universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

A comparison of lexical learning in CLIL and traditional EFL classrooms

Verfasserin

Eva Maria Seregély

angestrebter akademischer Grad

Magistra der Philosophie (Mag. phil.)

Wien, im Oktober 2008

Studienkennzahl lt. Studienblatt:	A 190 344 406
Studienrichtung lt. Studienblatt:	Unterrichtsfach Englisch
Betreuerin / Betreuer:	Ao. Univ.-Prof. Mag. Dr. Christiane Dalton-Puffer

To my late grandmother

Toda luna. Todo año.
Todo día. Todo viento.
Camina y pasa también.
También toda sangre llega
Al lugar de su quietud.
CHILAM-BALAM DE CHUMAYEL

Each moon. Each year.
Each day, and each wind.
Makes its way and also passes.
And in the same way, all living blood
Finds its place to rest.
[my translation]

Acknowledgements

The final stage of one's studies is an exciting journey and, unlike at the beginning, for a considerable time, you are travelling alone. Nonetheless, on my way towards completing this thesis, I have met quite a number of people who helped me carrying my luggage and walked along with me for a while. These people deserve special mention:

First and foremost, I would like to thank my supervisor Ao. Univ.-Prof. Mag. Dr. Christiane Dalton-Puffer, who motivated, advised and supported me from the very first day on, and who gave me the idea for the present thesis, and awakened my interest in Applied Linguistics, also in view of possible further projects in future.

Since this study is a replication, I am greatly indebted to Dr. Liss Kerstin Sylvén, who kindly provided her test materials and was always willing to help me with questions and problems concerning the administration and evaluation of my own empirical field work. In this connection, many thanks also go to Marianne Feldhofer for translating the Swedish questionnaires.

It goes without saying that studies of the present kind would not be possible without the contributions of the informants involved. I wish to acknowledge the debt I owe to all the students and teachers who painstakingly completed the lexical tests and questionnaires without receiving any consideration, apart from the opportunity to participate in a research project. I am also especially grateful to the coordinator of the school's bilingual programme, who supported me in organisational matters and afforded detailed insights into this particular educational setting, even beyond the scope of the present survey.

When it came to the statistical analysis of the test results, Mag. Peter Lucsanits has been a pleasant and helpful companion, as he taught me to draw on my second university subject and examine the linguistic data from the perspective of a mathematician. For this I am deeply thankful.

Furthermore, I owe a great deal to my colleagues in the diploma seminar, who, through their own research, suggestions and comments, have always encouraged and enlightened me. Special thanks also go to Mag. Dr. Christine Lechner, for proof-reading parts of my thesis and providing me with useful feedback regarding content.

Finally, I would like to express my sincere gratitude to my family and friends, who amused and distracted me, whenever this was necessary, and by doing so, were a genuine source of inspiration.

Above all, two people deserve even more honourable mention, that is, my parents, who supported me throughout my studies, constantly urged me on and, in particular, bore with all my emotional ups and downs during the last, strenuous stages of this thesis. For this I am eternally grateful!

Despite the help of all these wonderful people, whatever shortcomings remain are my responsibility alone.

Declaration of authenticity

I confirm to have conceived and written this paper in English all by myself.
Quotations from other authors and any ideas borrowed and/or passages paraphrased from the works of other authors are all clearly marked within the text and acknowledged in the bibliographical references.

Wien, im Oktober 2008 _____

Table of Contents

Abbreviations and symbols
used in the thesis

BNC	British National Corpus
CBI	Content-based instruction
CLIL	Content and Language Integrated Learning
CLT	Communicative Language Teaching
COBUILD	Collins COBUILD English dictionary
Coord.	Coordinator
CT	cloze test
DLP	Dual Language Programme
E	English
EFL	English as a Foreign Language
f	female
$F_{f_1,f_2;1-\alpha/2}$	critical value, F-test (= F-value)
FL	foreign language
F_{pr}	test value, F-test
G	German
H_0	null hypothesis
H_A	alternative hypothesis
i+1	input at a level slightly beyond the learners' linguistic competence (cf. Krashen 1985)
LEO	LEO Online Deutsch-Englisches Wörterbuch
L1	first language (= mother tongue, native language)
L2	second language (= foreign language)

m	male
MC	multiple choice
MED	Macmillan English dictionary
N, n	number of subjects in a given sample
NS	native speaker(s)
NST	native speaker teacher(s)
OALD	Oxford advanced learner's dictionary
OCD	Oxford collocations dictionary
ODEI	Oxford dictionary of English idioms
QR	questionnaire
$.^{r}$	reduced test battery
R+, R-	reader(s), non-reader(s) (of English texts)
r_s	Spearman rank order coefficient
SLA	Second Language Acquisition
SR	self-report
T	teacher
T. in ESC	time spent in English-speaking countries
$t_{n_1+n_2-2;\ 1-\alpha}$	critical value t-test (= t-value)
t_{pr}	test value, t-test
TR (I, II, III)	test round (I, II, III) in the original study (Sylvén 2004)
TRS	traditional (student/s)
VBS	Vienna Bilingual Schooling
WIC	words in context
+	with
\	without

1. Introduction

1.1. Background

"Warum nur 1 Sprache, wenn man 2 haben kann." [QR.VBS.6f]

This statement not only reflects a student's personal motive for deciding on a bilingual secondary school programme, but also mirrors the main incentive for the emergence of a series of innovative models for foreign language teaching in Europe in the past 25 years. In 1984, the European Community released a document setting out the demand that "[o]ne modern language in addition to the mother tongue should be studied in depth" (European Community 1984). Meanwhile, this claim has been revised. According to an action plan launched by the European Commission in 2003,

> [l]earning one *lingua franca* alone is not enough. Every European citizen should have meaningful communicative competence in at least two other languages in addition to his or her mother tongue. (European Commission 2003)

One possible way to achieving this ambitious goal has been found in the introduction of *Content and Language Integrated Learning*, in short *CLIL* (compare European Commission 2007). The basic idea behind this method is to use a foreign language in the teaching of content subjects, such as Biology, History or Maths. In the Austrian educational context, in the majority of the cases, the target language is English, but might just as well be Croatian, Czech, French, Hungarian, Italian, Slovak, Slovene or Turkish (Eurydice 2006: 5; 15-17).

The most essential feature of CLIL classes is that they provide students with a higher amount of foreign language input than would be possible in the traditional setting, that is, in the formal language classroom. What is crucial in this connection is that "CLIL is based on language acquisition rather than enforced learning" (Darn 2006), meaning that the development of linguistic skills should occur in a more natural way, resembling a child's acquisition of their mother tongue (compare *CLIL Compendium* 2001).

The origins of *CLIL* can be traced back to the year 1965, when a pedagogic experiment, which later came to be known as the "immersion method" (Lambert

1977), was initiated in the officially bilingual Canada, in order to promote students' French and English language proficiency (Sylvén 2004: 1). In the following, small-scale replications and adaptations of this project were implemented, and existed along with more structural and systematic approaches in countries all over the world (Sylvén 2004: 2). While for instance in Sweden, the first CLIL experiment was introduced as early as the 1970s (Sylvén 2004: 2), it was not until the beginning of the last decade of the 20th century that - with the establishment of the *Vienna Bilingual Schooling* concept - Austria, too, saw the integration of *Content and Language Integrated Learning* into its educational landscape (Eurydice 2006: 4; 6). Henceforth, CLIL programmes have been expanded and attracted growing interest.

> [T]he current situation in Austrian secondary schools is characterised by a wide spectrum of organisational forms ranging from 'mini-projects' with just a few lessons to bilingual schooling. (Eurydice 2006: 4-5)

While the number of CLIL schools and projects has been continuously increasing, for quite some time, research on the effectiveness of the method was lagging behind. The situation changed, more or less with the turn of the millennium, and from this point on, also the number of CLIL-focused studies has been growing steadily (for more details, see, e.g. Dalton-Puffer 2007).

One influential work in this area is Sylvén's *Teaching in English or English teaching?: on the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*, an empirical study conducted in 1999-2004 and involving four different Swedish upper secondary schools. As the title reveals, this research project dealt with one particular aspect of language learning, namely lexical competence. In order to discover whether CLIL students acquire a larger vocabulary than their peers taught in the traditional way, that is, with Swedish as the medium of instruction and English as a separate subject, Sylvén devised a test battery of four types of vocabulary tests (see, e.g. Sylvén 2004: 6-7). In addition, questionnaires concerning personal background were filled in by the students and the CLIL teachers.

Generally speaking, the results of Sylvén's study show that the CLIL learners outperform the traditional learners in all of the areas tested. Yet, other factors, above all, voluntary reading, were proven to have the same effect on the students' lexical development (compare Sylvén 2004: 224-226).

The present thesis aims at replicating Sylvén's work, albeit on a smaller scale. By adopting the original research methods and test materials, I will try to investigate whether in an Austrian school setting, the effects of *CLIL* on students' vocabulary knowledge are similar or even the same. The following section provides an overview of my main research interests.

1.2. Aims

Just as the original (compare Sylvén 2004: 4), the present thesis focuses on the study of general English vocabulary knowledge and lexical proficiency. [1] It is exclusively concerned with written language use. In contrast to Sylvén's work, which had a longitudinal dossier, this survey is restricted to one single test round. Hence, rather than on developmental aspects, my emphasis will be on students' lexical competence at one specific point in time. In this connection, my main research question is:

- ◊ Do CLIL students have a larger and more complex English vocabulary than traditional students?

As mentioned above, for the present investigation, I will adopt the materials used in the original study. Since these include four different kinds of lexical tests, an important issue to consider will be:

- ◊ To what extent is students' lexical performance dependent on/reflected by the test type and format used?

In the previous section, it has been indicated that English input from outside the CLIL or EFL classroom had a major influence on the Swedish students' vocabulary acquisition (see, e.g. Sylvén 2004: 224). English as the medium for international communication plays an essential role in today's everyday life, as people watch English movies, browse through English-speaking websites and travel all over the world. Thus, we will also analyse the following factors:

[1] Technical or subject-specific vocabulary will not be dealt with in the present work.

◊ Do leisure activities, such as reading, writing, watching TV/movies and playing computer games in English, have an impact on students' lexical proficiency?

and

◊ To what extent does time spent in English-speaking countries contribute to students' lexical proficiency?

Another interesting finding of Sylvén's study was that throughout the entire test sequence, the male students performed better than their female peers did (compare Sylvén 2004: 227). Consequently, in the present work, we also need to take into account the gender-aspect:

◊ Are there gender differences with regard to test performance and lexical proficiency?

For my research purposes, I will focus on an educational setting in Vienna that is generally known for its multicultural student population. It is therefore quite likely that the test groups under study will not exclusively consist of native speakers of German and/or English. The impact of this factor raises the question:

◊ To what extent do students' native languages affect their performance with regard to English vocabulary?

Evidently, in an investigation on (linguistic) performance, a special emphasis is always placed on those who do best. In other words,

◊ What is the key to acquiring a high-level lexical competence, i.e. what are the characteristic features of the highest performing students in the present test population?

It has been proven that one important benefit of the CLIL method is that it increases students' motivation and their willingness to learn (see also *CLIL Compendium* 2001; Abendroth-Timmer 2007; Dalton-Puffer 2007). Thus, in the present thesis, the following issue will also be of particular interest:

◊ Are there differences in motivation, attitude and self-assessment between CLIL students and traditional students?

Finally, we will also consider teachers' roles and perspectives in connection with *CLIL*:

◊ How do teachers estimate the success of their school's bilingual programme/the CLIL method in general?

1.3. Brief outline of study

The findings of my empirical research will definitely be the centre of attention in the present thesis. However, no case study will ever be valid without a theoretical grounding. Hence, in the chapter directly following this introduction, I will provide an overview of the history of lexical learning and the role of vocabulary in the context of *CLIL*. Moreover, I will also discuss the concept of lexical competence, as understood in three different theoretical works (Richards 1976; Carter 1987 and Henriksen 1999).

Chapter 3 is devoted to the methods, materials and test procedures employed. First, a few general comments will be made on the type of methodology used for the present study. Then the participants, their school setting, the process of data collection and the selection of test materials will be described. Finally, the evaluation and scoring of the lexical tests will be accounted for.

While presenting the overall test results, Chapter 4 also offers a detailed description of the statistical methods and analyses applied.
Chapter 5 is concerned with the results on the separate tests, which are analysed and discussed according to different factors, and compared and contrasted with the findings of the original work.
Chapter 6 then moves from the lexical tests to the outcomes of the learner questionnaires and tries to establish connections between the students' test performance and their extracurricular language experience.
This discussion will be complemented in Chapter 7, which focuses on further background factors, such as the learners' motivation, attitude and self-assessment, as well as the teachers' situation, aims and objectives with regard to *CLIL*.

As the final section, Chapter 8 summarizes the outcomes of the present study, together with some concluding comments and suggestions for future research.
The appendix provides examples of the information sheets, questionnaires and lexical tests used for the empirical fieldwork.

2. Theoretical framework

> "Without grammar very little can be conveyed,
> without vocabulary *nothing* can be conveyed." (Wilkins 1972: 111)

This quotation already hints at the enormous importance of vocabulary for the acquisition of languages (cf., e.g. Nation 1990: 2; Mobärg 1997: 201; Boyd Zimmerman 1997: 5). A learner involved in a survey on the role of specific aspects of English language learning (Pickett 1978) even goes as far as to state

> [V]ocabulary learning [...] to me always seems the key to any language. I am quite happy to pronounce badly and make grammatical mistakes but there is no escape from learning words. (Pickett 1978: 71)

This view is also echoed in other works. For instance, Widdowson (1978) argues that native speakers are well able to understand grammatically inaccurate utterances as long as the vocabulary is correct. On the contrary, statements that are grammatically 'impeccable' but contain incorrect vocabulary may cause considerable confusion. As Hedge (2002: 111) points out, "Sometimes the context of the utterance would lead the listener to question their first interpretation, but a chance response [...] gives [them] the wrong impression." In a like manner, Linnarud notes that a limited knowledge of the vocabulary of a foreign language not only gives rise to potentially embarrassing misunderstandings,

> but also to a less imaginative, dull and uninteresting composition with repetition of highly frequent lexical items and a simple and unelaborated theme. (Linnarud 1986: 3)

Considering these statements, it appears quite astonishing that for a long time, the teaching and learning of vocabulary, unlike that of phonology and syntax, has largely been neglected in the field of second language acquisition (compare Boyd Zimmerman 1997: 5; Long & Richards 1999: xi, 2001: xiii; Thornbury 2004: 13-14). In the following section, we will look at the development of lexical learning within different linguistic and pedagogic frameworks from the historical point of view. Next, we will attempt to explain the implications of these perspectives for the role of vocabulary in the CLIL classroom. The final section will then be concerned with the

question as to what constitutes lexical proficiency, more precisely, what is involved in 'knowing a word'.[2]

2.1. Lexical learning from the historical point of view

The *Grammar Translation Method,* which was introduced at the end of the 19th century and dominated foreign language instruction in Europe and the United States until into the 1920s (Boyd Zimmerman 1997: 5, 7), "had mastery of structures as [its] main goal" (Long & Richards 2001: xiii). Based on the teaching of Latin and Ancient Greek, the key element of the method was the translation of excerpts of classical literature (Rivers 1981; Howatt 1984). The target language was not supposed to be studied for practical use, but rather, learning was seen as a "mental exercise" (Boyd Zimmerman 1997: 5) that was indispensable for a sound humanistic education. In this connection, reading and writing of classical materials, standardized tests, and above all, detailed descriptions of grammatical rules, were of paramount importance (Rivers 1981; Howatt 1984). For the most part, "[l]anguage skill was judged according to one's ability to analyze the syntactic structure, primarily to conjugate verbs" (Boyd Zimmerman 1997: 6).

The development of vocabulary knowledge was regarded "as some kind of auxiliary activity" (Long & Richards 2001: xiii), facilitating the instruction of grammar. Very often, long lists of obsolete words and archaic patterns had to be memorized (Boyd Zimmerman 1997: 5-6). The vocabulary was selected according to one basic principle: It had to be suitable for the exemplification of morphological or syntactic structures. In other words, lexical items were only taught when they represented a grammatical rule (Kelly 1969; Howatt 1984: 136). Explicit vocabulary teaching was rare, and if it happened at all, the focus was mainly on etymology, which was perceived as "one way of discovering truth" (Kelly 1969: 130).

