

The CLISS Project: Receptive Vocabulary in CLIL versus non-CLIL Groups

LISS KERSTIN SYLVÉN
SÖLVE OHLANDER
University of Gothenburg

*“For to the one who has, more will be given, and he will have an abundance”
(Matthew 13:12)*

Abstract

Basically, Content and Language Integrated Learning, CLIL, aims at increasing language learners' exposure to a foreign language by using it as the medium of instruction when teaching ordinary school subjects, e.g. biology and history. It is nowadays a widespread educational approach in Europe and research into CLIL is attracting increasing interest. However, research on the effects of CLIL in the Swedish context is scarce. To remedy this to some extent, the large-scale, longitudinal CLISS project, focusing primarily on CLIL as well as non-CLIL students' proficiency and progress in written academic English and Swedish in upper secondary school, was launched in 2011. In this article, the CLISS project is accounted for in some detail, and the results from the first round of English receptive vocabulary test are presented. As this test, known as the Vocabulary Levels Test, was administered at the very outset of the CLIL experience for the CLIL students, these results represent baseline data. Findings reveal that already from the start, the CLIL students outperform the non-CLIL ones, and also that the males have a larger vocabulary than the females in both groups of students. Some possible reasons for these results are discussed.

Keywords: CLIL; academic language; vocabulary acquisition; second language acquisition, SLA; Vocabulary Levels Test

1. Introduction

Content and Language Integrated Learning, CLIL, is becoming increasingly widespread as an educational approach around Europe. It means, roughly, that another language – most commonly English – than students' first language (L1) is used as the medium of instruction in various school subjects. This article introduces a large-scale, longitudinal research project, *Content and Language Integration in Swedish Schools*, CLISS, which investigates, from different perspectives, with a primary focus on academic writing, the effects of the CLIL approach in Sweden, more specifically the use of English as the medium of instruction. We also account for and discuss the results obtained from the first test, within the project, of students' receptive English vocabulary. First, however, a brief background to CLIL is offered. The article ends with a discussion of CLIL in Sweden, especially in relation to the results from the first round of receptive vocabulary tests.

2. Background

The integration of language and content in educational settings is by no means a contemporary invention. On the contrary, throughout history this has been a way of imposing a majority language on citizens in colonized territories, or on linguistic minorities within a country. The Roman Empire is a case in point, where Latin was the language used by the ruling class, and thus used in educational settings throughout the empire; likewise, in Swedish schools well into the 20th century, the Sami language was banned. This kind of situation is also what many immigrant students experience on a daily basis in a new country where that country's majority language is used in school. In such cases, the use of another language than the student's L1 to teach content matter is not an optional choice made by the individual, but rather a forced situation. When, however, the immersion method was introduced in Canada in the mid 1960s, it was the result of English L1 parents' worries that their children did not learn French well enough in traditional school. Immersion was then presented as an option for English L1 children to learn French while at the same time learning the content of the school subjects concerned (Cummins, 1979; Lambert, 1977). This "two-in-one" approach was later implemented in various European contexts, under the umbrella term of Content and Language Integrated Learning, i.e. CLIL (Coyle, 2007; Dalton-Puffer, 2007; Marsh, 2000; Nikula, 1997), in many ways modelled on the Canadian immersion method.

The term CLIL subsumes all second language (L2) teaching approaches where another language than students' L1 is used as the medium of instruction, such as content-based learning, bilingual education, English-medium instruction, etc. (see also Tedick & Cammarata, 2012). Some scholars argue that immersion and CLIL are two separate approaches to the combination of content and language (Lasagabaster & Sierra, 2010), where immersion is used in a second language context with the target language readily available for learners to be exposed to also outside of the school context, whereas CLIL takes place in a foreign language (FL) context, with few opportunities to exposure outside of school. Others are of the opinion that there are so many variations of each approach that it is virtually impossible to keep them apart, and that indeed CLIL subsumes immersion (Cenoz, Genesee, & Gorter, 2013). Suffice it here to conclude that the basic rationale for CLIL, as well as for immersion, is to extend the time students are exposed to a foreign language in the "pure" language arts classes, often limited to a maximum of 2–3 hours per week, thus giving the learner more opportunities to receive input as well as produce output in the relevant language.¹

Research on CLIL has increased significantly in the last few decades, and is now to be regarded as an important field of study in its own right, not least within the larger domain of second language acquisition (SLA); cf., e.g., Ringbom

¹ In a wider historical perspective, CLIL may be seen as a latter-day offshoot of the tradition introduced by Comenius in the 17th century and referred to as the Natural Method, in which the method of learning an L2/FL mimicks that of learning one's L1 (Kelly 1969).

(2012). In a seminal study of CLIL in Austria, Dalton-Puffer (2007) illustrates the communicative benefits of using English as the medium of instruction. She emphasizes the sociocultural basis of CLIL, according to which language learning takes place in a social context. As communication, in order to learn a specific content, is in focus in a CLIL class, rather than language itself, the argument is that meaning is co-constructed among all participants in the classroom, while learning takes place (of both content and language). Nikula (2005) looks at classroom interaction in a Finnish CLIL setting, concluding that while CLIL students are viewed as competent language users in the content subject classroom, they are seen as language learners in the English as a foreign language (EFL) classroom. Findings show that in the CLIL classroom, students took active part in the interaction, as opposed to the EFL classroom, which to a larger extent was teacher-led. In a similar study in Sweden, however, Lim Falk (2008) shows that there is much less interaction in the CLIL classroom than in similar, non-CLIL classrooms. Thus, research findings are inconclusive – or even contradictory – as regards the effect of CLIL on classroom interaction.

CLIL in Spain, as well as other European countries, has been subject to extensive research. For instance, Navés and Victori (2010) found convincing evidence for a clearly beneficial effect on CLIL students' proficiency in the target language (TL), in this case English. Positive effects of CLIL have also been shown in reading comprehension (Admiraal, Westhoff, & de Bot, 2006; Navés, 2011) as well as in vocabulary knowledge (Jiménez Catalán, Ruiz de Zarobe, & Cenoz, 2006). Similar findings have been reported in studies from other contexts (Klippel, 2003; Zydatis, 2007). Furthermore, productive proficiency has been investigated in a number of studies, with various outcomes. Ruiz de Zarobe (2010) investigated, among other things, CLIL learners' productive skills, finding that they outperformed their non-CLIL peers on content, vocabulary, organization, language use and mechanics in a test of written production, and significantly so on content and vocabulary (p. 206). Writing ability is also the focus in Jexenflcker and Dalton-Puffer (2010), who show that while CLIL students perform better than non-CLIL students on general language ability and writing skills, there is no difference between the two groups as regards textual competence. Vollmer, Heine, Troschke, Coetzee, and Küttel (2006) detected no differences between CLIL and non-CLIL students in their written texts on geography, and Llinares and Whittaker (2006) claim that there is room for improvement in CLIL and non-CLIL groups alike regarding writing.

Of specific interest to the present paper are studies investigating receptive vocabulary knowledge. However, there are very few such studies, and those that exist do not report on baseline data, i.e. pre-CLIL levels of proficiency. For instance, Jiménez Catalán and Ruiz de Zarobe (2009) show that in 6th grade, female CLIL students outperform female non-CLIL students on receptive vocabulary tests. The tests were administered when the CLIL students had been taught a number of subjects in English for several years, and their estimated total time of exposure to English in school was 960 hours, compared to 629 in the non-

CLIL group. No information is given on entry-level differences between groups, however. An exception to the rule of not including baseline information is Admiraal et al. (2006), where CLIL students (N = 1,305) were reported to score higher on entry-level receptive vocabulary. After four years of CLIL instruction, this advantage remained intact, rather than growing, which would have been the expected outcome. As regards receptive vocabulary proficiency from a gender perspective, there are very few studies specifically addressing this issue. Jimenez Catalán (2010), however, analyses various English vocabulary tests performed by 12 year-old male and female students in Spain. The findings indicate no gender differences on the Vocabulary Levels Test (cf. sections 3.2 and 4 below) used to tap into receptive vocabulary proficiency among these informants.

Critical views are voiced on some of the studies on the effects of CLIL. Bruton (2011) argues that there are several issues in need of addressing before there can be any claims about the beneficial effects of CLIL. He points to four areas in particular, in response to research carried out mainly in a Spanish context. First of all, he criticizes the lack of pre-tests (where Admiraal et al [2006] is an exception), arguing that without access to details about baseline data, it is impossible to attribute any findings, positive or negative, to CLIL. Second, the lack of comparable control groups makes results difficult to substantiate. Third, Bruton (2011) brings up the very important factor of extra CLIL support, which is common in, for instance, many Spanish CLIL schools. Finally, the need to specify the amount of FL use in CLIL classes is discussed. In a similar vein, Rumlich (2013) finds that CLIL students in a German context are ahead of their non-CLIL peers even before CLIL has begun (in this case in year 7), arguing that this has not been acknowledged in previous studies showing the benefits of CLIL.

In contrast to the mainly positive findings in CLIL research in other European countries, no study in Sweden, as of yet, has been able to verify a positive effect of CLIL on students' English proficiency. Washburn (1997) showed that while CLIL students were ahead of their non-CLIL peers initially, their English proficiency was more or less on a par after two years of CLIL at upper secondary level. Sylvén (2004/2010) showed that CLIL students were indeed superior to their non-CLIL peers as regards English vocabulary proficiency. However, this was the situation already from the start, and rather than the CLIL approach in itself, it seems as though it was the amount of exposure to English (EE) outside of school that was decisive as regards the progress in proficiency observed during the two years of the study. Interestingly, though, in studies on CLIL in a Swedish context with German as the TL, very promising results have been found (Dentler, 2002; Terlevic Johansson, 2011). There are several likely explanations for the disparity in results, depending on which TL is used as the medium of instruction; however, delving into them is beyond the scope of the present article (for some discussion, see Sylvén 2013).

