economic conditions?
8. What is ecocentrism?
9. What was one of the biggest changes that the Maori had to adapt to in New Zealand after their arrival?*
10. What is a wero?*
11. Name two problems that currently exist in the Maori population.*
12. What are palisades?*
13. What did the first people arriving in New Zealand look for when deciding where to settle?*
14. What is the Treaty of Waitangi?*
15. What did the first settlers in New Zealand bring to the country?*
16. What is flax?*

1. Tot welke eeuw bleef de Maori-populatie groeien?***
2. Wat zijn musketten?*
3. Waardoor werden de grenzen van de grondgebieden van Maori-stammen bepaald?*
4. Wat is een whanau?***
5. Wat voor soort samenleving wilde de regering van Nieuw-Zeeland creëren?***
6. In welk deel van Nieuw-Zeeland werden de eerste nederzettingen van de Maori gesticht?***
7. Wat is een pa?***
8. Welke twee negatieve zaken brachten de Europeanen naar Nieuw-Zeeland?***
9. Hoeveel procent van de Nieuw-Zeelanders zijn Maori?***
10. Wat is een moa?***
11. Wat betekent Maori?*
12. Wat is een haka?*
13. Wat deden de Maori voor ze begonnen met jagen?*
14. Wat is genealogie?*
15. Uit welke materialen maakten de Maori vishaken?*

* Newly added questions

** Questions translated from Slovak to English by Dana Gablasova in personal communication, then translated into Dutch by the researcher