It was also at that time, when bilingual word lists, which had previously served as aids for the instruction of reading and grammar and were arranged into semantic fields, came to be consulted as common sources of reference (Kelly 1969).

[2] Even though frequency counts served as the basis for the design of the test materials under investigation, this aspect is not covered in the present thesis. For a detailed account of the underlying theoretical considerations, their practical application and the related test outcomes, confer with the corresponding chapters in the original study (Sylvén 2004: 29-30; 62-68; 82-83; 91-93; 100-101 and 228 respectively).

Even though the *Grammar Translation Method* continued to be used at the beginning of the 20th century, it was met with severe criticism from various sides. The most frequent complaint concerned the method's disregard of natural, everyday language (Boyd Zimmerman 1997: 6-7). In the 1880s, this wave of criticism eventually led to the establishment of the Reform Movement, an interest group, which limited the Latin-based aspects of teaching (Simensen 1998: 29-32) and placed a special emphasis on spoken language. Fluency was increasingly valued and generally understood as "the ability to accurately pronounce a connected passage and to maintain associations between a stream of speech and the references in the outside world" (Boyd Zimmerman 1997: 7).

Henceforth, words were no longer taught in isolation, or in order to illustrate grammatical rules, but embedded in the context of sentences or longer texts which lent themselves to practical use (Boyd Zimmerman 1997: 7). The main idea was to keep the language as simple and 'unexciting' as possible, since it was feared that appealing contents would easily distract the students from their compulsory learning tasks (Howatt 1984: 187).

One of the methods that developed out of the Reform Movement at the end of the 19th century, even though it was not based on the linguistic theory of any of its proponents, was the so-called *Direct Method* (Richards & Rodgers 2001). Its name was derived from the underlying concept that translation should be avoided and meaning expressed *directly* in the target language (Boyd Zimmerman 1997: 8). Interaction was seen as the key to successful language acquisition, and thus, the teaching occurred in small, intensive classes, allowing for a continuous series of questions and answers. Due to the fact that the method heavily relied on the exposure to everyday phrases and sentences, it was often misinterpreted as a 'natural' way of learning languages (compare Boyd Zimmerman 1997: 8 and Sylvén 2004: 27).

However, in the 'direct', monolingual classrooms, vocabulary acquisition was scarcely 'natural' or incidental. Most of the time, the contents of the question-response sequences were carefully selected, so as to 'point the learners into the right direction' and draw their attention only to particular lexical items (Howatt 1984: 201), which were then explicitly taught with the help of objects, gestures or visual representations. Abstract concepts were explained by means of simple descriptions referring to the students' existing word knowledge (Boyd Zimmerman 1997: 8-9).

Two other examples of teaching methods that were introduced in the first decades of the 20th century and used "a more or less systematic approach to vocabulary acquisition" (Sylvén 2004: 27)[3] are the *Reading Method* and *Situational Learning* (for details, see Boyd Zimmerman 1997: 9-10). The former of the two stressed the importance of lexical knowledge for the development of the learners' reading skills. It was recommended that teaching materials should be created on the basis of word-frequency lists, to ensure that the students only acquired and practised expressions that were of practical relevance for them (Boyd Zimmerman 1997: 9). Careful selection and adaptation were also the defining principles of the second approach. *Situational Learning* worked on the fundamental assumption that

> language should be taught by practising basic structures in meaningful situation-based activities; speech was the basis and structure that made speech possible (Boyd Zimmerman 1997: 10).

With its 'renewed' focus on grammar, the *Situational Teaching* movement was also closely linked to yet another method, which developed in the United States towards the end of World War II, namely *Audiolingualism* (cf., e.g., Richards & Rodgers 2001: 50-67). Influenced by the theories of structural linguistics and behaviourist psychology, the *Audio-lingual Method* was centred around the belief that "[f]oreign language learning [was] basically a process of mechanical habit formation" (Rivers 1964: 19-22, quoted in Richards & Rodgers 2001: 57). Thus, linguistic contents were primarily taught in spoken form. Intensive practice, learning by imitation, repetition and memorization were considered the most essential features of instruction (Richards & Rodgers 2001: 59). "Oral proficiency [was] equated with accurate pronunciation and grammar" (ibid: 58). Explicit grammatical explanations were rare. Instead, the students were taught syntactic structures by means of examples and numerous pattern drills (Boyd Zimmerman 1997: 10).

The teaching of vocabulary was subordinate to the development of oral skills (Richards & Rodgers 2001: 58). In a like manner as in the *direct* classrooms, the range of lexical items was supposed to be kept under strict control so that the learners could focus their undivided attention on the learning of the target structures (Brooks

[3] Following Sylvén (2004: 26), I will use the term *systematic* (as opposed to *incidental*) to refer to a language teaching situation "in which vocabulary, just like any other aspect of language is taught explicitly and in a conscious and structured fashion."

1964: 142; Boyd Zimmerman 1997: 11; Thornbury 2004: 14). New items were introduced "only in context" (Brooks 1964: 142), which meant through the drills, or because they were simple and "fitted neatly into the 'structure of the day'" (Thornbury 2004: 14).

At the end of the 1960s, *Audiolingualism* came under serious attack from both theoreticians as well as practitioners. On the one hand, the method was criticised for its deficient linguistic and psychological groundings. On the other hand, it was argued that the practical outcomes left much to desire. Students exclusively trained through oral drills in language laboratories very often found themselves incapable of applying their skills in real communicative situations in the 'outside world' (Richards & Rodgers 2001: 65). Hence, it was felt that "[w]hat was required was a closer study of the language itself and [...] the meanings and intentions [expressed by its] speakers and writers" (Howatt 1984: 280).

This major rethink was predominantly initiated by the publication of Noam Chomsky's groundbreaking work *Syntactic Structures* in 1957. Chomsky vigorously rejected the concept of language acquisition as mere habit formation (Richards & Rodgers 2001: 65). He claimed that every human being was endowed with a set of mental properties (*Universal Grammar*) that allowed them to acquire the language of their respective social environment within a particular period in their development (Lightbown & Spada 1999: 36). The innate, subconscious properties of a person's grammar were referred to as *competence*, their practical application as *performance* (Boyd Zimmerman 1997: 12; Lightbown & Spada 1999: 37).
While Chomsky emphasized the importance of linguistic creativity, innovation and the individuality of sentences (Richards & Rodgers 2001: 65; 153), he largely neglected "the nature of language use in real communication" (Boyd Zimmerman 1997: 12). This aspect was then covered by Dell Hymes (1972), who stressed the impact of *communicative competence* on the language learner.

After a period of extensive "adaptation, innovation, experimentation, and some confusion" (Richards & Rodgers 2001: 67), in which several alternative methods to *Audiolingualism* were proposed, in the late 1970s and early 1980s, the new communicative tendencies came to be generally known as the *Communicative Approach* or *Communicative Language Teaching* (*CLT*). Even though this term

subsumed a variety of different sub-forms and versions, they all strove towards one common goal that is, "communicative proficiency rather than [...] mere mastery of structures" (Richards & Rodgers 2001: 153). According to this perspective, meaning was more important than form. Fluency was given priority over accurate language use, and the development of the four skills (reading, writing, speaking and listening) had to be integrated into the teaching procedures from the very first day on. Basically, effective language learning meant the learning of effectual strategies to communicate (compare Finocchiaro & Brumfit 1983: 91-93).

Although individual linguists pointed out the communicative value of words (e.g., Widdowson 1978; Rivers 1983), "vocabulary has not been the focus of attention in communicative language research or methodology" (Boyd Zimmerman 1997: 13). Instead of focusing on lexical items in isolation, it was suggested to address the (lexical) system as a whole (Wilkins 1974 19-20, quoted in Boyd Zimmerman 1997: 14). Just as in connection with grammar, this was only achieved due to extensive exposure to the target language and activities allowing for the useful contextualisation of individual words (compare Boyd Zimmerman 1997: 14; Finocchiaro & Brumfit 1983: 91-93). Overall, the lexical 'aspect' of *Communicative Language Teaching* was based on the following underlying assumption:

> Since vocabulary development occurs naturally in L1 through contextualized, naturally sequenced language, it will develop with natural, communicative exposure in L2. (Boyd Zimmermann 1997: 14-15)

As the name already implies, this 'natural' view on second language acquisition (SLA) was also taken on and further developed within the framework of the so-called *Natural Approach*, introduced in the early 1980s by the American linguists Krashen and Terrell (1983). A supporter of Chomsky's (1965) ideas (see also Sylvén 2004: 28) and proponent of *CLT* (Boyd Zimmerman 1997: 15; Lightbown & Spada 1999: 40), Krashen set up his own theoretical model, comprising five different 'hypotheses', namely

(1) the *Acquisition-Learning Hypothesis*, stating that linguistic knowledge may be attained in two distinctive ways: on the one hand due to 'natural' *acquisition*, which resembles the process by which children 'pick up' their first language simply through exposure to it, and on the other hand, via conscious *learning* and close attention to structures and forms (Boyd Zimmerman 1997: 15; Lightbown

& Spada 1999: 38; Johnson 2001: 75 ff).[4] Acquisition was considered to be more important for the development of fluency:

> [M]any learners are quite fluent without ever having learned rules, while other speakers may 'know' rules but fail to apply them when they are focusing their attention on *what* they want to say more than on *how* they are saying it. (Lightbown & Spada 1999: 38)

(2) The *Natural Order Hypothesis* was based on the observation that, just as in the first language, grammatical structures in the second language tend to be acquired in a logical, foreseeable order, irrespective of their artificial sequence in the language classroom (Boyd Zimmerman 1997: 15; Lightbown & Spada 1999: 39). Thus, "the rules which are easiest to state (and [...] to 'learn') are not necessarily the first to be acquired" (Lightbown & Spada 1999: 39), as in the case of third person singular –s.

(3) Krashen third hypothesis, the *Monitor Hypothesis,* defines the minor function of the learned system, which basically consists in controlling, adapting and "polishing what the acquired system has produced" (Lightbown & Spada 1999: 38).

The most influential element of Krashen's *Monitor Model* seems to be[5]

(4) the *Input Hypothesis*, maintaining that language is only acquired when the learner is exposed to *comprehensible input* (Krashen 1985), which includes structures and forms just beyond their existing level of linguistic proficiency "and from which they can infer meaning" (Hedge 2002: 10). In later works, Krashen (1989; 1993a; 1993b) suggested that for more advanced students, reading for pleasure was the most favourable source of this kind of *i+1*-input.

Finally, a possible failure in the learner's attempt to achieve a high-level competence, even if *comprehensible input* was actually available, was explained by means of

(5) the *Affective Filter Hypothesis*, according to which negative emotions or 'affects' such as anger, anxiety or boredom might cause the learner to "'filter

[4] While fully acknowledging the relevance of this distinction, in the present thesis, I will nonetheless use the two terms *acquisition* and *learning* (mainly) interchangeably. Whenever the focus is placed explicitly on the one or the other, further explanations will be provided in the text.
[5] at least from the perspective of *CLIL* (see the discussion in section 2.2.)

> out' input, making it unavailable to acquisition" (Lightbown & Spada 1999: 39). On the contrary, a positive attitude and motivation would 'lower' the affective filter, thus allowing successful acquisition to take place (Lightbown & Spada 1999: 40).[6]

In contrast to other 'versions' of *CLT*, the *Natural Approach* acknowledged the enormous potential of vocabulary "as a bearer of meaning" (Boyd Zimmerman 1997: 15). Krashen & Terrell (1983: 155) pointed out that "acquisition [would] not take place without comprehension of vocabulary." Evidently lexical items were not supposed to be treated consciously, but, just as grammatical structures, they should attract the students' attention in the course of communicative activities. It was felt that this was the only way "to encourage true vocabulary acquisition" (Krashen & Terrell 1983: 156).

As the above outline has shown, in the history of SLA, vocabulary "certainly did not [always] attract the amount of research attention one might expect for such a basic linguistic building block" (Schmitt 1999: 3). Regardless of whether the underlying method or approach was 'formally' or 'communicatively focused', the majority of language classes were primarily designed according to grammatical syllabi (Thornbury 2004: 14). This tendency mainly results from the fact that, while syntax and morphology are undoubtedly governed by systematic rules, vocabulary appears to be scarcely more than a random or unstructured assortment of items (Sylvén 2004: 25; Thornbury 2004: 14). Taking into account the Chomskyan view (Chomsky 1957) that one single grammatical rule enables a speaker to create a variety of different sentences, it becomes understandable why grammar was (and very often still is) considered more productive than vocabulary: "Grammar multiplies, while vocabulary merely adds" (Thornbury 2004: 14).

It was not until the 1980s and 1990s that SLA researchers re-assessed this perspective and came to value the "centrality of the lexicon to language structure, second language learning, and language use" (Richards & Rodgers 2001: 132). As Thornbury (2004: 14) notes, this countermovement against the predominance of

[6] To a certain extent, these motivational and attitudinal factors will also be considered in the analysis of the present test results (see Chapter 7).

grammar was characterized by two essential developments: on the one hand, the introduction of lexical syllabi (e.g. Willis 1990), based on the assumption

> that the building blocks of language learning and communication are not grammar, functions, notions, or some other unit of planning and teaching but lexis (Richards & Rodgers 2001: 132),

and on the other hand, the acknowledgment of the significance of multi-word units, so-called *lexical chunks* (e.g. 'be in charge of', 'take a picture', 'For crying out loud!'). Both of these developments, within what has been generally referred to as the *Lexical Approach* (see Lewis 1993)[7], were triggered by insights gained from extensive analyses of computer-based language corpora, such as the Collins Birmingham University International Language Database (*COBUILD*) or the British National Corpus (*BNC*), containing more than 300 million words (Boyd Zimmerman 1997: 16; Richards & Rodgers 2001: 133). These studies in the new field of *Corpus Linguistics* stressed the relevance of lexical phrases, idiomatic expressions and collocations for the acquisition of (oral) fluency (Richards & Rodgers 2001: 133). Since lexical chunks were proven to be indispensable for everyday language use, as a consequence, they "should [also] be central to language teaching" (Boyd Zimmerman 1997: 17).

In this connection, a controversial question concerned the possible ways in which non-native students should be enabled to "internaliz[e] this massive inventory of lexical usage" (Richards & Rodgers 2001: 134). Not surprisingly, Krashen suggested that this could only be achieved through copious amounts of input (compare Richards & Rodgers 2001: 134). Other linguists assume(d) that the learners themselves "must take on the role of "discourse analyst[s]"" (Richards & Rodgers 2001: 136), thus working with computer databases and deriving collocations of lexical items from their occurrence in different texts. In this sense, 'lexical classrooms' "teach students to teach themselves" (Woolard 2000:35), which marks a significant shift away from the numerous pattern drills and excessive translations of the past.

Summing up, theoretical and pedagogical priorities have changed considerably throughout the course of history. According to Lewis (1993: 89), the linguistic focus

[7] although it seems that there is more than one specific, clear-cut version of lexically based language teaching (compare Boyd Zimmerman 1997: 16-17; Richards & Rodgers 2001 132 ff.).

of (E)FL classrooms has underwent a transition from "lexicalised grammar" to "grammaticalised lexis". Although this position may seem a bit exaggerated, it is certainly true that in today's classes and coursebooks,

> vocabulary is no longer treated as an 'add-on' [but m]uch more attention is given to the grammar of words, to collocation and to word frequency (Thornbury 2004: 14).

Schmitt (1999: 3) even notes that the 1990s recorded a surplus of books and articles on vocabulary. Nonetheless, hitherto, there are no clear-cut theoretical guidelines with regard to lexical learning in the context of *CLIL* (compare also Sylvén 2004: 26). The following section will offer a brief discussion on this particular issue.