In this necessarily brief overview of research findings concerning CLIL, results obviously differ a great deal depending on various factors (cf., e.g., Sylvén 2013). One crucial factor to control for in any CLIL research is the national context.

Consequently, research targeting CLIL from a plethora of perspectives within one and the same country is much needed in order to shed some further light on the overall framework, implementation and effects of CLIL.

3. The CLISS project

In 2011, the research project *Content and Language Integration in Swedish Schools*, CLISS, was launched.² The CLISS project came about primarily as a result of the relative scarcity of studies focusing on CLIL in a Swedish context (but cf. Alvtörn 2000, Lim Falk 2002, 2008; Kjellén Simes 2008, Sylvén 2004/2010, 2007, 2013). It is a large-scale, longitudinal and multi-perspective investigation into CLIL as implemented in Sweden. Below follows a fairly detailed account of the project.

3.1 Overall description

The CLISS project aims at illuminating the role of the language of instruction, English, in the development of different academic language competencies among upper secondary school students in Sweden. Informants are, on the one hand, students in CLIL programmes (henceforth: CLIL students), where English, apart from being a separate subject, is also the medium of instruction in several or all subjects, e.g. biology and history, and, on the other hand, students in programmes where Swedish is used as the medium of instruction throughout the school day and English is studied as a separate subject (henceforth: non-CLIL students). The main focus is on students' proficiency and progress in written academic language, English and Swedish, used in the school context, which in many ways deviates from more personal, oral, everyday communication (Schleppegrell, 2004).

The focus on academic language is by no means unique to this project. On the contrary, recent years have seen a growing interest in this particular field of English, vocabulary as well as grammar in a wide sense, also from an L2/FL perspective, as evidenced in the publication of, for example, the *Longman Exams Dictionary* (2006) (see Ohlander 2007) and a grammar like the *Cambridge Grammar of English* (Carter & McCarthy, 2006), where special attention is paid to academic English. This, in turn, may be explained by the spread of English as a global language, the world's foremost lingua franca (Graddol 1997, Crystal 2003), not least within European institutions of higher education in the wake of the Bologna Process (<http://www.ehea.info/>), making proficiency in the academic registers of English a necessity for students intending to continue their education at tertiary level. Of interest here is the distinction made by Cummins (1979) between basic interpersonal communication skills (BICS) and cognitively advanced language proficiency (CALP). The students involved in the CLISS project are all enrolled in theoretical, academically oriented programmes aiming at preparing students for higher education. It is therefore of great interest to investigate the possible increase in the CALP register among the informants.

² The CLISS project is funded by the Swedish Research Council, project no. 721-2010-5376.

More specifically, the main research questions, in which a comparison between CLIL and non-CLIL students – as well as between English and Swedish – is paramount, are as follows:

- How do productive and receptive competencies in academic language develop?
- How well are subject-specific terms in biology and history mastered?

Subsidiary questions are:

- To what extent, if any, are there differences between students with Swedish as their first language and students who speak Swedish as a second language?
- To what extent, if any, are gender differences in evidence?

Further, as CLIL is widespread in many countries, it is also of interest to consider the results obtained from an international perspective.

3.2 Method and Material

Since the overall purpose of the CLISS project, which runs from 2011 through 2014, is to provide as multifaceted a picture of CLIL in Sweden as possible, several different methods are employed. First of all, the study is longitudinal, spanning three years. Thus, students are followed from their start at upper secondary school, year 10, throughout the three years making up this educational level in the Swedish school system. This makes for ample comparative opportunities. More specifically, groups can be compared at any given point in time; also, throughout the period of the study, individuals can be compared with themselves over time.

Table 1 gives an overview of the types of data collected within the CLISS project.

Table 1. Overview of data collection.

	Year 1		Year 2		Year 3	
	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6
Student background questionnaire	X					
Motivation questionnaire	X					X
Vocabulary tests	Sw/Eng					Sw/Eng
Synonyms & Collocations		Sw/Eng				Sw/Eng

Reading comprehension	Eng	Sw			Sw/Eng	
Argumentative text			Sw/Eng		Sw/Eng	
Exploratory text			Sw/Eng	Sw/Eng		
Classroom observations	X	X	X	X	X	X
Student interviews	X	X	X	X	X	X
Teacher interviews		X		X		X
School board interviews	X			X		

As Table 1 illustrates, both quantitative and qualitative data are included in the study. Examples of quantitative data are the questionnaires and the vocabulary tests. Qualitative data are, among other sources, to be found in the interviews. It should be noted, however, that most of the quantitative data can, and will, also be analysed qualitatively. As the focus of the project is on students' written proficiency, all empirical data specifically relating to language skills are in writing.

The vocabulary tests used are the *Vocabulary Levels Test* (VLT; Laufer & Nation, 1999; Nation, 2001) for English and a similar test for Swedish. Further, the tests concerning synonyms and collocations are, for English, the *Depth of Vocabulary Knowledge Measure* and the *TOEFL Vocabulary Measure*, as described in Qian and Schedl (2004), and corresponding tests for Swedish. The reading comprehension tests in English are tests originally designed for use in the English Reading Comprehension (ERC) part of the Swedish Scholastic Assessment Test for higher education (SweSAT) and which, when tried out on fairly large numbers of test takers, have proved slightly too easy for inclusion in the high-stakes SweSAT (Ohlander 1996, Reuterberg and Ohlander 1999). Similar tests in Swedish have been constructed. The tests of free written production have been carefully designed, based on curriculum goals, teacher experience and pilot testing, in both Swedish and English. Apart from these tests, there is a background questionnaire covering students' language background and experience, parents' educational level, students' self-assessment of language proficiency, and other relevant areas, collected at the outset and the end of the three-year period. Finally, a questionnaire tapping into students' motivation, language anxiety and willingness to communicate (Dörnyei, 2009) is also included in the empirical data.

Apparently unique to the CLISS project is the fact that we follow students throughout their entire upper-secondary school experience, i.e. for three full school years, and also that we collect similar and simultaneous data in both English and Swedish from all our informants, i.e. both CLIL and non-CLIL students. However, as analyses remain to be performed on the bulk of the empirical data, the present article reports only on the baseline results obtained on

the VLT.

3.3 Informants

At the outset of the study, in 2011, the informants in the CLISS project were 15–16 years old. They had just finished their nine compulsory years of schooling and started their three years at upper secondary level, which, while not obligatory by law, is opted for by approximately 98% of all students in Sweden (Skolverket, 2012).

In all, 221 students in eight groups at three different schools were invited to take part in the study. A letter of consent, in accordance with the guidelines set up and authorized by the regional ethical review board at the University of Gothenburg (<http://www.epn.se/en/start/startpage/>), was given to all students. A total of 203 positive replies were returned, resulting in a participation ratio of 92% among the initial students in the project. As is always the case in longitudinal, classroom-based studies, things happen in the course of the study. As a result, some of the initial students involved in the project have left for other schools while others have joined in at a later date. In year 3 of the data collection, a total of more than 240 students were involved in the study.

The three schools differ slightly from one another. Two of them are located in medium-sized cities and have a fairly homogeneous body of L1 Swedish students, although students with other L1s are found in all groups. The third school is located in a larger city with students displaying a variety of L1s. In fact, at this school Swedish L1 students are in the minority. At schools 1 and 2, the use of English varies across the school day depending on subject and current topic of interest. At school 3, English is spoken throughout the school day, and is often also heard among students during breaks.

All students attend academically oriented study programmes, aimed at preparing them for higher education. At school 1, one CLIL and one non-CLIL group take part in the study. At school 2, there are two groups of each kind, whereas at school 3, two CLIL groups (and no non-CLIL group) are included; see Table 2.

Table 2. Overview of group distribution: CLIL vs. non-CLIL, male vs. female students.

	CLIL groups	CLIL students	Non-CLIL groups	Non-CLIL student	Female students	Male students	Total
School 1	1	33	1	32	43	22	65
School 2	2	43	2	52	60	35	95
School 3	2	61	0	0	43	18	61
Total	5	137	3	84	146	75	221

In Table 2, it can be seen that there are about twice as many females than males among the informants. As the CLISS project is carried out in intact classes, the distribution of students could not be controlled for. Furthermore, there are more CLIL groups (N=5) than non-CLIL (N=3), and CLIL groups in general tend to have a majority of female students (Lasagabaster & Ruiz de Zarobe, 2010; San Isidro, 2010; Sylvén, 2010). It should also be noted that students with another L1 than Swedish seem to prefer the CLIL strand as opposed to the non-CLIL one. In the CLISS project, students with another L1 make up approximately 23% of the total number of students, compared to approximately 4% in the non-CLIL groups (Sylvén and Thompson, in press).

3.4 Research team and research perspectives

Just as the student body is somewhat heterogeneous, so is the project team. The multifaceted nature of the research aims of the project is reflected in the composition of the research team, whose members represent various theoretical perspectives and subject domains. Apart from senior researchers, five PhD students are presently attached to the project.

The senior researchers on the team represent various areas of expertise and illustrate the variety of perspectives taken in the project. English as a second/foreign language (L2/FL) is, of course, an area of primary interest in this project. Large amounts of data are collected to cover this aspect, more specifically in the form of vocabulary and reading comprehension tests, written texts, classroom interaction and interviews. Further, motivation is of importance in any learning environment, and in the CLISS project it makes up a specific area of interest. The students filled in comprehensive motivation questionnaires at the beginning of the project and will also do so at the end of it. This material is analysed in depth, being used as a backdrop to the various test results obtained.

The majority of the informants in the project have Swedish as their L1; consequently, Swedish as a first language is an important area of study. To look into Swedish L1, text analyses (based on, e.g., Systemic Functional Linguistics) will be made, as well as analyses of vocabulary and reading comprehension tests. In this connection, classroom observations and interviews will also be essential. A fair number of the informants have Swedish as their L2, which makes Swedish as a second language another relevant field of study. Data similar to those for Swedish L1 will be used to cover this perspective; in addition, focus groups of Swedish L2 students have been formed in which group discussions and interviews have taken place. Classroom interaction in both CLIL and non-CLIL groups is another area meriting further study. Accordingly, classroom observations are made throughout the period of investigation, by means of field notes as well as audio and video recordings.

The PhD students also cover distinctly separate areas. One of them investigates the effects of extramural English (EE) on CLIL and non-CLIL students' written English, especially with regard to some specific textual aspects, such as the use of cohesive devices (e.g., *however*, *therefore*), in the texts produced by the students.

To accomplish this, a questionnaire tapping into students' EE activities is administered and texts produced by the students are analysed, drawing on a Systemic Functional Linguistics framework. Another PhD student focuses on teachers' views on and understanding of CLIL, by closely following them both in class and in qualitative, personal interviews. Classroom observations and semi-structured interviews are used to this end. A third PhD student is concerned with giving an overview of CLIL in Sweden at large – there are no updated figures since Nixon (2000) on the spread of CLIL in Sweden – while at the same time taking a student perspective by, for example, shadowing individual students during entire school days. The fourth PhD student is specifically interested in the development of male students' writing proficiency in Swedish, including both CLIL and non-CLIL students in her study. This will be done through classroom observations, interviews and text analysis. The fifth PhD student is focused on the assessment practices among CLIL and non-CLIL teachers. By collecting tests from teachers, the aim is to identify any differences between CLIL and non-CLIL classes in this respect, while teacher interviews are intended to give an insight into teacher thinking as regards assessment.

We now turn to the very first test administered to the students, aimed at receptive proficiency with regard to English vocabulary.

4. The Vocabulary Levels Test

This section introduces, in some detail, the basic properties of the vocabulary test (VLT) given to the students, both CLIL and non-CLIL ones, participating in the CLISS project (4.1). The main part of it, however, accounts for the quantitative results of the first test round, administered at the very outset of the project (4.2).

4.1 The VLT: some basic features

Vocabulary is often referred to as the most fundamental building block of language (Carter, 1987; Nation, 1990), and as such perhaps the most important part of language for learners to acquire. Other aspects, such as grammar and pronunciation, are also important, of course, but without words there can be no basis – nor any need – for grammar or pronunciation. Therefore, the level of vocabulary proficiency can be used as an indicator of the command of an L2/FL, among individuals as well as groups of learners. The specific learner groups of interest in this paper are, above all, CLIL and non-CLIL students, but also males and females.

In view of the importance assigned to vocabulary knowledge in an L2, the very first test given during the first year of the CLISS project was indeed a vocabulary test, viz. the *Vocabulary Levels Test* (VLT; Nation 2001), mentioned earlier (section 3.2). The VLT is a frequency-based test, comprising items from the 2 000 level, i.e. the 2 000 most frequent words in English (e.g. *victory* and *develop*), up to the 10 000 level, i.e. much less frequently occurring words (e.g. *benevolence* and *pacify*). The VLT administered within the project also contains a special group of words taken from the *Academic Word List* (AWL; Coxhead, 2000). The

AWL is a compilation of words which are more frequent in academic texts in general than elsewhere (e.g. *evidence* and *indicate*). It consists of 570 word families – defined by Bauer and Nation (1993) as a stem plus affixed forms, inflectional as well as derivational (e.g., **increase**: *increase* (v), *increase* (n), *increased*, *increasing*, *increasingly*; **family**: *familiar*, *unfamiliar*, *familiarity*, *familiarize*) – and drawn from a corpus of 3.5 million words from texts in various academic domains, such as the arts and humanities, commerce, law, and science. The inclusion of a section devoted to academic vocabulary in the VLT fits nicely in with the main purpose of the CLISS project, which is to investigate students' proficiency in written, academic language.

Originally, the VLT was used as a way of testing learners diagnostically, but it has also been employed for various other purposes (Beglar, 2010). As regards the test format of the VLT, six words are given, three of which are to be paired up with one of three definitions/explanations or synonyms, as illustrated below.

- | | |
|--------------------|---|
| 1. <i>business</i> | |
| 2. <i>clock</i> | _____ <i>part of a house</i> |
| 3. <i>horse</i> | _____ <i>something used for writing</i> |
| 4. <i>pencil</i> | _____ <i>animal with four legs</i> |
| 5. <i>shoe</i> | |
| 6. <i>wall</i> | |

In total, the VLT includes 150 test items, 30 of which are taken from the 2 000 level, 30 from the 3 000 level, 27 from the 5 000 level, 21 from the 10 000 level, and 39 from the AWL (see App. 1).

The VLT is one of the tests administered twice during the CLISS project. The first test occasion was at the very outset of upper secondary level, i.e. at the start of CLIL for the CLIL students; the second occasion was more than two school years later, in the third and final grade of upper secondary level. In focus in this paper are the results from the first test round only, those from the second round not being available at the time of writing.

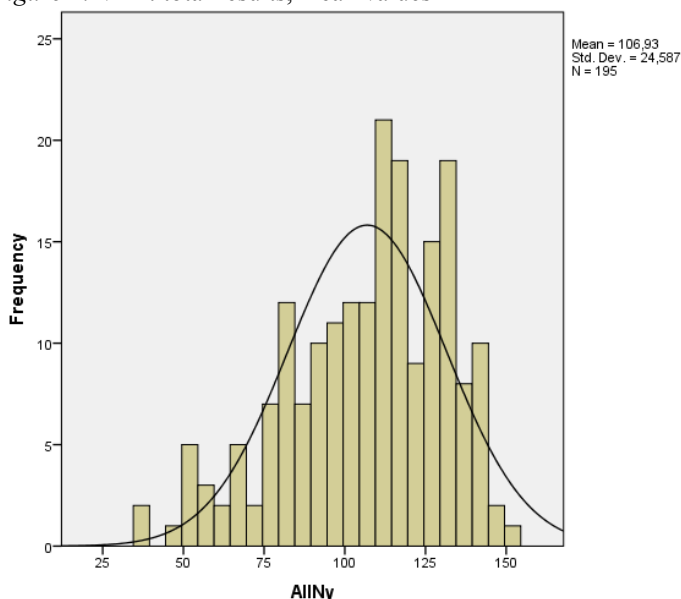
Our hypotheses, based on previous research on vocabulary size among CLIL and non-CLIL students (Sylvén, 2004; Washburn, 1997), are, first, that the CLIL students are likely to have a larger receptive vocabulary than the non-CLIL students already at the outset, before the start of CLIL; second, based on earlier findings on gender differences in vocabulary proficiency and reading comprehension (Herriman 1997, Reuterberg and Ohlander 1999, Sylvén and Sundqvist 2012), that male students are expected to outperform female ones; and third, based on the fact that up until the start of upper secondary school, the amount of input of English in school should be fairly equal between the three schools, i.e. there should be no differences between the schools involved in the project. Consequently, as regards CLIL versus non-CLIL students, our specific research questions for the English receptive vocabulary part of the CLISS project are:

1. Are there differences in receptive English vocabulary?
2. Are there gender differences?
3. Are there differences between schools?

4.2 Results

Figure 1 illustrates the overall results, for both CLIL and non-CLIL students, of the first VLT administered within the CLISS project in the autumn term 2011, including the normal curve.

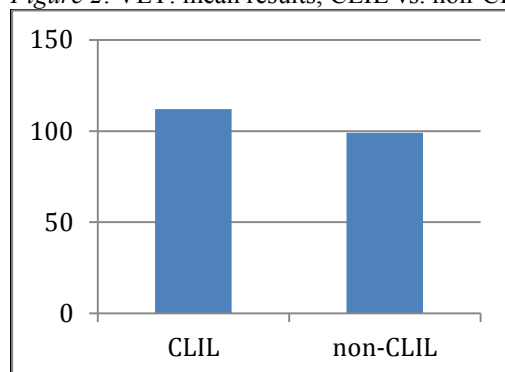
Figure 1. VLT: total results, mean values



As is evident in Figure 1, the normal curve is slightly skewed towards the right, indicating that the test overall was slightly too easy for the test takers. The mean is 107, but the variation in results is large, with a standard deviation (SD) of 25. It may be noted that one student managed to get all 150 items correct already at this first round of the test.

Using PASW software, statistical analyses were conducted. In Figure 2, the results from an independent samples T-test divided by group are illustrated.

Figure 2. VLT: mean results, CLIL vs. non-CLIL



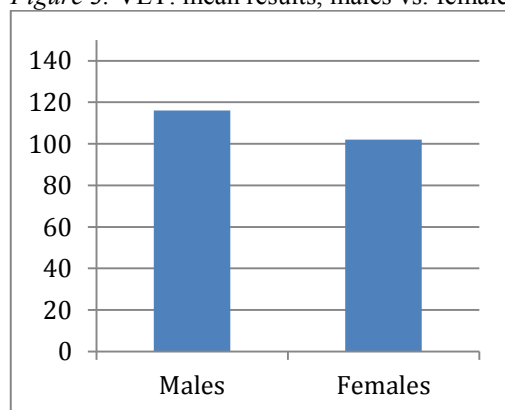
The group statistics behind the columns in Figure 2 are specified in Table 3, with details on mean results, number (N) of students and the SD in each group.

Table 3. VLT: group statistics, CLIL vs. non-CLIL

	Mean	N	SD
CLIL	112	119	22.8
non-CLIL	99	76	25.1

As shown in Figure 2 and Table 3, the CLIL students outperform their non-CLIL peers with a mean result of 112 vs. 99, which is a statistically significant difference ($p < 0.001$). Considered from a gender perspective, the results turn out as illustrated in Figure 3.