2.2. Vocabulary in the CLIL classroom

2.2.1. Defining the context

So far, we have primarily looked at different methods and approaches for the teaching of English as a subject per se. In other words, we have outlined historical perspectives on language (vocabulary) learning in the context of what we will from now on call 'traditional' classrooms. With regard to the focus of the present study, in an Austrian school setting, 'traditional' means that in all non-language subjects, German is used as the medium of instruction. On the contrary, *Content and Language Integrated Learning (CLIL)* is a "dual-focused educational context" (*CLIL Compendium* 2001) in which subjects such as Physics, Biology, Chemistry, Maths and History are taught completely or partially in English or any language other than the students' mother tongue. Besides, this foreign language is also taught separately, just as in the traditional classes (compare Mewald 2004:42). The term CLIL classroom/lesson, however, only refers to the teaching in the non-language subjects.

Concerning the terminology for the method itself, *CLIL* is actually an umbrella term comprising a variety of different sub-forms and 'approaches' in which language teaching "is organized around the content or information that students will acquire, rather than around a linguistic [...] syllabus" (Richards & Rodgers 2001: 204; compare also Sylvén 2004: 9; Darn 2006). Depending on the target group, the respective setting and the amount of content vs. language instruction, these *CLIL* variants have been labelled *bilingual education, Language/English across the*

Curriculum, Immersion Teaching, Teaching Content through a Foreign Language, Englisch als Arbeitssprache (English as a Medium of Instruction), Language Medium Teaching, Content Based Instruction etc. (compare Richards & Rodgers 2001: 205 ff.; Mewald 2004: 42 ff.; Sylvén 2004:9). The list would be endless if we also included the names that individual schools or other educational institutions have given to their own specific realisation of one or the other type of *CLIL*. For instance, *Vienna Bilingual Schooling (VBS)* and the *Dual Language Programme (DLP)* are two, relatively new Austrian (or, more precisely, Viennese) versions of bilingual education. However, it is not the aim of the present thesis to provide a definition of each single programme.[8] For more information on different forms of bilingualism, see e.g. Mewald (2004: 42 ff.).

The CLIL context for the present study is the above-mentioned *Vienna Bilingual Schooling* programme, which will be described in detail in Chapter 3. Throughout the thesis, both the terms *CLIL* and *bilingual education* will be used when the emphasis is on the method in general. *VBS* refers to aspects that are characteristic of this particular school's bilingual programme. For the distinction of the two test groups, the expressions CLIL and VBS vs. traditional will be used interchangeably (for more details, see section 3.2.2.).[9] Furthermore (only) in the present discussion on the learning situation in CLIL classrooms, I will follow Mewald's (2004: 43) example and draw on the methodological principles of *Content Based Instruction (CBI)*, fully acknowledging that it is not exactly the same as *CLIL*.[10]

2.2.2. Theoretical backgrounds and their application in practice

As indicated above, the defining feature of CLIL teaching is that it integrates language acquisition into the learning of content matter, such as psychology, history or mathematics (compare Richards & Rodgers 2001: 205; Sylvén 2004: 9). I have explicitly used the terms 'language **acquisition'** and 'content **learning**', since according to Krahnke (1987: 65),

[8] In any case, clear-cut definitions would be extraordinarily difficult, since the above-mentioned 'methods' to a large extent overlap and intersect (see also Mewald 2004: 42).
[9] The terms will be written in *italics* when they specifically refer to the method or programme itself. In expressions, such as CLIL/VBS group/class/context/teacher ... , they will be written in normal print.
[10] As mentioned previously, Austrian CLIL students typically also have English as a separate subject. In contrast, *CBI* "is not usually backed up by additional FL tuition" (Mewald 2004: 43). However, for the present purposes, this distinction seems to be irrelevant, since our focus is on the acquisition/learning of vocabulary in subject lessons, and not in the context of EFL.

> [i]t is the teaching of content or information in the language being learned with little or no direct or explicit effort to teach the language separately from the content being taught.

CLIL or CBI syllabi are therefore more likely to be content-subject, instead of linguistic, syllabi (Richards & Rodgers 2001: 212). This basic feature is derived from the assumption that students learn a foreign language more successfully when they use it in a meaningful context, "*rather than as an end in itself*" (Richards & Rodgers 2001: 207). Since English is not only a medium for international communication, but also the predominant language in natural science and economics, it has been argued that content-based programmes provide a more adequate preparation for the learners' future academic studies and professional careers than traditional EFL classes do (cf., e.g. Richards & Rodgers 2001: 207; Darn 2006). In this sense, "*Content-Based Instruction better reflects learners' needs for learning a second language*" (Richards & Rodgers 2001: 207). This also implies that when the teaching content "*is perceived as interesting, useful, and leading to a desired goal*" (ibid: 209), the students' motivation to acquiring the target language and, in the ideal case, also the related success, will increase (compare also Darn 2006). Relevant contents, interesting topics and authentic teaching material, as well as the usage of language as "seen in real-life situations" (Darn 2006) therefore constitute they key elements of *CLIL* (compare also *CLIL Compendium*; Mewald 2004; Sylvén 2004).

In this sense, the method seems to bear close resemblance to the *Natural Approach* described above and indeed, as Baker & Prys Jones (1998: 650) point out, "Krashen's Monitor Model has [...] directly fed teaching practice, teacher training and classroom strategies." While due to their curricular demands, 'normal' (E)FL classrooms, no matter how much they rely on communicative procedures, will always be limited in their degree of 'naturalness' (see Mewald 2004: 58), *CLIL* claims to offer such a "natural situation for language development" (*CLIL Compendium*).

Thus, with regard to vocabulary, the basic assumption is that - just as grammatical structures - it is "absorbed" (Schmitt 2000: 15) more or less incidentally, that is, in the same way as by a child acquiring their first language (compare Sylvén 2004: 28). However, unlike the infant who for the first time tries to make sense of lexical items, the naturalistic L2 learner "already has experience of making relevant connections

between lexical forms and meanings in his/her L1" (Singleton 1999: 48), which "facilitate[s] the classification of reality offered by the L2" (ibid). As mentioned above, in this respect, the (Austrian) CLIL learners are even more privileged, since apart from their previous L1 knowledge, they also have a greater awareness of the target language due to the additional instruction in the formal EFL classroom.

What is of crucial importance for both, the child as well as the L2 learner in whatever kind of naturalistic environment, is an extensive amount of linguistic input, as established in Krashen's *Input Hypothesis* (see also Singleton 1999: 48; Sylvén 2004: 28-29; Mewald 2004: 57 ff.). Evidently, in an educational setting, oral language input is primarily provided by the teacher. For instance, Ellis (1990: 70) has discovered that teacher-talk accounts for as much as 70-80 percent of the classroom time. Hence, in the context of *CLIL*, it proves indispensable for the teachers to have a high-level proficiency in the language used as the medium of instruction (Nikula 2002: 463). According to Hartiala (2000: 56-79), CLIL teachers must be qualified in both, the target language **and** the respective content subject. In addition, they should have at least some basic skills in the students' mother tongue(s).
In the majority of Austrian CLIL schools, these requirements are indeed fulfilled, since very often, the subject teachers involved in *CLIL* with English as the medium of instruction also have a teaching degree in EFL. Another common practice is the teaching in teams, consisting of one native speaker of English and one Austrian subject teacher each (compare Mewald 2004 and the description of the VBS setting in Chapter 3). For more details on the situation of the CLIL teachers in the present study, see Chapter 7.

Apart from speech provided by native and non-native speakers, another essential source of input in the CLIL classroom is the language in schoolbooks and other teaching materials. As with other elements in the content-based approach, these resources are not linguistically focused but based on the contents of the related subjects (see also Richards & Rodgers 2001: 215). Corresponding to the *CLIL* pursuit of 'naturalness' and 'real-life affinity' (cf., e.g. *CLIL Compendium* 2001), the most essential claim in connection with materials seems to be their authenticity (compare Mewald 2004: 65 ff.). In this context, the term 'authentic' is ambiguous (Richards & Rodgers 2001: 215): On the one hand, it refers to coursebooks and other teaching aids that have been designed for the content instruction of native speakers,

and on the other hand, it comprises text types, such as newspaper articles, magazines, travel guidebooks and technical manuals, as well as films, TV broadcasts and other audiovisual materials "that were not explicitly produced for language teaching purposes" (Brinton, Snow & Wesche 1989: 17).

As regards the former group, a major problem consists in finding the right balance between linguistic and subject-matter complexity (Mewald 2004: 65-66). While for instance, schoolbooks for English students at the elementary level, 'language-wise' seem to be perfectly suitable for EFL learners in their first (two) year(s) of lower secondary, the respective contents would be too simple and undemanding and thus fail to meet the Austrian curricular standards. What is more, a Swedish study (Ljung 1990, quoted in Sylvén 2004: 34) has shown that school-textbooks provide students with vocabulary that is suitable only for classroom use, but not necessarily for every day communication in the target language. The use of 'realia' (i.e. the second group of authentic resources), is therefore more recommendable, especially for learners at higher levels (compare Mewald 2004: 65; Richards & Rodgers 2001: 215).
Due to a lack of published resources for bilingual instruction, for the most part, Austrian CLIL teachers produce their materials themselves, frequently drawing on online sources and, if available, on their English-speaking colleagues' support. The basis is, indeed, very often provided by authentic texts (of type 2), which are adjusted according to specific teaching aims and simplified, if need be (see also Mewald 322 ff.).

In connection with materials, it should also be noted that

> CBI [CLIL] views language use as involving several skills together. In a content-based class, students are often involved in activities that link the skills, because this is how the skills are generally involved in the real world. Hence students might read and take notes, listen and write a summary, or respond orally to things they have read or written. (Richards & Rodgers 2001: 208)

Above all, reading proves to be the most essential skill in *CLIL* (Darn 2006), as it is in other contexts of naturalistic L2 acquisition (Singleton 1999: 49). In line with Krashen (1989; 1993a; 1993b, see above), Darn (2006) argues that "[t]he best and common opportunities [for vocabulary acquisition] arise through reading texts." In this respect, *CLIL* heavily relies on the *Lexical Approach*, "encouraging learners to notice language while reading" (ibid). The impact of reading on the development of

lexical skills is also taken into account in the present study, where the focus will be on the students' reading habits in English **outside** the classroom (see Chapter 6).

Most typically, the language in reading materials designed for English-based *CLIL* comprises three types of vocabulary (compare Darn 2006): general English terms, such as 'to rely on', 'to deny' or 'diary', academic expressions (e.g. 'utilise', 'notwithstanding', 'thereto'), and subject-specific terminology, for instance, in Mathematics 'normal distribution', 'field extension' or 'tangent plane'. It should be pointed out that the incidental approach to acquisition only applies with regard to the first two categories. Technical terms, on the other hand, receive special attention and need to be learned consciously, in the CLIL lesson, just as in the 'traditional' subject classroom with the students' L1 as the medium of instruction (see also Sylvén 2004: 4, 35; Nikula 2002). However, this is not within the scope of the present study.[11]

Summing up the above discussion, in an Austrian CLIL setting, overall, the students' linguistic, and in particular, their lexical, development turns out to be influenced by two major sources:

> proficiency gained through formal instruction primarily based on principles of the Communicative Approach supplemented by the use of course books, and proficiency gained through CBI based on the Natural and the Lexical Approach thus setting aside the focus on form and emphasising the learning of language through the study of subject matter. (Mewald 2004: 83-84)[12]

Thus, it seems that in the interpretation of the CLIL students' test results, both aspects, acquisition **and** learning need to be taken into account (ibid: 84).

[11] According to Sylvén (2004: 4), the present test material contains exclusively general English vocabulary. However, in view of the above-mentioned three-part distinction, I would argue that some of these general terms are actually academic expressions. Undoubtedly, subject-specific items are not included.

[12] In addition to *CLT*, the *Natural* and the *Lexical Approach*, at a later stage in her study, Mewald (2004: 257) also mentions *Experiential Learning* and the *Multi-sensory Approach* as the most essential underlying concepts for *CLIL*.

2.3. What does it mean to know a word?

2.3.1. Some basic definitions

In the previous sections, we have looked at how vocabulary is presented and learned or acquired in different educational settings, in particular in the context of *CLIL*. What is even more central to the present study, is the question as to what constitutes lexical competence, or the knowledge of words. Evidently, this issue is as complex as the history of lexical learning itself.

First of all, the greatest problem consists in defining what a *word* actually is. In his introductory chapter, Singleton (1999: 8-38) tries to approach this question by comparing and contrasting different specialist and non-specialist views. However, at the end of this lengthy discussion, he voices the fear that now the concept has become completely elusive and almost impossible to grasp (Singleton 1999: 37-38). Since the emphasis of the present study is not so much on exact, theoretical backgrounds, as it is on empirical results, I will draw on a definition that reflects the most common, every-day understanding of the term *word*:

> UNIT OF LANGUAGE [...] a single unit of language which means sth and can be spoken and written (*Oxford advanced learner's dictionary*[13] 2000: 1490)

Yet, in the lexical tests under examination, certain items contain more than one "single unit of language". In this case, it seems more adequate to speak of *lexical items* or *lexemes*:

> **lexeme/lexical item** A separate unit of meaning, usually in the form of a word (e.g. 'dog'), but also as a group of words (e.g. 'dog in the manger'). (Widdowson 1996: 129)

Widdowson's definition closely corresponds also to Cruse's (1986) concept of the *lexical unit*, as quoted in Sylvén (2004: 35):

> a lexical unit must be at least one semantic constituent
> a lexical unit must be at least one word.

Thus, in the present thesis, the terms *lexical item, lexical unit* (and *lexeme*) will be used interchangeably to refer to both, individual words as well as phrasal expressions. The term *word* will only appear occasionally, whenever an isolated, one-part unit is meant (compare also Sylvén 2004: 35).

[13] Henceforth referred to as *OALD*.

From the above discussion, it also becomes obvious what the system of *lexis* and a person's *lexicon* (or the *lexicon* of a language) are (see Table 2:1). In the previous sections, we have used the expressions *lexical* vs. *vocabulary* (learning, acquisition, knowledge, skills ...) in the same way. As Kemmeter (1997: 22) notes, this is slightly problematic, because

> Die Bezeichnung Wortschatz[14] bedeutet die Gesamtheit der lexikalischen Elemente einer bestimmten Sprache (auch Fremdsprache). Er beinhaltet die Vernetztheit der lexikalischen Elemente mit allen Zusatzinformationen […].
>
> Der Vokabelbegriff betont eher isolierte lexikalische Einheiten, wie sie im Moment der Erstbegegnung auf ein "individuelles Potential" treffen.

According to this view, *lexis* is the more comprehensive concept, and *vocabulary* is a specific part of it. However, if we again consult the *OALD*, we will notice that, in fact, the term *vocabulary* denotes both, a complex network or linguistic system (i.e. the 'Wortschatz') as well as its individual constituents (i.e. 'Vokabel', or if the items are grouped into specific categories '(Fach-)Vokabular').

Table 2:1 *Basic definitions: 'lexis' and 'lexicon' vs. 'vocabulary'*

Vocabulary	**Lexicon**
1 [C, U] all the words that a person knows or uses: *to have a wide/limited vocabulary ◊ your active vocabulary (= the words you use) ◊ your passive vocabulary (= the words you understand but don't use)* [...] **2** [C] all the words in a particular language [...] **3** [C, U] all the words that people use when they are talking about a particular subject [...] **4** (also *informal* **vocab** [...]) [C, U] a list of words with their meanings, especially in a book for learning a foreign language	**1** (also the lexicon) [sing.] (*linguistics*) all the words and phrases used in a particular language or subject; all the words and phrases used and known by a particular person or group of people
	Lexis
	[U] (*linguistics*) all the words and phrases of a particular language **SYN** VOCABULARY)
(*OALD* 2001: 1447)	(*OALD* 2000: 739)

Given the resulting overlapses with *lexis* and *lexicon* (marked in coloured print), it therefore seems perfectly justified to use the adjective *lexical* (meaning "connected with the words of a language", *OALD* 2000: 739) **and** the noun *vocabulary* as modifiers for 'acquisition' and 'learning', 'test', 'skills', 'competence', and any other expression relating to this specific linguistic area. Besides, in the present study, both the acquisition of individual items (in Kemmeter's terms 'Vokabelerwerb')[15] as well as the students' overall lexical competence ('Wortschatzkompetenz') are touched upon.