Figure 3. VLT: mean results, males vs. females



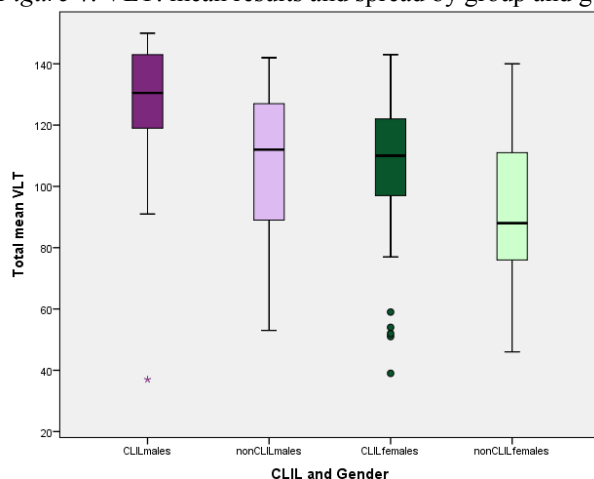
As can be seen from Figure 3, the males ($M = 116$, $SD = 25.6$) outperform the females ($M = 102$, $SD = 22.9$) significantly ($p < 0.01$).

So far, we have compared results from two groups, using independent T-test. In order to find out whether there are any statistically differences between CLIL and

non-CLIL students and the two gender subgroups within each of them, thus providing a basis for comparison involving four groups, we need to perform one-way analyses of variance, ANOVAs. To illustrate the results obtained in the ANOVA, box plots³ are used.

Figure 4 illustrates the total results on the VLT, broken down into the four-group division of group (CLIL vs. non-CLIL) and gender (males vs. females).

Figure 4. VLT: mean results and spread by group and gender



In Figure 4, we see that the CLIL males perform best of all four groups, and that the median for non-CLIL males and CLIL females coincides. Furthermore, it is evident that the results for both males and females in the CLIL group are more condensed, compared to the non-CLIL students, who exhibit more variation in results. In Table 4, the mean and standard deviation are presented.

Table 4. VLT: statistics for gender and group

Gender	Group	Mean	N	SD
Males	CLIL	126	30	22.8
	Non-CLIL	107	33	24.8
Females	CLIL	107	89	20.9
	Non-CLIL	92	43	23.7

As Table 4 shows, the CLIL males have the highest mean and the non-CLIL females the lowest, with the non-CLIL males and the CLIL females in-between,

³ In the box plot, a number of results can be found: first of all, the median (the thick line inside the box); second, the spread of the scores indicated by results in four percentiles; third, the 25th percentile (the lower whisker); fourth, the 50th percentile (the part in the box below the median line); fifth, the 75th percentile (the part in the box above the median line); and sixth, the 100th percentile (the upper whisker). Finally, the box plot also shows outliers (i.e. values greater than 1.5 interquartile ranges away from the 25th or 75th percentiles).

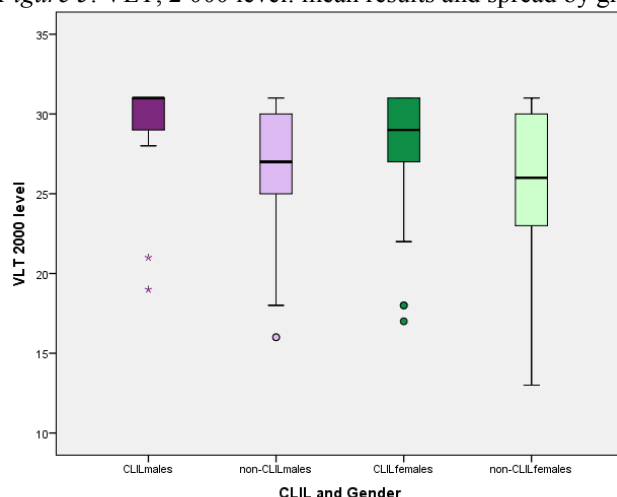
showing exactly the same mean. The ANOVA reveals that there is indeed a statistically significant difference between groups ($F[3, 191] = 13.45, p = .000$). Furthermore, a Tukey HSD post hoc test⁴ shows that the CLIL males score significantly higher than the non-CLIL males ($p = .005$), the CLIL females ($p = .001$), and the non-CLIL females ($p = .000$) on the VLT. The effect size⁵ is low ($\eta^2 = .174$), which is an indication that the strength of the measures for each specific group is not very high. This, in turn, can be explained by the fairly low number of participants in each group; see further below (section 5).

As explained above, the VLT is a frequency-based vocabulary test, and so may provide in-depth insight into different levels of vocabulary proficiency. In the following subsections, results per word frequency level are presented.

The 2000 level

This level of the VLT represents the most common words, and is thus to be considered the easiest part of the test. It consists of 30 items, such as *birth*, *debt* and *melt* (see App. 1). Figure 5 illustrates the results for the 2 000 level.

Figure 5. VLT, 2 000 level: mean results and spread by group and gender



As is evident from Figure 5, results across the four groups differ somewhat. The standard deviation in the two CLIL groups is smaller (CLIL males: $SD = 2.8$; CLIL females: $SD = 2.9$) than that of the two non-CLIL groups (non-CLIL males: $SD = 4.5$; non-CLIL females: $SD = 4.7$). In other words, there is less of a difference found in the results of the CLIL group than in those of the non-CLIL group. The ANOVA ($F [3, 191] = 10.38, p = .000$) shows a significant inter-group

⁴ In the post hoc test, the exact occurrence of the statistical differences detected in an ANOVA is indicated, i.e. whether there is a difference between all or just some of the groups included in the ANOVA (Green 2013).

⁵ Effect size is a complement to the inferential statistics obtained in a p-value, being a measure of "the strength of association between the two variables" (Green 2013).

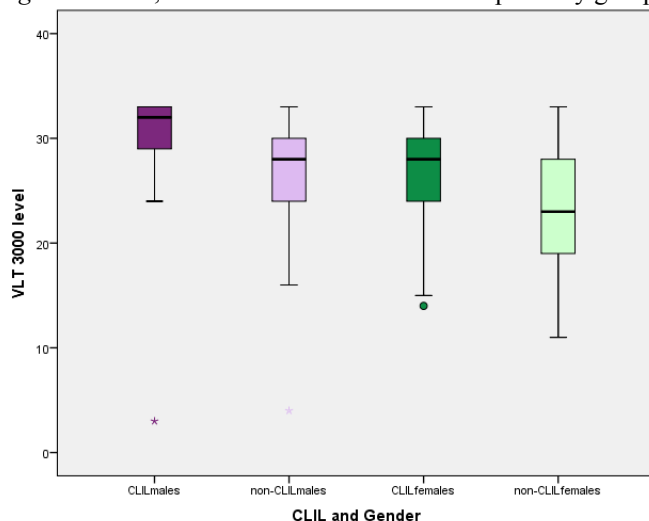
difference already at the 2 000 level, and the Tukey HSD post hoc test identifies the CLIL males as scoring significantly higher than both the non-CLIL males ($p = .004$) and the non-CLIL females ($p = .000$). Furthermore, the CLIL females score significantly higher ($p = .000$) than the non-CLIL females. The differences within both the CLIL and the non-CLIL group are non-significant. The effect size at the 2 000 level is low ($\eta^2 = .140$).

Thus, at the 2 000 level of the VLT, CLIL students score higher and more consistently than the non-CLIL students, whose results are lower and much more spread out.

The 3 000 level

At the 3 000 level, the words are still fairly common and include items such as *museum*, *blanket* and *grasp* (see App. 1). Here, too, the total number of items is 30. Figure 6 illustrates the outcome for this level of the VLT.

Figure 6. VLT, 3 000 level: mean results and spread by group and gender



As shown in Figure 6, the results among the CLIL males are found in the range 28–32 (SD = 5.7), with one outlier. The non-CLIL males are found in the range 24–28 (SD = 6.3), also with one outlier. The CLIL females score between 26 and 28 (SD = 4.2), with one outlier, while there are no outliers among the non-CLIL females, who score in the range 21–25 (SD = 5.4). Statistically significant differences, as shown by the Tukey HSD post hoc test, are found between the CLIL males and the non-CLIL males ($p = .011$), the CLIL females ($p = .028$) and the non-CLIL females ($p = .000$), as well as between the CLIL females and the non-CLIL females ($p = .001$). Once again, the effect size is at a low level ($\eta^2 = .148$).

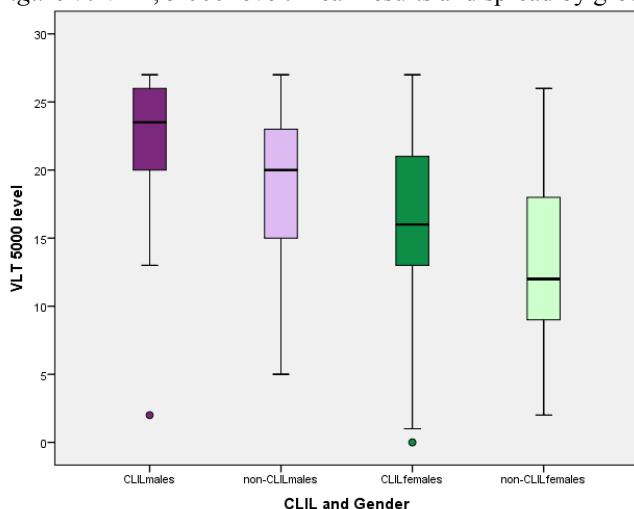
To sum up the results at the 3 000 level of the VLT, the CLIL males are clearly ahead of the other groups, with the CLIL females and non-CLIL males scoring

similarly in the middle, and the non-CLIL females exhibiting the lowest results. In addition, the results of the non-CLIL females are the most spread out, as indicated by the highest standard deviation.

The 5 000 level

At the 5 000 level of the VLT, the words are clearly more difficult than at the previous levels. A total of 27 items are included from this level in the test, e.g. words like *mortgage*, *summit* and *mansion* (see App. 1). Figure 7 illustrates the results for the 5 000 level.

Figure 7. VLT, 5 000 level: mean results and spread by group and gender



As is clear from Figure 7, the males in both groups are ahead of the females: the CLIL males are in the lead and the non-CLIL females have the lowest results. The ANOVA ($F [3, 190] = 14.47, p = .000$) shows significant inter-group differences, and the Tukey HSD post hoc reveals that the CLIL males score significantly higher than both the CLIL and non-CLIL females ($p = .000$), the non-CLIL males score significantly higher than the non-CLIL females ($p = .000$), and the CLIL females score significantly higher than the non-CLIL females ($p = .019$). The effect size is slightly larger than at the previous levels, but is still to be considered low ($\eta^2 = .182$).

In sum, the results at the 5 000 level mirror to a large extent those at the previous levels. The CLIL males score highest and the non-CLIL females lowest, with the CLIL females and the non-CLIL males in-between. However, this is the first level where the results of the non-CLIL males are significantly higher than those of the non-CLIL females.

The 10 000 level

The 10 000 level of the VLT consists of the most difficult items in this test. It

includes 21 items, e.g. words such as *benevolence*, *immerse* and *vindictive* (see App. 1). Figure 8 illustrates the results for this level.

Figure 8. VLT, 10 000 level: mean results and spread by group and gender

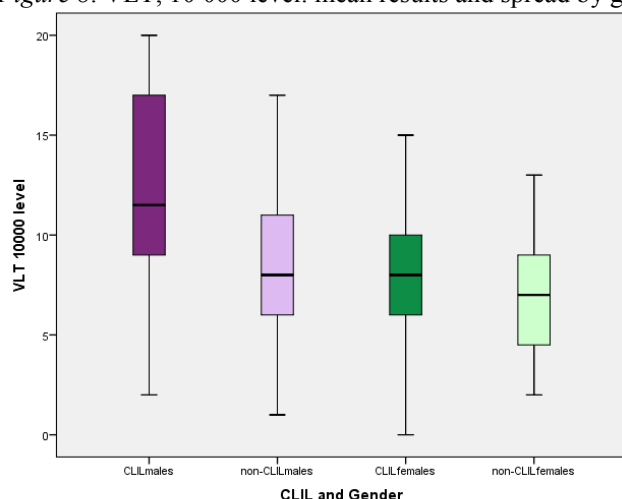


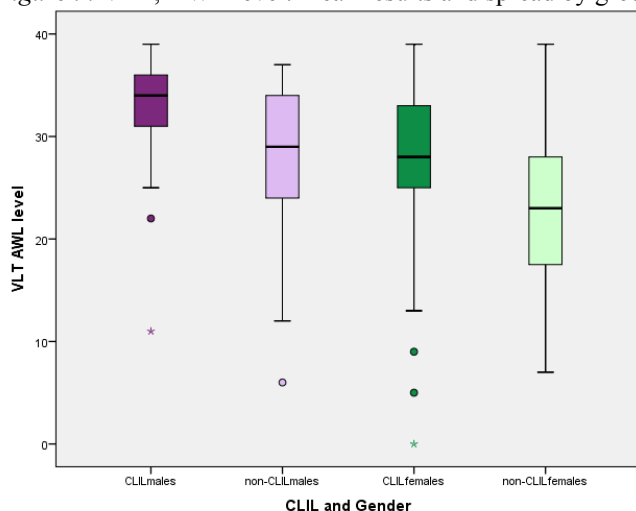
Figure 8 shows that the CLIL males are very clearly in the lead, with the three other groups at fairly similar levels. The ANOVA ($F [3, 189] = 12.03, p = .000$) demonstrates a statistically significant inter-group difference and the Tukey HSD post hoc reveals that the CLIL males score significantly higher than all the other groups ($p \leq .001$), and also that there are no other significant inter-group differences. The effect size is low ($\eta^2 = .159$).

To summarize the 10 000 level, the CLIL males maintain their leading position, significantly higher than the other three groups, who all perform at similar levels.

The AWL level

Apart from the frequency-based items, the VLT also includes a section with vocabulary items selected from the Academic Word List (Coxhead 2000), as mentioned earlier. There are a total of 39 such items, including words as *evidence*, *gender* and *exclude* (see App. 1), Figure 9 illustrates the results from this section.

Figure 9. VLT, AWL level: mean results and spread by group and gender



As shown in Figure 9, the results in this segment of the VLT are spread out to a larger extent than in the previous ones. The ANOVA ($F [3, 190] = 12.63, p = .000$) indicates a significant inter-group difference, and the Tukey HSD post hoc informs us that the CLIL males score significantly higher than all the other groups ($p \leq .01$). Further, the non-CLIL males score significantly higher than the non-CLIL females ($p = .043$), and the CLIL females likewise score significantly higher than the non-CLIL females ($p = .000$). The effect size is, once again, at a low level ($\eta^2 = .152$).

To put it briefly, in the AWL segment of the VLT, too, the CLIL males score the highest and the non-CLIL females the lowest.

Finally, the results will be looked at through the lens of the three different schools. As was seen in Table 2, schools A and B include both CLIL and non-CLIL strands, whereas at school C, there are only two CLIL classes. Furthermore, the student body is more heterogeneous as regards students' L1 at school C compared to schools A and B. Therefore, it is of interest to see if there are initial differences also at school level before upper secondary – and CLIL, for the CLIL students – starts. An ANOVA was run for this purpose, with the three schools as the independent factor. Very briefly, the results show that there are no statistically significant differences between the schools. A subsequent Tukey HSD post hoc confirmed this outcome. Thus, we can safely assume that, at school level, there are no statistically significant baseline differences.

After this statistical account of the results from the VLT at group, gender and school level, let us now discuss in some more depth, and in a wider context, how these data may be interpreted.

5. Discussion

So far in this paper, an overview has been given of the large-scale, longitudinal

CLISS project, the main purpose of which is to assess, at upper secondary school, students' proficiency and progress in written academic language skills in both English and Swedish. The project also aims at gaining a wider understanding of CLIL practices in a Swedish context, as well as comparing it with other national contexts. Further, the results of the first test administered within the CLISS project, the Vocabulary Levels Test have been accounted for, providing baseline vocabulary data at the onset of the CLISS project. This section starts out with a brief discussion of the results obtained from the VLT, as presented above. It then proceeds to a general discussion of some of the perspectives taken in the CLISS project.

First of all, a word of caution is in order as regards the generalizability of the results of the VLT. There are statistically significant differences between groups in several cases; however, the effect size is low. This means that there are indeed differences within the samples tested here, but because the groups are relatively small, these differences are not necessarily true across populations. Therefore, we need to be careful not to draw conclusions based only on the results reported here. Also, it should once again be noted that we only report findings from the initial VLT-test, which is the only test so far analysed, and the only data available at the time of writing.

As we have seen, the CLIL group significantly outperformed non-CLIL students. Thus, our first hypothesis is confirmed, and these findings are completely in line with previous research (Washburn, 1997, Sylvén 2004), thus only to be expected. Attending a CLIL class is an option decided on by individual students (and their parents), and thus, an active, voluntary choice. Students who find English difficult or uninteresting simply do not choose to attend CLIL classes, while students who enjoy English, are good at it and/or feel motivated by the challenge of being taught other school subjects through that language are more inclined to do so.

It was also found that male students outperformed female ones, confirming our second hypothesis, which may, at first glance, seem slightly more puzzling. However, research has indicated that students' extramural English (EE) activities (i.e. activities performed through the medium of English in students' spare time, outside of school) are of crucial importance in this connection ([Olsson 2011](#), [Oscarson & Apelgren 2005](#), [Sundqvist 2009](#), [Sundqvist & Sylvén 2012a](#), 2012b; [Sylvén & Sundqvist 2012](#)). Moreover, what type of EE students engage in is another important aspect to take into account. There is a distinction to be made between active and passive EE, in line with active and passive language skills. By active language skills, we refer to writing and speaking, while passive language skills relate to reading and listening. In a similar vein, when we talk about active EE, reference is made to, for instance, reading books, writing letters or playing digital games. Passive EE, on the other hand, involves such activities as listening to song lyrics or watching TV. The difference between active and passive activities is that, in the active ones, the individual is required to do something, e.g. speak or write, whereas in the more passive ones, the individual does not need to

perform but may only receive input in through, for instance, music or TV. More active EE would, generally speaking, seem to be more beneficial to vocabulary acquisition than passive EE, as there is a need for output of some kind to be produced. In the studies referred to above, it is clearly shown that males' EE is usually more active than females' (Sundqvist & Sylvén 2012b). For instance, even though both males and females play digital games, males tend to prefer so-called massively multiplayer online role-playing games, where social interaction is an integral part. Females, on the other hand, seem to prefer the more single-player, offline type of games. This is a possible explanation for the fact that, at the outset of the study, male students had a significantly larger English vocabulary than female ones. In subsequent analyses, possible correlations between the students' EE and their performance on the various tests administered in CLISS, including the VLT, will be investigated.

As the VLT is divided into different frequency levels, the results invite some interesting observations. First of all, we have seen that CLIL males score highest, and statistically significantly so, not only regarding the total result, but also at each single level of the VLT. This indicates that the male students who are attracted to and actively opt for the CLIL programme are those who, even before the start of CLIL, are highly proficient, at least as regards lexical competence. In this connection, several studies (Lasagabaster, 2011; San Isidro, 2010; Sylvén, 2004) note that female students make up a majority in CLIL programmes. It may be speculated that only males who are already quite proficient in English choose CLIL, whereas female students may view the CLIL option more as a way of pursuing their interest as well as improving their proficiency in English. This hypothesis also finds support in the fact that CLIL females and non-CLIL males display a close similarity in their results: in none of the levels is there a statistically significant difference between these two groups. Thus, there is indeed a possible selection effect, which would correspond to Rumlich's (2013) findings.