[14] ~ Lexik (*Duden* 1996: 457)
[15] Kemmeter 1997: 21 ff.

2.3.2. Different views

In her attempt at defining the concept of vocabulary knowledge, Sylvén (2004: 35-39) draws on the works by Richards (1976), Carter (1987) and Henriksen (1999).[16] Since my thesis aims at replicating Sylvén's study, I will only briefly summarize these three views, especially focusing on their relevance for the present (and the original) tests.

According to Meara (1996b), Richards' paper on the role of vocabulary for language teaching is not only one of the first, but also one of the most influential works in this specific field. Concerning the question as to what contributes to the development of lexical proficiency, **Richards** (1976: 83, quoted from Meara 1996b) laid out **eight** basic **assumptions**:

1. The native speaker [of a] language continues to expand his vocabulary in adulthood, whereas there is comparatively little development of syntax in adult life.
2. Knowing a word means knowing the degree of probability of encountering that word in speech or print. For many words, we also know the sort of words most likely to be found associated with the word.
3. Knowing a word implies knowing the limitations imposed on the use of the word according to variations of function and situation.
4. Knowing a word means knowing the syntactic behaviour associated with that word.
5. Knowing a word entails knowledge of the underlying form of [a] word and the derivatives that can be made from it.
6. Knowing a word entails knowledge of the network of associations between that word and the other words in [a] language.
7. Knowing a word means knowing the semantic value of the word.
8. Knowing a word means knowing many of the different meanings associated with the word.

As Meara (1996b) notes, Richards' model clearly reflects the theoretical concepts and research concerns predominant in the mid-1970s. The first assumption is based on studies about L1 acquisition, and does not necessarily apply to foreign languages learned at later stages in life (compare Sylvén 2004: 36). The second assumption is closely connected to the beginnings of Corpus Linguistics and the insights gained from extensive analyses of computer-based language corpora (Meara 1996b). It

[16] Another, more recent model of word knowledge is also offered by Nation (2001: 26 ff.). However, this will not be included in the present study.

unites the knowledge of word frequency and collocational patterns (Sylvén 2004: 36). Assumption three is linked with the fields of pragmatics and discourse analysis, and assumption four derives from developments in syntactic theory. Assumption five falls into the category of morphology. On the other hand, the sixth assumption refers to the knowledge of synonyms, antonyms and other types of paradigmatic relations between individual words (compare Meara 1996b; Sylvén 2004: 36). The last two assumptions "concern the basic aspect of knowing a word, viz. what the word means" (Sylvén 2004: 36).

Richards does not cite any reasons for the internal hierarchy of these eight points. However, it seems strange that the two meaning-related categories should come last, while assumption one, which has little to do with the notion of lexical competence, and assumption two, which is similarly insignificant, appear on the top of the list (see also Meara 1996b; Sylvén 2004: 36).
As regards the lexical tests under study, only the notion of a word's "semantic value" (i.e. assumption eight) and its "different meanings" (assumption seven), as well as to some extent also syntactic features (assumption 4) are taken into account (compare Sylvén 2004: 36).

Eleven years after Richards, **Carter** (1987: 187) introduced the following **seven-part model** of lexical competence:

1. It means knowing how to use it productively and having the ability to recall it for active use, although for some purposes only passive knowledge is necessary and some words for some users are only ever known passively.
2. It means knowing the likelihood of encountering the word in either spoken or written contexts or in both.
3. It means knowing the syntactic frames into which the word can be slotted and the underlying forms and derivations which can be made from it.
4. It means knowing the relations it contracts with other words in the language and with related words in an L1 as well.
5. It means perceiving the relative coreness of the word as well as its more marked pragmatic and discoursal functions and its style levels.
6. It means knowing the different meanings associated with it and, often in a connected way, the range of its collocational patterns.
7. It means knowing words as part of or wholly as fixed expressions conveniently memorized to repeat – and adapt – as the occasion arises.

Carter's list seems to be more relevant to the concept of word knowledge. Nonetheless, his and Richards' ideas partly overlap. Richards' assumption two combines Carter's aspects two and four. Conversely, Carter's third point comprises both assumptions four and five of Richards' model. Carter's aspect six reflects the ideas expressed in Richards' assumptions six and eight, and to some extent, Richards' third and Carter's fifth aspect also coincide (see Sylvén 2004: 37).

Carter's model actually reflects a particularly high level of lexical competence, as it defines the understanding and usage of fixed phrases and expressions (i.e. aspect seven) as one of the basic requirements for knowing a word (Sylvén 2004: 37). In the present study, this aspect is covered in the two *cloze* tests (see sections 3.3.1. and 3.4.4. for details). Moreover, this particular test type also proves suitable for measuring word knowledge as defined in point one. According to Sylvén (2004: 38), aspects three, four and five are relevant for both the *words in context* (see 3.4.2.) and the *self-report* test (3.4.1.), and I would argue that in the latter, the first aspect, too, seems to be applicable. Carter's sixth aspect is important for all four test types. Only aspect two is not integrated in any of the materials or procedures applied in the present study (Sylvén 2004: 38).
Needless to say, in a limited test battery such as the present one, where the focus is exclusively on written language, it is virtually impossible "[t]o cover the entire scope of the notion of "knowing a word"'' (Sylvén 2004: 38).

As a third perspective on the concept of lexical proficiency, Sylvén (2004: 38-39) describes **Henriksen's** (1999) **"Three dimensions of vocabulary knowledge"**. This model combines both global (e.g. vocabulary *size* and *organization*, as proposed by Meara 1996a) and individual aspects of word knowledge (as those illustrated above), and emphasises the connection between the two (Henriksen 1999: 303-304). All three dimensions may be regarded as continua along which the development of lexical knowledge can be explained (Henriksen 1999. 315; Haastrup & Henriksen 2000: 222).

The first dimension, the *partial-precise knowledge dimension*, reflects "different levels of comprehension of the same lexical item" (Haastrup & Henriksen 2000: 222). It is an imaginary scale along which the degree of internalization of such a lexeme into the learner's mental word store is measured (compare Haastrup &

Henriksen 1998). On the *partial knowledge* end, we find the informant who recognizes the item and has a rough idea of its meaning. On the other hand, the *precise* end of the scale is represented by the learner who does not only know the exact meaning of the word and is able to pronounce, translate and rephrase it correctly, but who can also "identify the domain[17], provide word associations, and suggest other forms of the word" (Henriksen 1999: 305).

The second dimension, the so-called *depth of knowledge dimension*, involves the semantics, syntax and morphology of the word (see Henriksen 1999: 305-306; Haastrup & Henriksen 2000: 222). In order to 'reach the top end of the scale', the informant must be knowledgeable about the synonyms, antonyms and hyponyms (i.e. paradigmatic relations), as well as the collocational patterns (syntagmatic relations) of the word. Furthermore, s/he must also be familiar with the syntactic and morphological features of the item and (if any) with possible restrictions concerning its use (Henriksen 1999: 305-306).

As the name implies, dimension three, the *receptive-productive dimension*, reflects the continuum between mere word recognition and the productive use of the item in written or spoken discourse. Hence, this is the dimension "where the quality of the output is manifested" (Sylvén 2004: 39).

From their description, it becomes obvious that none of the above dimensions is actually a closed system on its own, but that the three continua are closely related and intertwined. Dimensions one and two both refer to different aspects of the semantization process (Henriksen 1999: 312). While they are directly linked with the acquisition of word knowledge,

> [d]imension 3 is essentially a control continuum that describes levels of access or use ability, which may be operationalized through different types of receptive and productive tasks (Henriksen 1999: 314).

Henriksen stresses that the distinction between receptive and productive knowledge is not a strict matter of either-or, but becomes blurred with the learner's gradual acquisition of familiarity (Henriksen 1999: 313). Sylvén (2004: 39) notes that in this connection, it would have been useful to consider also the pragmatic aspect, as it was

[17] Henriksen (1999) does not further specify the concept of "domain". For more details on this particular aspect, see Sylvén (2004: 38).

covered in Richards' assumption three (see above): "Being aware of the pragmatic properties of words is certainly important for the L2 learner." (Sylvén 2004: 39)

Summing up, unlike Richards, whose assumptions are only applicable to a certain extent, and Carter, who sets the level of word knowledge a bit too high (see Sylvén 2004: 37), Henriksen fully takes into account the different stages and problems involved in the development of lexical proficiency. Thus, her three-dimensional model proves to be most relevant for explaining the different tasks and procedures involved in the present test battery. Chapter 3 offers more details on this particular issue.

3. Research design and test administration

This chapter presents the methods, materials and test procedures applied in the present study. First, a few general comments will be made on the type of methodology used. Following this introduction, the test groups involved in this study, as well their educational background, in terms of school setting, will be described. The third part of the present chapter will then be devoted to the selection of test materials and the practical organisation of the lexical tests and questionnaires. Finally, the process of evaluation and test scoring will be accounted for.

3.1. Research methododology[18]

As already mentioned in the introductory chapter, the present thesis aims at replicating an earlier study conducted by the Swedish researcher Sylvén in the years 1999-2004. The main aim of this original work was to investigate the influence of *Content and Language Integrated Learning*, as opposed to traditional English language teaching, on Swedish learners' incidental lexical acquisition (cf., e.g. Sylvén 2004: 40).

For the purpose of the study, the CLIL method was regarded as a kind of 'permanent experiment', acknowledging that, as an approach to language teaching, it was still "new and unproven" (Sylvén 2004: 40), even though it cannot be regarded as an experiment in the strict and controlled sense (compare Sylvén 2004: 40-41). Given the special quality of this experimental situation, a combination of quantitative as well as qualitative elements was deemed most appropriate for analysing the test results and addressing the underlying research questions (Sylvén 2004: 41). As for the quantitative aspect, a battery of four different test types was developed, in order to provide a fairly detailed picture of the students' lexical proficiency (see Sylvén 2004: 51 ff.). This purely linguistic data was complemented by information gained from questionnaires concerning the students' and teachers' sociolinguistic backgrounds, their extracurricular English language experience as well as their attitudes towards CLIL in general. For the most part, these original materials have also been adopted in the present research context.
Based on this starting point, the test groups for the present study have been selected:

[18] As defined in Sylvén (2004: 40)

3.2. Test groups

In contrast to Sylvén's thesis, which provides a fairly representative picture of the "CLIL situation in Sweden as a whole" (Sylvén 2004: 41), the scope of this replication only allows for a relatively small group of participants. Instead of a selection of four schools, located in different parts of the country (Sylvén 2004: 42-43), one single school had to suffice for the present research purposes.

As in the original context, the most decisive criterion for participation was that the school should have an upper secondary section with at least one group of CLIL learners and a parallel group of traditional students, taught with German "as the normal medium of instruction and with English as a separate subject" (Sylvén 2004: 42). Since the English Department of the university collaborates with a series of Austrian, in particular Viennese, schools on various language projects, the selection process was quite straightforward. The decision was made in favour of a school that had already been involved in several other studies on bilingual education and was thus ready to provide access to test groups for the present evaluation.

The following section describes the specific context of the participating school and also explains why it was deemed appropriate for this study.

3.2.1. Description of the school context

As intimated above, the aim of the present thesis is not to give an overview of the effects of *CLIL* on the lexical skills of Austrian students in general, but to analyse the situation of one specific educational setting, namely a school that forms part of the so-called *Vienna Bilingual Schooling* programme:

Having its origins in the early 1990s, *VBS* is a comprehensive concept,

> covering the whole spectrum of the Austrian school educational system from Kindergarten to upper secondary school ("VBS Middle Schools" 2005: 1).

In the school under study, the first German-English bilingual classes for students from the age of 10 to 14 were introduced in the year 1998. At the same time, the school also conducted a trial run for the first Viennese bilingual upper academic school (Gierlinger 2002: 1).[19] Since the project was met with high approval, it was

[19] A so-called "bilingual upper commercial college (HAK)" ("VBS Middle Schools" 2005: 1) had already been introduced two years earlier (compare also Stadtschulrat für Wien: 22 August 2008a).

continued and extended in the following years, and finally also included into the regular school programme in 2005/2006 (compare VBS-Protokoll 2006). June 2002 saw the administration of the first bilingual Matura exams. Ever since, two bilingual upper secondary classes have graduated each year (Poisel 2003?[20]: 2).

In its 10-year history, the school has experienced an enormous increase in popularity. The overwhelmingly positive reactions to the extended, 8-year model of bilingual schooling also provided new incentives for other Viennese schools, which, up to that point, had only offered such language programmes for students at the lower secondary level. According to current information by the local school authority, at the present moment, as many as four Viennese grammar schools have VBS classes from grade 5 to 12 or 13 onwards (Stadtschulrat für Wien: 22 August 2008b).

One of the most essential distinguishing features of the VBS programme is that, unlike other (Austrian) sub-forms of CLIL instruction, it is intended, not only for Austrian students who are especially gifted and particularly interested in English or other foreign languages, but also for students with English as their L1 or medium of communication, who already have some basic skills in the German language ("Bilinguale Schule": 22 August 2008).[21]

Just as the VBS programme itself is a relatively new and innovative educational concept, so are the methods and techniques applied in the bilingual classes. Even though the teaching contents for both the bilingual lower secondary as well as the upper secondary closely adhere to the guidelines of the Austrian national curriculum, the VBS teachers are still relatively free with regard to the actual realisation of specified teaching aims ("Bilinguale Schule": 22 August 2008).

Table 3:1 presents the most important facts and figures of the school's specific approach to *VBS* or *CLIL* in general:

[20] Unfortunately, the exact year of publication is unknown (either 2003 or 2004).
[21] Henceforth, these students will be referred to as 'international students'.

Table 3:1 *Details of the participating school*[22]

Bilingual education is applied from grade level	5 and/or 9 (10)
Programme	Lower secondary: natural sciences Upper secondary: natural sciences or modern languages
Admission criteria	Orientation talk including various language tasks
Subjects taught bilingually	All except German/English Music, Arts and Physical Education are taught in either of the two languages.
Total amount of bilingual instruction	100% ca. 50% English – 50% German
Native speaker teachers	4

(Adapted from Sylvén 2004: 44)

As mentioned previously, at this particular school, students may start their bilingual education either at the beginning of lower secondary (grade 5), or during the first two years of upper secondary (grade 9 and 10). In both cases, the programme is based on the curriculum of the so-called 'Realgymnasien' with a special focus on natural sciences. Besides, the bilingual upper secondary is also offered in the form of a 'Gymnasium' with French as an additional foreign language (cf., e.g. Poisel 2003?: 2). Parallel to the VBS classes, at each grade level, the school has at least one, in most of the cases two traditional forms, which are either taught according to the curriculum of the 'Realgymnasium' or that of the 'Gymnasium' with French or Latin from grade 7 onwards ("Unterrichtsorganisation" : 22 August 2008).

The first step towards admission to a VBS class consists in an orientation talk, which "gives the staff the opportunity to assess the child's ability in English and German" ("VBS Middle Schools" 2005: 1). At the upper secondary level, this assessment comprises a listening and reading comprehension, as well as a grammar check-up, a writing task and several speaking exercises. ("Bilinguale Schule": 22 August 2008) Students from regular lower secondary classes, or from other middle schools or academic schools, who intend to enrol in the bilingual upper secondary have the opportunity of taking crash courses, in order to prepare for the entry 'exams' and the requirements of the bilingual school setting.

[22] Based on "Bilinguale Schule" (22 August 2008).

As for the actual teaching itself, the co-operation of teacher-teams, each consisting of one Austrian subject teacher and one English native speaker teacher, is an essential component in all bilingual Biology, Physics, Chemistry, Mathematics, Geography, History, Psychology and Philosophy lessons at this particular Viennese school. In the so-called 'subsidiary subjects', the teaching contents are presented in language-specific chunks throughout the whole school year, i.e. certain topics are predominantly taught in German whereas others are mainly introduced in the English language (cf., e.g. Poisel 2003?: 2). Mathematical tasks, on the other hand, are alternately assigned in English and German during one and the same lesson (personal observation in the course of an internship in winter term 2005/06).
In Music, Arts and Physical Education, the teaching is conducted in either of the two languages. As a matter of course, the same also applies to German and English as separate subjects, however taking into account that each of the two languages is on the one hand taught as a mother tongue and on other hand as a second or foreign language. Upon approval by the corresponding church or officially recognised religious community, the subject Religion may also be taught bilingually ("Bilinguale Schule": 22 August 2008).

Although the actual amount of the English language input, as well as the number of lessons taught by native speaker teachers slightly varies from class to class, as a guiding principle, it has been established that both languages are used in approximately equal shares ("Bilinguale Schule": 22 August):

> Diesen Maßnahmen liegen folgende Überlegungen zugrunde. Die Zusammenarbeit österreichischer und englischsprachiger Lehrer macht es möglich, daß [*sic!*] sowohl fremdsprachige Kompetenz als auch das Bildungsziel erreicht werden können. ("Bilinguale Schule" : 22 August 2008)

At the time of the test administration (school year 2006/07), the school employed a total of four native speakers of English, who, apart from "help[ing] CLIL teachers with their English" (Sylvén 2004: 44), taught a variety of different subjects, ranging from History, Science and English to Computer Studies, Arts and Film.

Corresponding to the VBS slogan "Taking a world view. Einen Blick für die Welt bekommen." (see, e.g., "VBS Middle Schools" 2005: 1), the bilingual teaching staff at the school under study has set themselves the ambitious goal of providing a sound "bilingual German/English general secondary education to a culturally diverse

student population" (VBS Middle Schools" 2005: 1), and supply their students with "the necessary linguistic skills and the educational knowledge necessary to compete successfully in the international workplace" ("VBS Middle Schools" 2005: 1). In this connection, *Content and Language Integrated Learning*, along with enhanced second or foreign language instruction, is of equal importance as an overall "education towards tolerance and openmindedness." ("Bilinguale Schule": 22 August 2008) The school philosophy, pledging an atmosphere of intensive intercultural exchange and mutual understanding ("Bilinguale Schule": 22 August 2008), is not merely a theoretical concept, but constitutes an integral part of the students' daily experience, as the following statement by the schools' VBS coordinator shows:

> Auf Grund der Vielfalt der Herkunftsländer der SchülerInnen zeichnen sich alle VBS-Klassen durch ein hohes Maß an interkultureller Kompetenz aus. Für sie ist es selbstverständlich, mit Vertretern fremder Kulturen zusammenzuarbeiten, ihre Freizeit zu verbringen, zahlreiche Feste gemeinsam zu feiern und deren Sitten und Bräuche hoch zu schätzen. Vielleicht gerade wegen der bunten Mischung verschiedenster Kulturkreise bilden sich enge Freundschaften und Klassengemeinschaften. Allgemein konnten wir beobachten, dass in diesen Klassen weniger Aggression vorkommt, dass sich die SchülerInnen gegenseitig nicht nur in schulischen Angelegenheiten helfen, sondern sich auch dafür verantwortlich fühlen Neuankömmlinge in die „österreichische" Schulpraxis einzuführen und Anfangsschwierigkeiten zu überbrücken. Sie machen begeistert bei Projekten vor allem mit kulturvergleichenden Schwerpunkten mit und organisieren gemeinsam mit LehrerInnen und Eltern der VBS-Klassen jedes Jahr ein Halloweenfest Ende Oktober und ein Multikultifest am Ende des Schuljahres. (Coord.VBS, unpublished article: 4) [23]

Documentations of these projects can be found throughout the whole school building, in the corridors, just as well as in the classrooms, thus contributing to an overall warm and friendly atmosphere, in which also university students, researchers and other people from outside of the school community are most welcome (personal observation in the course of an internship in winter term 2005/06).

Given the appropriateness of the educational background and the students' (and teachers') openness towards innovative projects and experimental teaching, this particular school seemed to provide an adequate environment for the present research project to take place.

[23] Quote anonymized

3.2.2. Selection of the participants

Having finally decided on a school setting, the question remained as to which specific participants should be involved in the present study. Taking into account that the original work analysed the lexical development of CLIL students and traditional students in the course of two years of their upper secondary education (Sylvén 2004: 46)[24], the first idea was to imitate the longitudinal outline of the study and thus, focus on two test groups at both grade levels 9 and 12 respectively. However, after an initial meeting with the school's VBS coordinator (in November 2006), these plans had to be re-assessed, since, for various pragmatic reasons, the desired target groups were not considered suitable for the research context. Instead, the coordinator suggested two classes at grade level 11, one of which was taught according to the VBS programme, and the other one received instruction in the traditional manner. [25]

As Table 3:2 illustrates, with a total number of 21 students, the VBS group was considerably larger than the traditional group, comprising only 12 students. Accordingly, in the former class, the ratio between the two genders was more balanced: 10/11 for the VBS females/males compared to 9/3 for the corresponding traditional subset. All of the students in the regular 11th form were native speakers of German, whereas seven participants of the CLIL class were so-called 'international students', four of whom had English, the remaining three Albanian, Croatian and Swedish respectively, as their mother tongue. The VBS class followed a teaching programme with a specific focus on natural sciences. The traditional class, on the other hand, was more specialized in modern languages, in particular French.

[24] In Sweden, upper secondary education starts with grade level 10 and continues up the 12th form (ESTIA 2003).

[25] Henceforth, in the description of the two test groups, the following terms will be used interchangeably and without bias: VBS = bilingual = CLIL; traditional = conventional = regular = normal students/group/class/(sub-)set/assemblage. 'Native Austrians' or 'typical Austrian students' are students who were born and raised in Austria and who have German as their mother tongue.

Table 3:2 *Details of the selected test groups*

Group	Grade level	Number of participating students	Female/male ratio	Students' L1	Programme
VBS	11	21	10 females, 9 males	14 German, 4 English, 1 Albanian, 1 Croatian, 1 Swedish	Natural sciences
Traditional	11	12	9 females, 3 males	all German	Modern languages

Throughout the entire test sequence, the number of students remained more or less stable. Only one of the traditional students happened to be absent during the last test.

In the following section, the organisation and practical administration of the test sequence will be described in more detail. First, the actual choice of test materials will be accounted for:

3.3. Choice of materials and test administration

As indicated at earlier stages, fortunately, Dr. Sylvén has kindly agreed on the re-use of her test sheets and questionnaires for the purpose of the present study. Since the scope for this MA thesis is more limited, only certain parts of the original materials have been adopted. The following paragraphs briefly outline the selection of these resources. Further details on the individual test types will be given in section 3.4., where the scoring mode and test evaluation are accounted for.

3.3.1. Lexical tests

As a longitudinal study, spanning a period of two full school years, Sylvén's (2004) work comprised three individual test rounds, one of which was conducted at the beginning of the 10th form and the other two at the end of year 10 and 11 respectively. On each of these occasions, a battery of four different vocabulary tests was used, namely

- a *self-report* test, in which the students had to estimate their knowledge of particular lexical items,

- a *words in context* test, in connection with a reading task, where the learners were asked to derive the meaning of a set of words from their use in a newspaper article,
- a *multiple choice* test, and
- a *cloze* test, focusing on lexical phrases and idioms (compare Sylvén 2004: 6-7 and 51 ff.).

This strategy was employed in order to make sure "that the test[ing was] not biased towards one particular method or to one particular sort of learner" (Alderson, Clapham & Wall 1995:45).

Each of the four test types consists of 30 lexical items, which are fairly equally distributed among word-classes, frequency groups and, with regard to the *self-report* test, also across the alphabet (compare Sylvén 2004: 52-58). The selection of the test items followed a highly complex scheme: First, a comprehensive word list (Thorén 1976) was consulted and subsequently compared to the frequency ratings provided by the *Collins COBUILD English dictionary* (1995).[26] On the basis of these two sources, a provisional list of possible test items was compiled, which was then submitted to a small group of students for a trial test run (cf., e.g. Sylvén 2004: 66 and 73). After a few additional modifications, the ultimate list of test items was established. "For a fair comparison between test rounds, the level of difficulty was kept the same in all three rounds" (Sylvén 2004: 66), meaning that the distribution with regard to frequency ratings was similar in all tests of the same kind. Besides, one item of the *multiple choice* test and ten out of the 30 *cloze* test items were so-called *anchor items* and thus identical in all test rounds (see Sylvén 2004: 66 and 68 respectively). For a detailed account on the selection of the lexical items and the actual distribution pattern, see Sylvén (2004: 62-68 and Tables 4:5 and 4:6).

Apart from providing a more general picture of the learners' lexical development over a longer period of time, each individual test type also measures a particular aspect or dimension of the students' lexical competence. Referring to Henriksen's (1999) "Three dimensions of vocabulary development" (see section 2.3.2.), Sylvén (2004: 53-58) argues that the *self-report* test, the *words in context* test, and especially also the *cloze* test, primarily focus on the levels of *partial to precise*, as well as

[26] Henceforth referred to as *COBUILD* (1995).

receptive to productive knowledge. In contrast, the *multiple choice* test is only suitable for testing the former of these two dimensions (Sylvén 2004: 55). The so-called *depth of knowledge dimension*, which is supposed to reflect the learner's "knowledge of semantics, syntax and morphology" (Sylvén 2004:55), does not seem to be covered by any of the four test types, even though "the *self report* could easily be expanded into measuring also this dimension" (Sylvén 2004: 54).

When the aims and objectives for this MA thesis had been defined, it was quite obvious that, in order to obtain reliable results, all four of the above-mentioned test types had to be included in the present research context. However, instead of the original three test rounds, the battery was reduced to a single sequence, conducted in the middle of school year 2006/2007. Given the fact that, at this point, the Austrian students were already in the third year of their upper secondary education, the materials of Sylvén's test round III seemed to provide the most adequate basis for comparison. In addition to the third 'versions' of the Swedish *self-report* (see Sylvén 2004: Appendix 3, 270-271), *words in context* (Sylvén 2004: Appendix 3, 274-276), *multiple choice* (277-278) and *cloze* test (272-273),[27] a second 'gap-filling exercise' was included from Sylvén's test round II (see Sylvén 2004: Appendix 3, 262-263). [28] The reason for this selection is that, according to the students involved in the Swedish pilot study, the *cloze* tests present the most difficult tasks in the entire test sequence (cf., e.g. Sylvén 2004: 58) and are thus quite effective tools for distinguishing the highest performing students from their intermediate or low-achieving peers.

Having decided on this five-part test battery, as a next step, the numbering of the original sheets, as well as certain test instructions had to be changed in order to fit into the Austrian context. With regard to the sociolinguistic questionnaires, a few more modifications were required. The following section illustrates the particular design of these question sheets:

[27] Henceforth referred to as *cloze test 1*.
[28] Henceforth referred to as *cloze test 2*.

3.3.2. Questionnaires

Since in today's (Swedish) society, the influence of the English language is by no means restricted to the classroom setting, in the original study, the linguistic data gained from the above-mentioned vocabulary tests was supplemented by a survey on the students' extracurricular experience with English (cf., e.g., Sylvén 2004: 5, 41 and 69). The related question sheets were distributed at the end of test rounds I and III respectively (Sylvén 2004: 47). They addressed issues such as the students' mother tongue, their social background, time spent in English-speaking countries and, above all, leisure interests regarding the English language (for full details and examples of these materials, see Sylvén 2004: 69-70 and Appendix 4, 279-288). In this connection, the students were asked to indicate whether they read English texts/watched English movies/TV programmes, with and without Swedish subtitles/visited English-speaking websites on "a daily, weekly, monthly or annual basis" (Sylvén 2004: 69). Another central concern was the way the learners felt about speaking in front of an audience in Swedish and English respectively.

In the second version of the questionnaire, the students also had to indicate whether they used the English language for any conversations outside of the CLIL or EFL classroom (Sylvén 2004: 69). Moreover, they were asked about whether they wrote any e-mails, letters, chat entries or other kinds of texts in the English language. Besides, the question concerning the students' Internet or PC habits was complemented by an additional one, focusing on role-plays, strategy games and other forms of electronic entertainment including instructions in English (Sylvén 2004: 70).

In addition to their leisure preferences, the learners also had to indicate how they evaluated their own progress in the fields of English lexis and grammar, as well as speaking, writing and understanding written and spoken English, in the course of their first two years of upper secondary education. Furthermore, they were asked about their marks in the subjects Swedish, English and Mathematics, and about how they liked, and whether they would recommend their particular choice of programme (see Sylvén 2004: 70).

In order to gain even more details on the educational backgrounds of the Swedish participants, on the final test occasion, questionnaires were also distributed among the teachers (see Sylvén 2004: 70-71 and Appendix 4, 289-291). Apart from general

issues, such as mother tongue, gender and year of birth, a special emphasis was placed on the teachers' formal education and proficiency in the English language, as well as on their teaching experience in connection with *CLIL*. Besides, the teachers were also asked about how **they** estimated their students' development in the above-mentioned areas. This offered quite interesting opportunities for discussion. In an open-ended section at the end of the questionnaire, the teachers were finally invited to state their personal thoughts, views, aims, suggestions and perspectives with regard to the CLIL method (Sylvén 2004: 71).

Given the specific, multicultural setting of the Austrian VBS school, a comparable analysis of background factors was considered indispensable. In order to design questionnaires that were suitable for the present research context, as a first step, the Swedish materials had to be translated into German. Next, a decision was made as to which of the original questions should be adopted or rejected. Apart from a few minor modifications regarding layout and structure, on the whole, the student questionnaires were kept in the same form as they were in Sylvén's test round III.

Just as in the original study, a separate question sheet was used for the VBS students and the traditional students respectively (for an example of both questionnaires, see Appendix 2 in the present thesis). However, these two questionnaires only differed with regard to one particular question, that is, question (2): While the VBS students were asked about the reasons for their choice of programme ("Warum hast du dich für den VBS-Zweig entschieden?"), the corresponding question for the traditional students read, "Hast du überlegt, den VBS-Zweig zu besuchen? – Ja/Nein, weil ...".

Apparently, certain formulations had to be adapted in order to fit into the Austrian school context. For instance, pertaining to the question "Welche Ausbildung haben deine Eltern abgeschlossen?", instead of the original distinction between elementary/compulsory school, grammar school and university, the following answer options were offered:

Figure 3:1 *Student questionnaire, answer options, question 11*

	Mutter	Vater
Pflichtschule		
Lehre		
Fachschule oder Berufsbildende Mittlere Schule (ohne Matura)		
Allgemeinbildende oder Berufsbildende Höhere Schule (mit Matura)		
Kolleg, Akademie		
Universitäts- oder Fachhochschulstudium		

A similar adjustment had to be made in connection with questions (14) "Wie würdest du die Verbesserung deiner Englischkenntnisse in folgenden Bereichen seit deinem Eintritt ins Gymnasium beurteilen?" and (16) "Wie bist du mit der Wahl deines Schultyps/Unterrichtsschwerpunktes zufrieden?". Here, the original evaluation had been carried out according to a five-point scale, ranging from "1 = [very] poor" to "5 = very good" (Sylvén 2004: 70). However, this would have been slightly confusing for the participants in the present study, given that the Austrian grading system functions exactly in the reverse order, and so, the instruction was transformed into "Bewerte nach Schulnoten".[29]

As for the teacher questionnaires (once again, there were two of them, one in German and one in English for the native speaker teachers, see Appendix 2), the above-mentioned questions concerning gender, mother tongue, English language competence, formal education and teaching experience were more or less adopted in the original format. Furthermore, just as their Swedish colleagues, the Austrian teachers were asked to estimate the students' improvement in the fields of English vocabulary, grammar, reading, writing, speaking and listening skills, as achieved by means of *Content and Language Integrated Learning*. Again, this evaluation was based on the scale of the Austrian marking system.

In order to gain more detailed insights into the daily routines and practices at the VBS school under examination, the teachers' former English language experience

[29]Nonetheless, when the corresponding outcomes of the two studies are compared, the figures will be calculated according to the Swedish system (see Chapter 7).