The non-CLIL males' overall results are more or less identical to those of the CLIL females, but at the 5 000 and 10 000 levels as well as in the AWL section of the test, they score slightly higher. However, there are no significant differences between these two groups at any of the levels. These results indicate that the more difficult – i.e. less frequent – the words are, the more proficient are the non-CLIL male students as compared to the CLIL females. However, the results also show that, overall, the spread within both of these groups is much larger than that in the group of CLIL males. This, in turn, indicates that in the non-CLIL groups, the males are much more heterogeneous with regard to proficiency in English, whereas the CLIL groups include only those who, from the outset, are relatively proficient. From an educational and classroom point of view, the non-CLIL group thus seems to present a greater challenge for the EFL teacher than does the CLIL group.

Just as it is very clear that the CLIL males are in the lead when it comes to vocabulary as measured by the VLT, so is the fact that the non-CLIL females are those who consistently score lowest in terms of mean results. However, the box

plot illustrations in the previous section allow us to see the distribution of the results in each group, and it then becomes obvious that the widest overall spread is actually found in the group consisting of non-CLIL females. Why this is so is not clear. There are, of course, a multitude of potentially underlying factors – a lack of interest among some students in performing well on tests that are not directly school-related, divergent L2 English backgrounds both in and out of school, other individual differences, etc. – to be controlled for if this state of affairs is to be fully understood. The CLISS project may, at least to some extent, contribute to their further illumination.

Among other instruments used in the project, the motivation questionnaire (Ryan, 2009) may help us understand some of the factors underlying students' achievements on the various tests. The analysis of the first round of the motivation questionnaire reveals, among other things, that the non-CLIL females display the greatest amount of language anxiety and, further, the smallest amount of self-confidence in using English (Sylvén & Thompson, in press). It may be the case that the relatively poor performance of the non-CLIL females on the VLT is a reflection of their low levels of self-confidence and high levels of language anxiety. As mentioned above, the motivation questionnaire will be repeated towards the end of the three years of upper secondary school, and comparisons between the two rounds will be made. Do the differences exhibited in the first round of the questionnaire remain constant over the three years of the study, or do they change? If the latter proves to be the case, in what ways are such changes evident? Further analyses, it is hoped, will bring to light inter-group differences, which, in turn, may help us understand why these differences exist and also how best to address them. These findings will not only be interesting *per se* and for the research community, they will also be of relevance for practicing teachers. If, for instance, the non-CLIL females continue to lag behind while at the same time displaying high levels of anxiety and low self-confidence, perhaps a great deal of attention should be paid to these individual characteristics, at the same time as the L2/FL is being taught.

Looking at the VLT results in some more detail, at each frequency level, we note that the CLIL males are in the lead at all levels. There are, however, some intriguing differences between the groups at the various levels, which call for some comments. At the 2 000 level, both CLIL males' and females' results are very "condensed", whereas the non-CLIL results are more spread out, with some very high as well as some very low results. In other words, at this level the results basically mirror the overall VLT results. However, at the 3 000, the 5 000, and the 10 000 levels, as well as in the AWL segment, while the CLIL males still perform fairly equally within their group (except at the 10 000 level, where they actually exhibit the largest spread of all groups), the profile of the CLIL females looks more similar to that of the non-CLIL males. These levels, of course, represent a more advanced vocabulary than the 2 000 level. Thus, it seems as though there are indeed quite a few male students in the non-CLIL group with a fairly advanced command of English vocabulary. Yet they did not opt for the CLIL programme; a

relevant question is why. Other empirical data in the CLISS project, not yet analysed, may shed some light on this question, in particular the background and motivation questionnaires. At this point, though, only speculation is possible: for instance, it may, for some reason, seem more natural for female students to choose the CLIL strand than it does for males. The overrepresentation of females in CLIL is by no means specific to the CLISS project, but appears to be the rule rather than the exception (Lasagabaster, 2011; San Isidro, 2010; Sylvén, 2004). Consequently, this imbalance seems to be in need of attention from those responsible for marketing CLIL programmes – provided, of course, that a gender balance is aimed for.

The results of the non-CLIL females offer some further food for thought. Even though they are indisputably the weakest group of the four, there are individuals among them who perform very well. Both at the 2 000 and the 3 000 levels, as well as in the AWL segment, the best results in this group coincide with the best among the CLIL females. And both at the 5 000 and the 10 000 levels, there are some CLIL females who actually score lower than any of the non-CLIL females. Similarly, as discussed above in relation to the advanced L2 English non-CLIL males' choice of programme, it will be interesting to find out more about these females' backgrounds, learning more about possible reasons for their preference for the non-CLIL strand.

In response to our third research question regarding possible differences between the participating schools, the results indicate that there are no such initial differences. Thus, our third hypothesis is also validated.

The findings reported on in this paper are of interest as they indeed verify what has long been expected, viz. that CLIL programmes by and large attract students who are already fairly proficient in the target language involved, in the present case English. This finding, obviously, should be highly relevant at the policy level where decisions on CLIL are made. Do we want CLIL to be a path for those who "already have an abundance" of knowledge in the target language, allowing them to become even more proficient and giving them a head start in higher education? Or, should CLIL also provide a way of motivating less proficient students to learn the target language? These are questions that need to be addressed, having important implications for the future of CLIL in Sweden – and possibly elsewhere.

6. Concluding remarks – looking ahead

At the forefront of the CLISS project is a comparison between CLIL and non-CLIL students. A detailed account of the project has been presented above, along with the results from the first round of the Vocabulary Levels Test. As is clear from the project description, a large amount of data is collected during a total of three school years. The results concerning students' receptive vocabulary reported here may serve as a baseline, to which results from other types of tests, as well as from a second round of the same test, will be related, in accord with the call for greater critical awareness in CLIL-related research, voiced by Bruton (2011) and

others. For instance, it will be of great interest to compare students' results on receptive vocabulary knowledge with their level of productive writing competence. Another intriguing comparative parameter relates to gender differences. As noted above, male students outperform female ones significantly on the VLT. An especially relevant question in this connection is whether there will be a similar difference when written productive competence is focused on.

In addition to the tests on vocabulary proficiency, tests on collocations and synonyms (Qian & Schedl 2004) are administered to the CLISS informants. In this way, further insights into their English receptive proficiency will be gained.

This article has mainly focused on English. However, as explained above, the CLISS project does not only look at students' proficiency in English, but also in Swedish, the first language of the majority of the informants. As mentioned earlier, similar tests are administered in both English and Swedish; thus, a test corresponding to the English VLT has been given on Swedish vocabulary. Obviously, the results on these two tests will be interesting to compare, especially longitudinally, where student progress will be evidenced. Do CLIL students make more progress in English receptive academic vocabulary than in Swedish, i.e. does English-medium instruction in any way "harm" Swedish receptive academic vocabulary? Put differently, does CLIL entail a domain-loss effect for Swedish? (Hyltenstam 2002, Josephson 2004). Is it the other way around with the non-CLIL students? Other comparisons to be made are, needless to say, those involving written productive skills, as demonstrated in the various text assignments, both in English and in Swedish, throughout the project. It remains to be seen if the findings from elsewhere (Admiraal et al 2006, Moreno 2009, Ruiz de Zarobe et al 2006, Zydatiss 2007) are mirrored in the CLISS project.

Not only, however, will the CLISS project produce a multitude of results on receptive and productive language skills. There will also be a wealth of data on classroom interaction, how students perceive and respond to CLIL and how teachers experience CLIL "in action".

The kind of longitudinal, comprehensive, and contrastive data generated within the CLISS project has been on the wish list among CLIL scholars for quite some time (Cenoz et al 2014, Dalton-Puffer, Nikula & Smit 2010, Lasagabaster 2011). Findings about the effects of CLIL over time will improve our knowledge about the demands and challenges facing CLIL teachers in the classroom. This, in turn, will provide a firmer grounding for designing teacher training courses, both pre- and in-service ones, specifically tailored to this reality. And only when there are specially trained CLIL teachers will we be able to see the full potential of CLIL, as envisaged by many scholars, in actual practice.

References

Admiraal, W., G. Westhoff & K. de Bot (2006), "Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English", *Educational Research and Evaluation*, 12(1), 75–93.

- Alvtörn, L. (2000), *Språk- och innehållsintegrerad undervisning under gymnasiet – en fallstudie*, unpubl. term paper. Lund: Department of Nordic Languages, Lund University.
- Bauer, L. & I.S.P. Nation (1993), "Word families". *International Journal of Lexicography*, 6, 253–279.
- Beglar, D. (2010), "A Rasch-based validation of the Vocabulary Size Test", *Language Testing*, 27(1), 101–118.
- Bruton, A. (2011), "Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010)", *Applied Linguistics*, 32(2), 236–241.
- Carter, R. (1987), *Vocabulary. Applied linguistic perspectives*, London: Allen & Unwin.
- Carter, R. & M. McCarthy (2006), *Cambridge grammar of English. A comprehensive guide. Spoken and written English. Grammar and usage*, Cambridge: Cambridge University Press.
- Cenoz, J., F. Genesee & D. Gorter (2014), "Critical analysis of CLIL: Taking stock and looking forward", *Applied Linguistics* 35(3), 243–262.
- Coxhead, A. (2000), "A new academic word list", *TESOL*, 34(2), 213–238.
- Coyle, D. (2007), "Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies", *The International Journal of Bilingual Education and Bilingualism*, 10(5), 543–562.
- Crystal, D. (2003), *English as a global language*, 2nd ed., Cambridge, UK: Cambridge University Press.
- Cummins, J. (1979), "Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters", *Working Papers on Bilingualism*, 19, 198–203.
- Dalton-Puffer, C. (2007), *Discourse in Content and Language Integrated Learning (CLIL) classrooms, language learning and language teaching*, 20, Amsterdam: John Benjamins.
- Dalton-Puffer, C., T. Nikula, & U. Smit (eds.) (2010), *Language use and language learning in CLIL classrooms*, Amsterdam: John Benjamins.
- Dentler, S. (2002), "Case 17: mixed ability – secondary", in Marsh, D. (ed.), *CLIL/EMILE - the European dimension*, Jyväskylä: University of Jyväskylä, 167–171.
- Dörnyei, Z. (2009), "The L2 motivational self system", in Dörnyei, Z. and Ushioda, E., eds., 9–42.
- Dörnyei, Z. & E. Ushioda (eds.) (2009), *Motivation, language identity and the L2 self*. Clevedon: Multilingual Matters.
- Green, R. (2013), *Statistical analyses for language testers*, Basingstoke: Palgrave Macmillan.
- Graddol, D. (1997), *The future of English*, London: The British Council.
- Herriman, J. (1997), *Vocabulary and reading comprehension*, Gothenburg: University of Gothenburg.
- Jexenflicker, S. & C. Dalton-Puffer (2010), "The CLIL differential. Comparing