(e.g. due to some work or study experience abroad), as well as their attitudes and opinions about the programme and *CLIL* in general, an additional block of questions was developed, which followed the example of a previous study by an Austrian researcher (compare Mewald 2004: Appendix 2, 22-28). This new section included questions, such as:

- *Which foreign languages do you speak and at what level? (beginner - intermediate – advanced – native speaker level)*
- *Have you worked or studied abroad? If yes, in which country/-ies? How long?*
- *In how many classes do you teach bilingual lessons/How many hours per week/ In which subjects?*
- *Which tasks do you have within the VBS project? What are the main teaching methods/strategies you are using in your bilingual lessons?*

In addition, the teachers were asked to evaluate the VBS programme according to statements, such as:

- *BI (= bilingual instruction) makes the subject lessons richer.*
- *There is sufficient in-service training for native speakers teaching BI.*
- *There are sufficient teaching materials for BI.*
- *The students seem to miss out on/ improve their subject knowledge due to BI.*
- *BI increases the students'/my own motivation.*

The corresponding answer options were "true", "mainly true", "partly true" and "not true".

As in Sylvén (2004), the questionnaire was completed with an open-ended section, where the teachers could express their approval or criticism of the VBS programme, and state their personal aims and future perspectives for the field of CLIL instruction.

We now turn to the practical administration of the questionnaires and lexical tests described above.

3.3.3. Test administration

Before the actual test sequence could be conducted, a series of organisational and ethical issues had to be considered. Having decided on the test groups and materials, an outline of the research project was submitted to the school's headmaster and the Viennese Board of Education. Furthermore, the participating students received an information sheet including a short description of their 'research task', together with

the note that the test results would remain anonymous and had no influence whatsoever on the students' marks. The same letter was also forwarded to the parents, who were asked for their permission to test the students' lexical skills in the English language. Examples of these information sheets are included in Appendix 1 at the end of the thesis.

Following these organisational matters was a two-month period of uncertainty, which ended by mid-February 2007, when the local school authorities finally gave their official consent to the present research project. The next step was to approach the two form teachers, so that the schedule for the actual testing could be arranged. In accordance with the Swedish original, for each of the five lexical tests (and the questionnaire respectively), a time limit of 30 minutes was planned, thus amounting to a total of at least three full school lessons per class. At first, the form teachers thought that this would be simply unfeasible, but, fortunately, they managed to organise a series of spare lessons, in which I should stand in for the absent subject teachers and conduct the questionnaires and lexical tests. The entire test round was then performed during a period of approximately three weeks at the beginning of the summer term of school year 2006/2007 (February/March 2007). Table 3:3 shows the exact test sequence for both groups:

Table 3:3 *Sequence of lexical tests and questionnaires according to test groups*

	VBS group **(February 2007)**	**Traditional group** **(March 2007)**
Day 1	Questionnaire	Questionnaire Self-report test }
Day 2	Self-report test Words in context test Cloze test 1 }	Words in Context test } Cloze test 1 Multiple choice test
Day 3	Multiple choice test } Cloze test 2	Cloze test 2

(... } = two consecutive school days)

As illustrated above, the order of the individual tasks was equal in both forms: After a brief introduction (in German), during which the students were informed about the organisation of the test procedure, first, the sociolinguistic questionnaires were handed out. Since, in the VBS group, this was done in the second half of one school lesson, there was no time left for an additional task, and so the *self-report* test had to

be postponed to 'day 2', which was, in fact, almost one week later. In the traditional group, on the other hand, the *self-report* test directly followed the questionnaire. In both the CLIL class and the regular class, the *words in context* test and *cloze test 1* were conducted during one single day, in February and March 2007 respectively. On the same day, the traditional class also completed the *multiple choice* test. *Cloze test 2* was the last test in the entire sequence and thus performed on the respective 'day 3' for each of the test groups. The main reason for 'separating' the two most difficult tests was to allow the students some time to relax and refresh their knowledge, if necessary (see also section 5.5.5.).

All of the tests were conducted by myself in both test groups. Prior to each individual task, the instructions were read out and explained to the students. Once again, the learners were reminded that the results would be treated as confidential, but in order to assign each individual test to the respective participant, they should, nonetheless put down their names on the sheets. In the beginning, some of students did not take this directive quite seriously, as they invented fantasy names and wrote down silly remarks on the questionnaires. However, this problem was easily resolved, when I stressed the importance of their contribution and encouraged them to do their best.

Apart from this minor incident, the test sequence could be carried out without hindrance. As indicated above, only one student happened to be absent during the last day of the test administration. Thus, absenteeism hardly presented any major problems, as it did in the Swedish study (compare Sylvén 2004: 48-51).

What proved to be slightly more problematic was the collection of the teacher questionnaires. On the first test occasion (i.e. 'day 1' in the VBS group) 20 of those were distributed among the teachers and native speakers at this particular school. Some of these participants turned out to be very co-operative, as they immediately filled in their question sheets and showed considerable interest in the present research project. However, this overall eagerness declined as soon as I had completed the last lexical test and was no longer present at the school. Even though several teachers had promised to hand in their questionnaires later, they were only willing to do so upon the VBS coordinator's repeated request. By the beginning of May 2007, which was the utmost deadline for returns, eventually, I had received at least 11 out of the original 20 questionnaires. Although, with 55 percent, the return

rate for the teacher sheets is relatively low, it still seems to be sufficient for the present purposes, given the small scope of the study, and the fact that its main focus is on the students' achievement in the lexical tests.

Before the corresponding results will be presented, the process of correction and test scoring has to be accounted for.

3.4. Test scoring

In her thesis, Sylvén notes that

> [i]deally, more than one person should be involved in the scoring of tests that are not purely psychometric in design. For pragmatic reasons [...], this is not always possible, as in the case of the present study. In order to avoid any bias in the scoring of the non-psychometric tests [...], all possible solutions for each item were listed. This list was consulted for any answer that was not clear-cut correct or incorrect. (Sylvén 2004: 58)

Taking into consideration that this study is a replication of the aforementioned study, the list of possible solutions would have been extremely useful for corrective procedures within the current research context. Unfortunately, this was administratively unfeasible, thus, a separate catalogue of solutions was compiled for each of the five vocabulary tests integrated in the present work. For this purpose, a series of different monolingual as well bilingual sources were consulted, for instance the *Oxford advanced learner's dictionary* (*OALD*), the *Macmillan English dictionary* (henceforth referred to as *MED*), the *Oxford collocations dictionary* (henceforth referred to as *OCD*), the *LEO Online Deutsch-Englisches Wörterbuch* (henceforth referred to as *LEO*), the *Roget's new millennium thesaurus* (henceforth referred to as *Thesaurus.com*) and the *BNC online sampler*, to enumerate some of them.

In the following section, the scoring scheme of each individual test type will be accounted for. In addition, wherever required, any modifications to the original will be explicitly illustrated.

3.4.1. The self-report test[30]

In the *self-report* test, the students had to judge their own knowledge of a set of target words and provide practical evidence for their estimation by writing either a synonym, a translation, or a full sentence using the item under examination. The underlying structure of this test type can be traced to the so-called *Vocabulary Knowledge Scale (VKS)*, an instrument developed by Paribakht & Wesche (1997) in order to discern different stages in students' acquisition of lexical knowledge.

The *VKS* is based on a 5-point scale that merges aspects of self-perception and linguistic performance to assess estimated competence and actual knowledge of a set of words within a written test form (Paribakht & Wesche 1997: 179). Table 3:4 illustrates the five answer categories:

Table 3:4 *Answer categories in the self-report test*

Self-report categories	Level of knowledge
A I don't remember having seen this word before.	*total unfamiliarity* word and meaning unfamiliar
B I have seen this word before, but I don't know what it means.	*partial unfamiliarity* word familiar, meaning unfamiliar
C I have seen this word before and I think it means _____. (synonym or translation)	*partial recognition* word familiar, word meaning (synonym or translation) is guessed
D I know this word. It means _____. (synonym or translation)	*total recognition* word familiar; correct meaning (synonym or translation) is given
E I can use this word in a sentence: _____. (Write a sentence.) *If you do this section, please also do section D.*	*total recognition and correct word use* as in D In addition, the word is used with semantic and grammatical correctness in a sentence.

(Adapted from Paribakht & Wesche 1997: 179-180 and Sylvén 2004: 52)

Corresponding to the scale ratings, test scores range from 1-5 points per item. Answer categories A and B only require an indication of self-perceived word knowledge, which yields a score of 1 and 2 points respectively, while for any higher score to be awarded, a concrete proof of knowledge is necessary. (Paribakht & Wesche 1997: 179-180). Incorrect responses in either of the categories C-E result in

[30] See Appendix 3 for an example of the *self-report* test.

a score of 2. A score of 3 signifies that a correct synonym or translation has been written in one of the categories C or D. 4 points are assigned if the target word is used within a sentence which clearly displays the learner's understanding of its meaning in that particular context but contains some grammatical error (e.g. the target word is used in another word class, misspelled or incorrectly conjugated as in 'catched' instead of 'caught'). For statements which demonstrate "both semantically and grammatically correct use of the target word, even if other parts of the sentence contain errors" (Paribakht & Wesche 1997: 180) the maximum score of 5 points is assigned.

In order to emphasize the distinction between perceived knowledge and actual knowledge, Sylvén (2004: 59) slightly altered this original scoring system. According to her, in category C, only the scores of 2 points, for incorrect answers, and 3 points, for correct answers, are possible. A score of 4 is not awarded, except from category D onwards (For an illustration of this scheme, see Sylvén 2004: 59, Figure 4:3.)

In theory, this modified scoring system was also supposed to be adopted in the present replication. In practice, however, even though the students had clearly been instructed about how to interpret the five different answer columns, there were quite a number of test responses that could hardly be categorized according to the above-mentioned principles. For these deviating answers, a new mode of correction had to be developed, on condition that the test scores would remain within the original 1-5 point scale. In total, twelve different types of **problem sources** could be detected:

General problem sources:[31]

Problem No. 1: **No answer** is given **at all.**

Description: The student does not choose any of the five answer possibilities A-E but leaves an empty line next to the word under examination.

Solution:

Since the original test design does not allow for any 0-point responses, and we could rule out the possibility that the participant did not know any English at all and had therefore not understood the test instructions, this was interpreted as an instance of

[31] This term refers to deviating answers resulting from careless mistakes or an ignorance of the test instructions.

answer category A: "I do not remember having seen this word before." For responses of this particular kind, a score of **1 point** was awarded. This solution was also welcomed by Sylvén (e-mail: 13 March 2008) and the supervisor of the present thesis.

Problem No. 2: **Demonstration of knowledge** is **missing.**

Description: One of the columns C, D or E has been marked with a cross but no further information is given.

Solution:

This was regarded as an instance of *partial unfamiliarity* (see Table 3:4 above). As in answer category B "I have seen this word before, but I do not know what it means.", a score of **2** was assigned.

Meaning-based problem sources:[32]

Problem No. 3: **Correct word use** is **guessed.**

Description: A sentence showing the correct use of the item in question is written in column E, but the corresponding synonym or translation appears in column C (instead of D).

Solution:

If the **meaning** of the target word was guessed **correct**ly, **5 points** were awarded, as in:

(1)[33] **Item:** *(to merge)* (21)[34]
Answer: (C) 'vermischen'
(E) 'We get the result when we merge these two substances.'
[VBS.SR.19f][35]

If, on the other hand, an **incorrect** answer was given in column C, the student only received **2 points**, as in:

(2) **Item:** *(to commence)* (7)
Answer: (C) 'carry on'
(E) 'Don't stop. Please commence.'
[VBS.SR.13f]

[32] This term covers all of those responses where the meaning of the target item is not understood or expressed in the required manner.
[33] Number of the linguistic example
[34] Number of the item in the respective test under consideration
[35] In the following, this type of code will be used whenever an example is quoted from the students' test answers.

In the above example, the same sentence use might have been suitable for the correct synonym 'to begin'.

Problem No. 4: Target word is used within a **sentence** but **no further information** is given.

Description: The student has written a sentence in column E, but no synonym or translation is given in column D.[36]

Solution:

If the sentence showed that the student was well aware of the word's meaning and its **correct** grammatical use, the maximum score (**5 points**) was awarded. Examples of such answers are:

(3) **Item:** *(a fraud)* (14)
Answer: (E) 'He was sentenced for two years because of fraud.'
[VBS.SR.2f]

(4) **Item:** *(to illuminate)* (17)
Answer: (E) 'The light illuminates the room.'
[VBS.SR.11m]

However, if the target word was used within a sentence which contained some **inaccurate grammar** (i.e. target word used within the wrong word class), as in

(5) **Item:** *(to disguise)* (10)
Answer: (E) 'You are the devil in disguise.'
[VBS.SR.11m]

... or which was **ambiguous** in terms of semantics, a score of **4** was assigned. An example of ambiguous word use is:

(6) **Item:** *(to commence)* (7)
Answer: (E) 'He commenced the race.'
[VBS.SR.2f]

In the above context, both the correct meaning 'to begin', as well as the incorrect one 'to continue, to carry on', which was more frequently used by the students, would be possible.

Sentences which clearly indicated that the student had **misinterpreted** the **meaning** of the item under consideration yielded a score of **2**, as in:

[36] This was one of the most frequent problem sources in the group of the VBS students.

(7) **Item:** *(to illuminate)* (17)
Answer: (E) 'We need to illuminate our dog.'
[VBS.SR.18m]

In example *(7),* the student had probably been thinking about some violent act, as he clearly confused the test item with 'to eliminate'.

The 2-point solution for *Problem No. 4* is fairly similar to

Problem No. 5: **Correct sentence** but **incorrect synonym/translation.**

Description: Column E contains a sentence which shows the (seemingly) correct use of the target word; however, the corresponding translation or synonym in column D is incorrect.

Solution: Just as for all other incorrect answers, **2 points** are assigned, as in:

(8) **Item:** *(in conclusion)* (18)
Answer: (D) 'in Übereinstimmung'
(E) 'In conclusion, you can say that ...'
[TRS.SR.3m]

Problem No. 6: **Sentence** in column E; **both correct and incorrect synonym/translation** written in column D.

Solution:

If the sentence revealed a stronger 'tendency' towards the **correct** option, the student received **5 points**, as in:

(9) **Item:** *(a refuge)* (28)
Answer: (D) 'Flucht/Rückzugsort'
(E) 'He hid in his refuge until they were gone.'
[TRS.SR.10f]

On the contrary, sentences such as

(10) **Item:** *(an editiorial)* (11)
Answer: (D) 'Editorial/Inhaltsverzeichnis'
(E) 'I still had to write the editorial for the book report.'
[VBS.SR.1f]

... where the content was more 'inclined' towards the **wrong** synonym or translation, yielded a score of **2**.

Apart from these *meaning-based problem sources*, a series of answers were also problematic with regard to grammatical features:

Grammatical problem sources:

Problem No. 7: **Semantically correct, but grammatically inaccurate use** of the target word **within** a **sentence.**

Description: Both columns D and E are filled in. Column D provides a correct synonym/translation of the target word. In column E, the word is used within a sentence demonstrating the learner's knowledge of its meaning in that context but with inaccurate grammar (Paribakht & Wesche 1997: 180).

Solution:

In accordance with the original scoring system, **4 points** are awarded (compare Paribakht & Wesche 1997: 180 and Sylvén 2004: 59). Examples of this specific answer type are:

(11) **Item:** *(adjacent)* (1)
Answer: (D) 'anliegend'
(E) 'The line is adjecent to the circle.'
[VBS.SR.4f]

(12) **Item:** *(prediction)* (26)
Answer: (D) 'Voraussagung
(E) 'She predicted the future.'
[VBS.SR.7m]

The same also applied when the word-class confusion or grammatical **error** occurred **in column D**, whereas column **E** showed the semantically and grammatically **correct** use of the target word (= **Problem No. 8**), as in:

(13) **Item:** *(hostile)* (16)
Answer: (D) 'an enemy'
(E) 'The military gained ground on hostile grounds.'
[VBS.SR.12m]

(14) **Item:** *(a bargain)* (4)
Answer: (D) 'to make profit'
(E) 'The car I bought was a real bargain.'
[VBS.SR.13f]

However, if the **word class** was **confused in** both columns **D and E**, an additional point was subtracted, thus resulting in a score of **3** (= **Problem No. 9**). An example of such an answer is:

(15) **Item:** *(to disguise)* (10)
Answer: (D) 'a camouflage'
(E) 'The military uses a disguise to not be seen.'
[VBS.SR.12m]

Grammatical errors in answer category **C** consequently resulted in a score of **2**, even if the meaning of the corresponding item under examination was guessed correctly (= **Problem No. 10**), as in:

(16) **Item:** *(a clue)* (6)
Answer: (C) 'eine Ahnung haben'

(17) **Item:** *(hostile)* (16)
Answer: (C) 'Feind'

[TRS.SR.2m]

The remaining two problem sources relate to aspects such as sentence complexity and collocations. They may be subsumed into the category

Stylistic problem sources:

Problem No. 11: The item is used with semantic and grammatical **correct**ness in both columns **D** and **E** but the corresponding **sentence** is **fragmentary**.

Solution:

Since the instruction for column E reads "... write a full sentence which shows how the word is typically used" (see *Self-report* test, Appendix 3), instead of the maximum score, only **4 points** were assigned for fragments such as:[37]

(18) **Item:** *(a rate)* (27)
Answer: (D) 'Rate'
(E) 'unemployment rate'

[TRS.SR.4f]

(19) **Item:** *(to rely)* (29)
Answer: (D) 'sich verlassen'
(E) 'to rely on sth.'

[TRS.SR.10f]

According to Sylvén's (e-mail: 13 March 2008) suggestion, stylistic errors and **inappropriate collocations** (= **Problem No. 12**), as in

(20) **Item:** *(a poll)* (24)
Answer: (D) 'Umfrage'
(E) 'The government started giving out a poll in which ...'

[TRS.SR.3m]

[37] Errors of this kind more typically occurred in the traditional class.

(21) **Item:** *(a rate)* (27)
Answer: (D) 'Rate'
(E) 'The inflation rate is very big.'
[VBS.SR.3m]

... were treated as those "other parts of the sentence" (Paribakht & Wesche 1997: 180) that were allowed to be erroneous without affecting the test score. For these instances, the maximum of **5 points** was awarded.

Table 3:5 provides an overview of the self-report scoring applied in the present study and briefly summarizes the above-mentioned, original and 'deviating, answer categories:

Table 3:5 *Summary of test scores according to categories in the self-report test*

Answer type	Answers per response column					Score
	A	B	C	D	E	
Category A	X					**1**
Category B		X				**2**
Category C			incorrect			**2**
			correct			**3**
Category D				incorrect		**2**
				correct		**4**
Category E				incorrect	incorrect	**2**
				correct	correct	**5**
Problem No.1	-	-	-	-	-	**1**
Problem No.2			X	-	-	**2**
				X	-	**2**
			-	-	X	**2**
Problem No.3			incorrect	-	correct	**2**
			correct	-	correct	**5**
Problem No.4				-	incorrect	**2**
				-	wrong gr./ambiguity	**4**
				-	correct	**5**
Problem No.5				incorrect	correct	**2**
Problem No.6				inc./corr.	incorrect	**2**
				inc./corr.	correct	**5**
Problem No.7				correct	wrong grammar	**4**
Problem No.8				wrong grammar	correct	**4**
Problem No.9				wrong grammar	wrong grammar	**3**
Problem No.10			wrong grammar			**2**
Problem No.11				correct	fragment	**4**
Problem No.12				correct	wrong collocation	**5**

3.4.2. The words in context test

As the name already implies, this test type was used to measure "the students' ability to understand words in the context of a longer text" (Sylvén 2004: 55), in this case a newspaper article about the success of a particular university programme in the United States (see Appendix 3 for an example of the *words in context* test). Just as the *self-report* test, the *words in context* test consists of 30 lexical items, which are evenly distributed among word classes and frequency groups and, besides, numbered and underlined within the text (Sylvén 2004: 56).

When the test was administered, in addition to the newspaper article, the students received a separate sheet where they had to fill in their answers for each respective lexical item. Similar to the original (Sylvén 2004: 56), the test instructions were as follows:

> Read the article and explain the following words as they are used in the text. You can either translate the word into German, give a synonym in English, or explain the meaning in German or English.

This variety of answer possibilities allowed for considerable freedom of expression, and consequently, when the test was marked, priority was given to semantic features, rather than to grammar errors or spelling mistakes (compare also Sylvén 2004: 60). In accordance with the original study, the scoring was based on a three-part scale, with 2 points for each correct answer, 1 point for an answer which was considered acceptable "but not quite on target yield" (Sylvén 2004: 60) and 0 points being assigned whenever an answer was incorrect, or not given at all.

Examples of the first response category are:

(22) **Item:** *age* (8)
Context: *Released last October [...], the report said, "The beginning of the 21st century appears to be another golden age for MBA students."*
Answer: 'Zeitalter' [VBS.WIC.11m]

or

(23) **Item:** *adhere* (16)
Context: *Harvard adheres to the case study as its sole teaching method, while Wharton is not tied to any single teaching practice.*
Answer: 'sticks to' [VBS.WIC.10m].

A translation which reflects the correct meaning of the word under examination, but does not exactly fit into the immediate context of the article, was for instance:

(24) **Item:** *graduate* (2)
Context: *[T]here are currently about 100,000 students graduating each year with a master of business administration degree.*
Answer: 'maturieren' [TRS.WIC.9f]

Similarly, answers where the student only wrote the German version of the (Latinate) item in question, without having understood its meaning, yielded a score of 1 point, as in:

(25) **Item:** *presence* (20)
Context: *Wharton recently announced that [...] it will soon establish a presence on the U.S. West Coast.*
Answer: 'Präsenz' [VBS.WIC.4f]

This particular item actually proved to be the most difficult one in the entire *words in context* test. Nonetheless, at least four students came up with the following 2-point responses:

(26) **Item:** *presence* (20)
Context: see above
Answers: 'Einrichtung' [TRS.WIC.8f]
'Sitz/Gebäude' [TRS.WIC.10f]
'Anwesen' [VBS.WIC.7m]
'Niederlassung' [VBS.WIC.17m]

Finally, typical examples of incorrect, 0-point answers, are:

(27) **Item:** *performance* (6)
Context: *... MBA programmes were "basking in the success engendered by the strong economic performance of that decade."*
Answer: 'Aufführung, Vorstellung' [VBS.WIC.9f]

as well as:

(28) **Item:** *eventually* (13)
Context: *A school's reputation continues to be the most important criterion, both for those seeking enrollment and those who will eventually hire them.*
Answer: 'möglicherweise' [VBS.WIC.3m]

Here, the test item is clearly mistaken for the German word 'eventuell', meaning 'probably'.

3.4.3. The multiple choice test[38]

In terms of Henriksen's (1999) three dimensions of lexical knowledge, this test exclusively measures the level of *partial-precise* knowledge (compare Sylvén 2004: 55). The students are asked to identify the best of five alternative synonyms for one particular lexical unit within each of the 30 test sentences.[39] In the examples below, the items under examination are underlined.

Following Sylvén (2004: 60), the scoring of this test type is based on a binary system, affording a straightforward process of evaluation: Each correct answer yields a score of 1 point, as for instance:

(29)

Sentence:	The document was <u>abbreviated</u>. (1)				
Possible Answers:	A	B	C	D	E
	burnt	*lengthened*	*recycled*	*shortened*	*translated*
Answer:	D				

If an incorrect answer is given, 0 points are awarded, as in:

(30)

Sentence:	Her <u>spouse</u> was a dentist. (25)
Possible Answers:	A *cousin* B *friend* C *husband* D *profession* E *son*
Answer:	D

On principle, the original *multiple choice* test format allows for only **one** correct answer per sentence (Sylvén 2004: 54-55). However, the following two test items had to be excepted from this basic convention:

(31)

Sentence:	The <u>authorities</u> cancelled the demonstration. (3)		
Possible Answers:	A *government*	B *officers*	C *protesters*
	D *shop-owners*	E *writers*	

(32)

Sentence:	Her <u>untimely</u> death shocked the entire nation. (27)				
Possible Answers:	A	B	C	D	E
	dramatic	*painful*	*too early*	*unexpected*	*violent*

With regard to the former sentence, it can be argued that, strictly speaking, what is under scrutiny here is the students' knowledge of the world, rather than their lexical skills. A close analysis reveals that these two sources are actually quite contradictory. While practical experience and common sense suggest that protest marches are typically dispersed by the police or by other executive forces, i.e. *officers*, a

[38] See Appendix 3, for an example of this particular test type.

[39] A lexical unit may consist of a single word as well as a set of words, as in
(26) Germany has <u>a surplus of</u> teachers. (see also Chapter 2)

considerable number of references (cf., e.g., *MED* or *Thesaurus.com*) favour the term 'authorities' as being used to denominate an organization or institution of the legislative body, such as the "cabinet, council" (*Thesaurus.com*: 13 March 2008) or the *government*. Besides, representatives of the public service might indeed be involved in a demonstration, albeit more often indirectly, that is, as the reason or immediate cause for the protest. In view of this apparent ambiguity, it seems reasonable to accept both options, A and B, as valid 1-point answers.

Concerning sentence *(27)*, a preliminary check with any monolingual (for instance *OALD* 2000: 1426) or bilingual dictionary (*PONS Globalwörterbuch Teil 1 Englisch-Deutsch* 1999: 1350) [40], clearly defines alternative C *too early* as the correct synonym for the word to be tested. Yet, on closer examination, it turns out that, likewise, option D *unexpected* takes into account the premature and inopportune nature of the person's death. The *MED* actually illustrates that the word 'untimely' is typically used to refer to

> the death of someone who dies at a time that makes their death extremely sad or **unexpected**, especially because they are young (*MED* 2002: 1577).

Thus it appears that both answers, C and D, should be considered acceptable. This view has also found support in the dialogue with other students, teachers, lecturers and even two native speakers of English. Most of them argue that while *too early* is the semantically more accurate synonym, *unexpected* seems more appropriate in the immediate syntactic context of the sentence.

Evidently, it would be necessary to perform a trial test run with a larger set of native speakers, in order to be able to fully resolve this conflict. However, for pragmatic reasons, such as limited time and resources, this is beyond the scope of the present study.

[40]Henceforth referred to as *PONS Global.*

3.4.4. The cloze tests[41]

In these two tests, the students had to demonstrate their knowledge of fixed lexical phrases and idiomatic expressions. In each of the sentences under consideration, only **one** particular word was missing (compare also Sylvén 2004: 61). If this specific word was guessed correctly, a score of 2 points was awarded, as in:

(33) We have worked enough, so let's call it a day and go home. [VBS.CT2.2f]

or in:

(34) "Is the dress to big?" "No, not at all. It fits like a glove." [VBS.CT2.21f]

If, instead, another word was given, "which was syntactically and semantically feasible" (Sylvén 2004: 61), but still not the exact item that was required, the answer yielded a score of 1. As Sylvén (2004: 61) notes, this strategy was employed in order to promote the students' inventiveness. For instance, linguistically more 'creative' versions for the above examples were:

(35) We have worked enough, so let's call it a siesta and go home. [TRS.CT2.2m]

and

(36) "Is the dress to big?" "No, not at all. It fits like a second skin." [VBS.CT2.17m]

As mentioned at the beginning of section 3.4., when the two *cloze tests* were marked, a series of different dictionaries and similar sources were consulted. Just as in Sylvén's study, especially the *BNC online sampler* proved to be extremely helpful when the correctness of the students' answers was in question, as in:

(37) I can't watch horror movies. They give me the horror. [TRS:CT2.5m]
chills. [VBS.CT2.8f]
thrill. [VBS.CT2.11m]

Indeed, this online reference provided evidence for the usage of all the three above-mentioned items in the expression 'to give so. ...'.

[41] For examples of both, *cloze test 1* and *2*, see Appendix 3.

On the contrary, no occurrences of the target word 'opinion' could be found with regard to the phrase in example *(38)*:

(38) *I have to write an essay about history and since you know a lot about it but I don't, do you mind if I pick your opinion?*

[VBS.CT1.1f]

As a consequence, this answer yielded a score of 0, which was also awarded for clear-cut incorrect answers, such as:

(39) *Speak up a bit! I'm a bit death of hearing, you see.* [TRS.CT1.5m]

-- and whenever a student did not offer any answer at all.

As the above examples have shown, the *cloze* test design definitely requires a profound knowledge of English phrases and idioms, which, in turn, according to the *Oxford dictionary of English idioms* (*ODEI* 1993: x) equals an almost native-like command of the language. By including two tests of this particular kind, we therefore hope to discern the most proficient language learners in the present student population.

Before we turn to the results on the individual lexical tests, the methods used for the statistical analyses and, more importantly, the overall test outcomes will be accounted for.

Chapters 4 and 5 illustrate the outcomes of the five vocabulary tests conducted at a Viennese bilingual school in the period February – March 2007. First, the results of the total test battery will be analysed with regard to the aspects of group, gender, research context and mother tongue (see Chapter 4).
As a second step, these analyses will be extended to the separate test types. Wherever possible, the students' answers are examined according to their distribution in the categories *correct, acceptable, incorrect* and *no answer*. (Chapter 5).
Chapter 6 presents the results on the learner questionnaires and shows how the students' lexical performance relates to extracurricular factors such as reading, TV and Internet habits, time spent in English-speaking countries and parents' education.

4. Overall results and statistical grounding

This chapter discusses the total mean results of the entire test sequence. As a starting point, initial insights will be provided into the scoring of the two groups. On the basis of these preliminary findings, the statistical methods used for the quantitative analyses in the present thesis will be accounted for. Subsequently, the overall test results will be analysed with regard to the above-mentioned aspects.

4.1. Basic insights

When the scores for all five vocabulary tests are added up, the maximum score is 360 points (150 for the *self-report* test, 30 for the *multiple choice* test and 60 points each for the *words in context* test as well as *cloze test 1* and *2* respectively). Figure 4:1: shows the total mean scores for the two test groups:

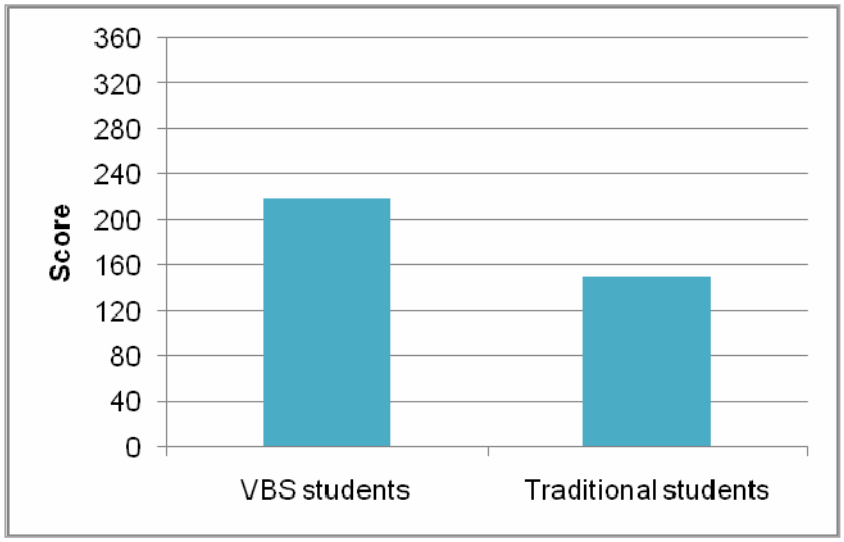


Figure 4:1 *Total mean score*

The VBS students are well ahead of their traditional peers, with a mean score of 218.14 as compared to 150.75 points. This superiority becomes more evident if the total test score of each individual participant of the CLIL group is contrasted with that of their conventionally taught contemporary, as is the case in the following figure.