- the writing of CLIL and non-CLIL students in higher colleges of technology", in Dalton-Puffer, C., T. Nikula, and U. Smit (eds.), 169–189.
- Jiménez Catalán, R. M. (2010), "Gender tendencies in EFL across vocabulary tests", in Jiménez Catalán, R. M. (ed.), *Gender perspectives on vocabulary in foreign and second languages*, Basingstoke: Palgrave Macmillan, 117–138.
- Jiménez Catalán, R. M. & Y. Ruiz de Zarobe (2009), "The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction", in Ruiz de Zarobe, Y. & R.M. Jiménez Catalán (eds.), 81–92.
- Jiménez Catalán, R. M., Y. Ruiz de Zarobe & J. Cenoz (2006), "Vocabulary profiles of English foreign language learners in English as a subject and as a vehicular language", *Vienna English Working Papers*, 15(3), 23–27.
- Kelly, L. G. (1969), *25 centuries of language teaching*, Rowley: Newbury House.
- Kjellén Simes, M. (2008), *Room for improvement? A comparative study of Swedish learners' free written production in English in the foreign language classroom and in immersion education*, unpubl. PhD thesis, Karlstad University Studies, 32, Karlstad University.
- Klippel, F. (2003), "New prospects or imminent danger? The impact of English medium instruction on education in Germany", *Prospect*, 18(1), 68–81.
- Lambert, W. E. (1977), "Cognitive consequences of bilingual schooling", *Linguistics*, 18(3), 13–133.
- Lasagabaster, D. (2011), "English achievement and student motivation in CLIL and EFL settings", *Innovation in Language Learning and Teaching*, 5(1), 3–18.
- Lasagabaster, D. & Y. Ruiz de Zarobe (eds.) (2010), *CLIL in Spain: Implementation, results and teacher training*, Newcastle upon Tyne: Cambridge Scholars Publishing.
- Lasagabaster, D. & J.M. Sierra (2010), "Immersion and CLIL in English: More differences than similarities", *ELT Journal*, 64(4), 367–375.
- Laufer, B. & Nation, P. (1999), "A vocabulary-size test of controlled productive ability", *Language Testing*, 16(1), 33–51.
- Lim Falk, M. (2002), *Språk- och innehållsintegrerad inläring och undervisning i praktiken: meningsfull målspråksundervisning*, MINS / Meddelanden från institutionen för nordiska språk vid Stockholms universitet, 51, Stockholm University.
- Lim Falk, M. (2008), *Svenska i engelskspråkig skolmiljö: ämnesrelaterat språkbruk i två gymnasieklasser*, unpubl. PhD thesis, Stockholm University.
- Llinares, A. & R. Whittaker, (2006), "Linguistic analysis of secondary school students' oral and written production in CLIL contexts: Studying social science in English", *Vienna English Working Papers*, 15(3), 28–32.
- Longman Exams Dictionary* (2006), Pearson/Longman.
- Marsh, D. (2000), *Using languages to learn and learning to use languages*, Jyväskylä: University of Jyväskylä.
- Moreno, S. (2009), "Young learners' L2 word association responses in two different learning contexts", in Ruiz de Zarobe, Y. & R.M. Jiménez Catalán,

- (eds.), 93-111.
- Nation, P. (1990), *Teaching and learning vocabulary*, New York: Newbury House.
- Nation, P. (2001), *Learning vocabulary in another language*, Cambridge: Cambridge University Press.
- Navés, T. (2011), "How promising are the results of integrating content and language for EFL writing and overall EFL proficiency?", in Ruiz de Zarobe, Y., Sierra, J. M. & F. Gallardo del Puerto (eds.), 155–186.
- Navés, T. & M. Victori (2010), "CLIL in Catalonia: An overview of research studies", in Lasagabaster, D. & Y. Ruiz de Zarobe (eds.), 30–54.
- Nikula, T. (1997), "Terminological considerations in teaching content through a foreign language" in Marsh, D., B. Marsland & T. Nikula (eds.), *Aspects of implementing plurilingual education*, Jyväskylä: University of Jyväskylä, 5–8.
- Nikula, T. (2005), "English as an object and tool of study in classrooms: Interactional effects and pragmatic implications", *Linguistics and Education*, 16(1), 27–58.
- Nixon, J. (2000), *Content and language integrated learning and teaching in Sweden*, Stockholm: Swedish National Agency for Education.
- Ohlander, S. (1996), "Forskning kring ordförrådet i det engelska läsförståelseprovet", in *Högskoleprovet: Genom elva forskares ögon*, Stockholm: Högskoleverket, 27–31.
- Ohlander, S. (2007), "Longman Exams Dictionary – en ordhandbok av akademiskt intresse", *LexicoNordica*, 14, 275–292.
- Olsson, E. (2011), "Everything I read on the Internet is in English". *On the impact of extramural English on Swedish 16-year-old pupils' writing Proficiency*, unpubl. PhL thesis, University of Gothenburg.
- Oscarson, M. & B.M. Apelgren (2005), *Nationella utvärderingen av grundskolan 2003 (NU-03). Engelska. Ämnesrapport till rapport 251*, Stockholm: Skolverket.
- Qian, D. D. & M. Schedl (2004), "Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance", *Language Testing*, 21(1), 28–52.
- Reuterberg, S.-E. & S. Ohlander (1999), *Engelsk läsförståelse i högskoleprovet*. Göteborg: Göteborgs universitet.
- Ringbom, H. (2012), "Review of recent applied linguistics research in Finland and Sweden, with specific reference to foreign language learning and teaching", *Language Teaching*, 45(4), 490–514.
- Ruiz de Zarobe, Y. (2010), "Written production and CLIL. An empirical study", in Dalton-Puffer, C., T. Nikula & U. Smit (eds.), 191–212.
- Ruiz de Zarobe, Y. & R.M. Jiménez Catalán (eds.) (2009), *Content and language integrated learning: Evidence from research in Europe*. Bristol: Multilingual Matters.
- Ruiz de Zarobe, Y., J.M. Sierra & F. Gallardo del Puerto (eds.) (2011), *Content and foreign language integrated learning*, Bern: Peter Lang

- Rumlich, D. (2013), "Students' general English proficiency prior to CLIL", in Breidbach, S. & B. Viebrock (eds.), *Content and Language Integrated Learning (CLIL) in Europe. Research perspectives on policy and practice*, Frankfurt: Peter Lang, 181–202.
- Ryan, S. (2009), "Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English", in Dörnyei, Z. & E. Ushioda, E. (eds.), 120–143.
- San Isidro, X. (2010), "An Insight into Galician CLIL: Provision and results", in Lasagabaster, D. & Y. Ruiz de Zarobe (eds.), 55–78.
- Schleppegrell, M. (2004), *The language of schooling. A functional linguistics perspective*, London: Lawrence Erlbaum Associates.
- Skolverket (2012), *Internationella språkstudien 2012*, Stockholm: Fritzes.
- Sundqvist, P. (2009), *Extramural English matters: Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary*, unpubl. PhD thesis, Karlstad University Studies, 55, Karlstad University.
- Sundqvist, P. & L.K. Sylvén (2012a), *Computer-assisted L2 English language-related activities among Swedish 10-year-olds*, translated by Bradley, L. and S. Thouésny, University of Gothenburg: Research-publishing.net, 280–285.
- Sundqvist, P. & L.K. Sylvén (2012b), "World of VocCraft: Computer games and Swedish learners' L2 vocabulary", in Reinders, H. (ed.), *Digital games in language learning and teaching*, Basingstoke: Palgrave Macmillan, 189–208.
- Sylvén, L.K. (2004/2010), *Teaching in English or English Teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*, *Gothenburg Studies in English*, 97, Göteborg: University of Gothenburg.
- Sylvén, L.K. (2007), "Swedish CLIL students' extracurricular contact with English and its relation to classroom activities", in Marsh, D. & D. Wolff (eds.), *Diverse Contexts – Converging Goals. CLIL in Europe*, Frankfurt: Peter Lang, 237–252.
- Sylvén, L.K. (2013), "CLIL in Sweden – Why does it not work? A metaperspective on CLIL across contexts in Europe", *International Journal of Bilingual Education and Bilingualism*, 16(3), 301–320.
- Sylvén, L.K. & P. Sundqvist (2012), "Gaming as extramural English L2 learning and L2 proficiency among young learners", *ReCALL*, 24(3), 302–321.
- Sylvén, L. K. & A.S.Thompson (in press), "Language learning motivation and CLIL: Is there a connection?", *Journal of Immersion and Content-Based Language Education*.
- Terlevic Johansson, K. (2011), *Erfolgreiches Deutschlernen durch CLIL?*, unpubl. PhD thesis, University of Gothenburg.
- Vollmer, H. J., L. Heine, L., R. Troschke, D. Coetzee & V. Küttel (2006), "Subject-specific competence and language use of CLIL learners: The case of geography in grade 10 of secondary schools in Germany", in *ESSE8 Conference*, London.
- Washburn, L. (1997), *English immersion in Sweden. A case study of Röllingby*

Liss Kerstin Sylvén & Sölve Ohlander – "The CLISS Project: Receptive Vocabulary..."

high school, 1987-1989, unpubl. PhD thesis, Stockholm University.

Zydatiss, W. (2007), *Deutsch-Englische Züge in Berlin (DEZIBEL). Eine Evaluation des bilingualen Sachfachunterrichts in Gymnasien: Kontext, Kompetenzen, Konsequenzen.*, Frankfurt: Peter Lang.

Appendix 1

VOCABULARY LEVELS TEST

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning.

Example

- | | |
|------------|----------------------------------|
| 1 business | |
| 2 clock | _____ part of a house |
| 3 horse | _____ animal with four legs |
| 4 pencil | _____ something used for writing |
| 5 shoe | |
| 6 wall | |

You answer it in the following way.

- | | |
|------------|-------------------------------------|
| 1 business | |
| 2 clock | <u>6</u> part of a house |
| 3 horse | <u>3</u> animal with four legs |
| 4 pencil | <u>4</u> something used for writing |
| 5 shoe | |
| 6 wall | |

Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, these words are *business, clock, shoe*.

The 2000 Level

1.

- 1 copy
 2 event _____ end or highest point
 3 motor _____ this moves a car
 4 pity _____ thing made to be like another

- 5 profit
 6 tip

2.

- 1 accident
 2 debt _____ loud deep sound
 3 fortune _____ something you must pay
 4 pride _____ having a high opinion of yourself

- 5 roar
 6 thread

3.

- 1 birth
 2 dust _____ game
 3 operation _____ winning
 4 row _____ being born
 5 sport
 6 victory

4.

- 1 clerk
 2 frame _____ a drink
 3 noise _____ office worker
 4 respect _____ unwanted sound
 5 theatre
 6 wine

5.

- 1 dozen
 2 empire _____ chance
 3 gift _____ twelve
 4 opportunity _____ money paid to the government
 5 relief
 6 tax

6.

- 1 admire
 2 complain _____ make wider or longer
 3 fix _____ bring in for the first time
 4 hire _____ have a high opinion of someone

- 5 introduce
 6 stretch

7.

- 1 arrange
 2 develop _____ grow
 3 lean _____ put in order
 4 owe _____ like more than something else

- 5 prefer
 6 seize

8.

- 1 blame
 2 elect _____ make
 3 jump _____ choose by voting
 4 manufacture _____ become like water
 5 melt
 6 threaten

9.

- 1 brave
 2 electric _____ commonly done
 3 firm _____ wanting food
 4 hungry _____ having no fear
 5 local
 6 usual

10.

- 1 bitter
 2 independent _____ beautiful
 3 lovely _____ small
 4 merry _____ liked by many people
 5 popular
 6 slight

The 3000 Level

11.

- 1 bull
- 2 champion _____ formal and serious manner
- 3 dignity _____ winner of a sporting event
- 4 hell _____ building where valuable objects are shown
- 5 museum
- 6 solution

12.

- 1 blanket
- 2 contest _____ holiday
- 3 generation _____ good quality
- 4 merit _____ wool covering used on beds
- 5 plot
- 6 vacation

13.

- 1 comment
- 2 gown _____ long formal dress
- 3 import _____ goods from a foreign country
- 4 nerve _____ part of the body which carries feeling
- 5 pasture
- 6 tradition

14.

- 1 apartment
- 2 candle _____ a place to live
- 3 draft _____ chance of something happening
- 4 horror _____ first rough form of something written
- 5 prospect
- 6 timber

15.

- 1 administration
- 2 angel _____ group of animals
- 3 frost _____ spirit who serves God
- 4 herd _____ managing business and affairs
- 5 fort
- 6 pond

16.

- 1 atmosphere
- 2 counsel _____ advice
- 3 factor _____ a place covered with grass
- 4 hen _____ female chicken
- 5 lawn
- 6 muscle

17.

- 1 abandon
- 2 dwell _____ live in a place
- 3 oblige _____ follow in order to catch
- 4 pursue _____ leave something permanently
- 5 quote
- 6 resolve

18.

- 1 assemble
- 2 attach _____ look closely
- 3 peer _____ stop doing something
- 4 quit _____ cry out loudly in fear
- 5 scream
- 6 toss

19.

- 1 drift
- 2 endure _____ suffer patiently
- 3 grasp _____ join wool threads together
- 4 knit _____ hold firmly with your hands
- 5 register
- 6 tumble

20.

- 1 brilliant
- 2 distinct _____ thin
- 3 magic _____ steady
- 4 naked _____ without clothes
- 5 slender
- 6 stable

21.

- 1 aware
- 2 blank _____ usual
- 3 desperate _____ best or most important
- 4 normal _____ knowing what is happening
- 5 striking
- 6 supreme

The 5000 Level

22.

- 1 analysis
- 2 curb _____ eagerness
- 3 gravel _____ loan to buy a house
- 4 mortgage _____ small stones mixed with sand
- 5 scar
- 6 zeal

23.

- 1 concrete
- 2 era _____ circular shape
- 3 fibre _____ top of a mountain
- 4 loop _____ a long period of time
- 5 plank
- 6 summit

24.

- 1 artillery
- 2 creed _____ a kind of tree
- 3 hydrogen _____ system of belief
- 4 maple _____ large gun on wheels
- 5 pork
- 6 streak

25.

- 1 chart
- 2 forge _____ map
- 3 mansion _____ large beautiful house
- 4 outfit _____ place where metals are made and shaped
- 5 sample
- 6 volunteer

26.

- 1 contemplate
- 2 extract _____ think about deeply
- 3 gamble _____ bring back to health
- 4 launch _____ make someone angry
- 5 provoke
- 6 revive

27.

- 1 demonstrate
- 2 embarrass _____ speech given by a priest in a church
- 3 heave _____ break suddenly into pieces
- 4 obscure _____ make someone feel shy or nervous
- 5 sermon
- 6 shatter

28.

- 1 correspond
- 2 embroider _____ exchange letters
- 3 lurk _____ hide and wait for someone
- 4 penetrate _____ feel angry about something
- 5 prescribe
- 6 resent

29.

- 1 decent
- 2 frail _____ weak
- 3 harsh _____ concerning a city
- 4 incredible _____ difficult to believe
- 5 municipal
- 6 specific

30.

- 1 adequate
- 2 internal _____ enough
- 3 mature _____ fully grown
- 4 profound _____ alone away from other things
- 5 solitary
- 6 tragic

The AWL Level

31.

- 1 area
- 2 contract _____ written agreement
- 3 definition _____ way of doing something
- 4 evidence _____ reason for believing something is or is not true
- 5 method
- 6 role

32.

- 1 construction
- 2 feature _____ safety
- 3 impact _____ noticeable part of something
- 4 institute _____ organization which has a special purpose
- 5 region
- 6 security

33.

- 1 debate
- 2 exposure _____ plan
- 3 integration _____ choice
- 4 option _____ joining something into a whole
- 5 scheme
- 6 stability

34.

- 1 access
- 2 gender _____ male or female
- 3 implementation _____ study of the mind
- 4 license _____ entrance or way in
- 5 orientation
- 6 psychology

35.

- 1 accumulation
- 2 edition _____ collecting things over time
- 3 guarantee _____ promise to repair a broken product
- 4 media _____ feeling a strong reason or need to do something
- 5 motivation
- 6 phenomenon

36.

- 1 adult
- 2 exploitation _____ end
- 3 infrastructure _____ machine used to move people or goods
- 4 schedule _____ list of things to do at a certain time
- 5 termination
- 6 vehicle

37.

- 1 alter
- 2 coincide _____ change
- 3 deny _____ say something is not true
- 4 devote _____ describe clearly and exactly
- 5 release
- 6 specify

38.

- 1 correspond
- 2 diminish _____ keep
- 3 emerge _____ match or be in agreement with
- 4 highlight _____ give special attention to something
- 5 invoke
- 6 retain

39.

- 1 convert
- 2 design _____ keep out
- 3 exclude _____ stay alive
- 4 facilitate _____ change from one thing to another
- 5 indicate
- 6 survive

40.

- 1 bond
- 2 channel _____ make smaller
- 3 estimate _____ guess the number or size of something
- 4 identify _____ recognizing and naming a person or thing
- 5 mediate
- 6 minimize

41.

- 1 explicit
- 2 final _____ last
- 3 negative _____ stiff
- 4 professional _____ meaning 'no' or 'not'
- 5 rigid
- 6 sole

42.

- 1 analogous
- 2 objective _____ happening after
- 3 potential _____ most important
- 4 predominant _____ not influenced by personal opinions
- 5 reluctant
- 6 subsequent

43.

- 1 abstract
- 2 adjacent _____ next to
- 3 controversial _____ added to
- 4 global _____ concerning the whole world
- 5 neutral
- 6 supplementary

The 10000 Level

44.

- 1 alabaster
- 2 chandelier _____ small barrel
- 3 dogma _____ soft white stone
- 4 keg _____ tool for shaping wood
- 5 rasp
- 6 tentacle

45.

- 1 benevolence
- 2 convoy _____ kindness
- 3 lien _____ set of musical notes
- 4 octave _____ speed control for an engine
- 5 stint
- 6 throttle

46.

- 1 bourgeois
- 2 brocade _____ middle class people
- 3 consonant _____ row or level of something
- 4 prelude _____ cloth with a pattern of gold or silver threads
- 5 stupor
- 6 tier

47.

- 1 contaminate
- 2 cringe _____ write carelessly
- 3 immerse _____ move back because of fear
- 4 peek _____ put something under water
- 5 relay
- 6 scrawl

48.

- 1 blurt
- 2 dabble _____ walk in a proud way
- 3 dent _____ kill by squeezing someone's throat
- 4 pacify _____ say suddenly without thinking
- 5 strangle
- 6 swagger

49.

- 1 illicit
- 2 lewd _____ immense
- 3 mammoth _____ against the law
- 4 slick _____ wanting revenge
- 5 temporal
- 6 vindictive

50.

- 1 indolent
- 2 nocturnal _____ lazy
- 3 obsolete _____ no longer used
- 4 torrid _____ clever and tricky
- 5 translucent
- 6 wily