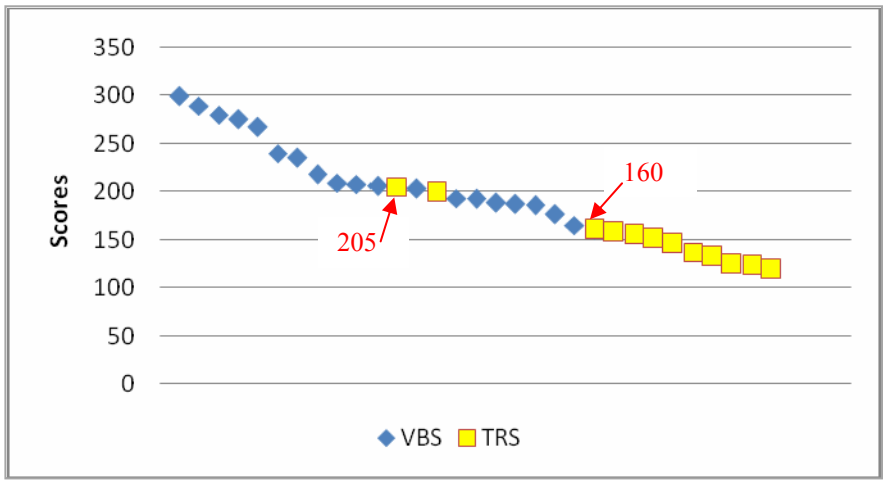


Figure 4:2 *Total test scores, ranking of individual values*

As Figure 4:2 shows, the test results of the bilingual group extend from a maximum value of 299 points to 160 points, while traditional students' total scores range from 205 to 119 points, which is the lowest score in the whole sample. Furthermore, the graph reveals that only three out of the twelve pupils attending the regular 11th form

have achieved test grades that lie within the scope of their VBS peers. The two scales actually intersect twice, at the scores of 205 and 160, which rank first and third on the yellow scale and occupy position 12 and 21 on the blue scale respectively (see Figure 4:2).

In view of the expressiveness of the above illustration, one is tempted to conclude that it provides sufficient evidence for the assumption that the VBS group clearly outperforms the control group with regard to lexical knowledge and, thus, that *CLIL* is the method to be advocated in vocabulary learning. However, in order to ensure that the numerical differences in total and mean values detected between the two test sets, also reflect real-world differences, a sound statistical analysis proves to be indispensable.

4.2. Statistical foundations

For the purpose of the present study, two specific statistical tests, the *F-test* and the *t-test*, have been of great significance. The underlying concepts of these methods will be explained in detail in the present section. To perform the statistical analyses of the test results, the *Microsoft Office Excel 2007* programme has been used.

Considering Figures 4:1 and 4:2, it can easily be discerned that there is considerable variation, not only between the mean scores of the two groups, but also between individual scores **within** each of the groups, in particular within the VBS group. These internal deviations indicate that any claims about the average test grade of the bilingual class as opposed to the traditional class have to be made with the utmost care. As Rietveld & van Hout (1993: 14) point out, it is indeed possible

> that the variation in mean scores is brought about by the variation between the individual subjects. In that case, the variation between the mean scores only reflects individual differences, not differences between the [teaching] methods used.[42]

In order to rule out this worst-case scenario, at first, an analysis of standard deviation or variance has to be performed, to determine whether the internal variation within each of the two groups is decisive for the overall test result or simply "caused by irrelevant individual differences between the subjects" (Rietveld & van Hout 1993:

[42] In the present study, the term 'teaching methods' refers to *Content and Language Integrated Learning* as contrasted to traditional instruction.

16).[43] The *standard deviation s* is a measure of the degree to which all values within a specific data sampling diverge from the mean (Timischl 1996: 36).[44] It is calculated by means of the following formula

$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad ,$$

where $x_i - \bar{x}$ represents the distance or deviation between each observed score and the mean, and n the total number of values within the sample. In the present example, the value of the standard deviation amounts to 41.66 points in the VBS group and 27.77 points in the control group. Apparently, this means that the test scores of the bilingual students are more dispersed than those of their traditional peers, as has been illustrated in Figure 4:2. However, this statement does not yet provide a sound scientific proof. This is precisely the point where the statistical model of the *F-test* comes into effect (see also Timischl 1996; Deutsche Gesellschaft für Qualität[45] 1993):

4.2.1. The F-test

Instead of comparing the magnitudes of the two standard deviations we want to determine with a specified degree of certainty whether the difference in standard deviation between the two groups is significant or simply based on chance (Rietveld & van Hout 1993: 15). In other words, a hypothesis, the so-called *null hypothesis* H_0, which suggests that there is no significant variation between the two values under observation, has to be tested and, if necessary, also be rejected. In that case, another hypothesis, the *alternative hypothesis* H_A that there is such a variation is accepted instead (see also Rietveld & van Hout 1993: 15; Timischl 1996: 148; Sylvén 2004: 72).

Focussing on variance and standard deviation, the hypotheses for the present problem can be expressed in the following way:

[43] In Sylvén (2004: 71), the similarity of the two standard deviations is assumed as a precondition for the t-tests to follow. However, as this approach often leads to erroneous results (compare Rietveld & van Hout 1993: 15), in the present thesis, an analysis of variance is considered indispensable.
[44] Its square is called the *variance* s^2
[45] Henceforth referred to as DGQ

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ gegen } H_A : \sigma_1^2 \neq \sigma_2^2 .$$ (Timischl 1996: 156)

The terms *sigma 1* (σ_1) and *sigma 2* (σ_2) stand for the variances of two normally distributed data samples.[46] As asserted by Sylvén (2004: 71), a normal distribution of the variable is assumed as precondition for the statistical tests under scrutiny. Indeed, visual representations, obtained via *Derive* or a similar mathematical programme, demonstrate that the numerical results of the various language tests correspond closely to the bell-shaped graph of the normal distribution function. [47]

Based on the above hypotheses, as a next step, the test proportion F_{pr}, relating the variances s_1^2 and s_2^2 of the present sample, is computed: $F_{pr} = \frac{s_1^2}{s_2^2} = \frac{41.66^2}{27.77^2} \approx 2.25$.[48]

In addition, a second value, the so-called *critical value* or *F-value*, is determined.

For this purpose, the statistical model of the *F-distribution* has to be applied. The structure of this probability distribution is a function of two parameters f_1 and f_2, referred to as the *degrees of freedom* (Timischl 1996: 157).[49] These are obtained by reducing the total number of entries within each data set by a value of one. In this particular instance, the number of records is equivalent to the amount of students per class, i.e. $f_1 = n_{VBS} - 1 = 20,\ f_2 = n_{TRS} - 1 = 11$.

Apart from the grades of freedom, the definition of the critical quantity $F_{f_1, f_2; 1-\alpha/2}$ also depends on a specific level of statistical significance α, which has previously been alluded to as 'degree of uncertainty'. As Rietveld & van Hout explain,

> [t]he alpha level is the probability of rejecting the null hypothesis (=H_0) when this hypothesis is in fact true. The corresponding error is called the Type I error. (Rietveld & van Hout 1993: 4)[50]

[46] On the contrary, the concrete values for (or realisations of) the variances of a given sample are referred to as s_1^2 and s_2^2

[47] According to common statistical conventions, strictly speaking, the normality of the variable would have to be verified by means of a separate test. For the present purposes, however, the basic assumption of a normally distributed data set is sufficient.

[48] In the quotient s_1^2/s_2^2, the higher variance always goes into the numerator (Timischl 1996: 157).

[49] For a more detailed description on the *F distribution,* see Rietveld & van Hout (1993: 22-25).

[50] The *α-level* is contrasted to the so-called *β–level*, which defines "the probability of accepting the null hypothesis when this hypothesis is in fact false. The corresponding error is called the Type II error". (Rietveld & van Hout 1993: 4)

In order to preclude such an erroneous outcome, the significance level has to be fixed before the actual test performance (Timischl 1996: 148). As the DGQ (1993: A4, 11) points out, it is the researcher's task to choose an α-level that is appropriate for the respective test environment. For the t-tests, Sylvén (2004: 72) for instance proposes an alpha level of 0.05. However, to be on the safe side and avoid biased results, it is highly recommended to conduct the experiment at different levels of significance:

Figure 4:3 *Common levels of statistical significance*

$\alpha = 0{,}05$	bzw. $1-\alpha = 0{,}95$	$(1-\alpha/2 = 0{,}975)$
$\alpha = 0{,}01$	bzw. $1-\alpha = 0{,}99$	$(1-\alpha/2 = 0{,}995)$
$\alpha = 0{,}001$	bzw. $1-\alpha = 0{,}999$	$(1-\alpha/2 = 0{,}9995)$

(adapted from DGQ 1993: A4, 11)[51]

Given all the information required, the critical values can either be derived from the existing, universally valid, F-scales (compare for instance Timischl 1996: 324-326), or be obtained computationally via *Excel*. In both cases, the present problem yields the following results:

Table 4:1 *Paired F-test: analysis of variance, total results*

VBS vs. Traditional $n_1 = 21$, $n_2 = 12$	test value F_{pr}	α	F-value $F_{f_1,f_2;1-\alpha/2}$
	2,250147542	0,05	3,22614478
		0,01	4,85522025
		0,001	8,14150397

The actual test procedure now consists in comparing each of the F-values with the above calculated quotient of the two variances and drawing adequate conclusions:

> Liegt die Größe F innerhalb des zweiseitigen Zufallsstreubereichs, so handelt es sich um ein typisches oder normales Ergebnis für den Fall [$\sigma_1^2 = \sigma_2^2$]. […] Die Annahme [$\sigma_1^2 = \sigma_2^2$] kann also akzeptiert werden.
>
> Liegt die Größe F dagegen außerhalb des zweiseitigen Zufallsstreubereichs, so ist dies nicht typisch oder normal für den Fall [$\sigma_1^2 = \sigma_2^2$]. […] In diesem

[51] The expression 1-α is called *level of confidence*. It is used whenever the statistical test under consideration is *unpaired* or 'one-sided'. *Paired tests* draw on the quantity 1- α/2 instead (DGQ 1993; Rietveld & van Hout 1993; Timischl 1996). Since the present problem concerns the (possible) difference between the two variables σ_1 and σ_2, a paired test has to be applied (Timischl 1996: 156).

> Fall wird also die Annahme [$\sigma_1^2 = \sigma_2^2$] nicht akzeptiert. Es liegt ein signifikanter Unterschied vor. (DGQ 1993: A 3.6, 4)

In other words, the null hypothesis is rejected if the test value exceeds the critical value, i.e. $F_{pr} > F_{f_1,f_2;1-\alpha/2}$, and accepted if both values are equal or if F_{pr} is smaller than $F_{f_1,f_2;1-\alpha/2}$, i.e. $F_{pr} \leq F_{f_1,f_2;1-\alpha/2}$ (Timischl 1996: 157).

As illustrated in Table 4:1, the quantity $F_{pr} \approx 2.25$ clearly remains below each of the F-values for the significance levels 0.05, 0.01 and 0.001 respectively. Consequently, there is no reason for discarding H_0. The variation between the standard deviations of the two data samples can thus be said to be purely accidental.

Having established that the dispersion of test scores **within** the two student groups does not influence the overall test result, the differences in mean scores have to be evaluated from a statistical point of view. This is generally accomplished by means of a *t-test*:

4.2.2. The t-test

Similar to the F-test, illustrated in the preceding section, this statistical instrument can be applied in a paired as well as an unpaired manner. As Sylvén (2004: 71) explains,

> [t]he unpaired t-test compares the mean values of two groups for a single variable. The assumptions for this test are a normal distribution of the variable and a fairly similar standard deviation in the two groups. [52]

Proceeding on these basic conditions, the *unpaired t-test* proves to be the most appropriate method for assessing whether there exists significant variation between the average test scores of VBS as contrasted with traditional students, or female compared to male students (Sylvén 2004: 71).

The *paired t-test*, on the other hand, is employed for analyses of "two sets of scores, when the scores are matched or paired" (Wright 1997: 49), as for instance the results of Sylvén's (2004: 72) three separate test rounds. Since the present thesis focuses on one test sequence only, this procedure appears to be irrelevant.

[52] In the present study, whenever necessary, the latter is proven by means of a paired F-test.

Returning to the unpaired mode, as a first step, the working hypotheses have to be formulated. In contrast to the F-test, conducted in order to discover **any** kind of difference in the variances or standard deviations of the two samples, the unpaired t-test aims at verifying a **concrete** assumption, namely that the mean value of VBS scores, μ_1, considerably **exceeds** the corresponding value μ_2 of the traditional group. Thus, for the present problem, the hypotheses are:

$$H_0 : \mu_1 \leq \mu_2 \text{ versus } H_A : \mu_1 > \mu_2 \text{. (compare DGQ 1993: A 3.6, 19)}$$

In the same way as in section 4.2.1, the actual test performance consists in comparing a particular quantity t_{pr} with the so-called *t-value*, that is, the critical value $t_{n_1+n_2-2;\ 1-\alpha}$, depending on the definition of α. Wishing to minimize the risk of statistical errors, once again, the experiment is conducted with regard to each of the three common significance levels 0.05, 0.01 and 0.001 respectively.

For the test value, the difference of the two group means $\bar{x}_1 - \bar{x}_2$ has to be divided by the standard deviation s_d of the mean differences, which is calculated according to the following complex formula:

$$s_d = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot \frac{s_1^2 \cdot (n_1 - 1) + s_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}} \text{ (DGQ 1993: A 3.6, 19).}$$

As mentioned in the previous section, the variables n_1 and n_2 represent the total number of students in each of the two groups. The symbols s_1^2 and s_2^2 are used to denote the related variances. Entering these pieces of information, the test quantity $t_{pr} = \frac{\bar{x}_1 - \bar{x}_2}{s_d}$ yields a result of approximately 4.99.

In the following, the *t-values* are computed or obtained from the corresponding distribution scale (see for instance Timischl 1996: 321). All in all, the unpaired t-test produces the following outcomes:

Table 4:2 *Unpaired t-test, total mean score*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS vs. Traditional $n_1 = 21$, $n_2 = 12$	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,695518742
				0,01	2,45282418
	67,39285714	13,50801997	4,989099608	0,001	3,37489928

As described in connection with the F-test, the interpretation of the above results is determined by a relation of magnitude between the test value and the t-value: While the inequation $t_{pr} \leq t_{n_1+n_2-2;\ 1-\alpha}$ signifies the validity of the null hypothesis, the mathematical expression $t_{pr} > t_{n_1+n_2-2;\ 1-\alpha}$ provides evidence for rejecting H_0 and accepting the alternative hypothesis instead (DGQ 1993: A 3.6, 19), as is the case for each of the three significance levels in the present issue.

What is more, Table 4:2 demonstrates that the difference in mean scores between the two groups is actually *highly significant*, meaning that the value of t_{pr} 'contradicts' the null hypothesis, already at the lowest α-level of 0.001.[53] In other words, there is a 99.9% chance of the VBS mean exceeding the average score of traditional students. Thus, the assumption made at the beginning of section 4.1. has been substantiated.

To sum up, the preceding sections have offered an outline of the statistical methods utilized for the quantitative analysis of test results in the present study. Similar to the mean test scores, other research aspects, such as gender differences or variations with regard to the individual vocabulary tests, will be evaluated according to the previously established techniques.

[53] For a detailed account of the evaluation of test results at different significance levels, consult DQG (1993: A 4, 11-12).

4.3. Further (statistical) considerations

Having ascertained that, on the most basic level, the difference in mean scores between the VBS class and the traditional class is statistically highly significant, the question arises as to whether this relation still holds if the figures of the original sampling are subjected to certain changes. The following passages investigate how factors such as student's mother tongue and gender may influence the overall test result. An additional emphasis is placed on the comparison of the present findings with the outcomes of Sylvén's (2004) longitudinal study.

4.3.1. Mother tongue

As illustrated in Chapter 3, one of the distinguishing features of a *Vienna Bilingual School* is that it offers secondary education for a considerable number of 'international students' with native languages other than German. Apart from their various mother tongues, including English, members of this particular school community have learned either German or English as their second or third language respectively, or have acquired both languages simultaneously, in connection with their bilingual education.

Given the relevance of such multilingual backgrounds, it is one main concern of the present study to discover whether the degree of variation among the total means of the two test groups will significantly decline, if international students' scores are excluded from the VBS population. In order to obtain detailed insights, this specific situation is actually analysed in two consecutive stages:
At first, English speakers' values are subtracted from the overall scoring. This sanction affects four out of the 21 participants in the CLIL group. As a next step, the test grades of three additional students, whose L1 is neither German nor English, are exempted from the evaluation. The resultant second subset, thus, only comprises 'typical Austrian students', in other words, participants with no other mother tongue than German.

Figure 4:4 compares the total mean scores of the original VBS assemblage and its two subgroups with the average value of the traditional group:

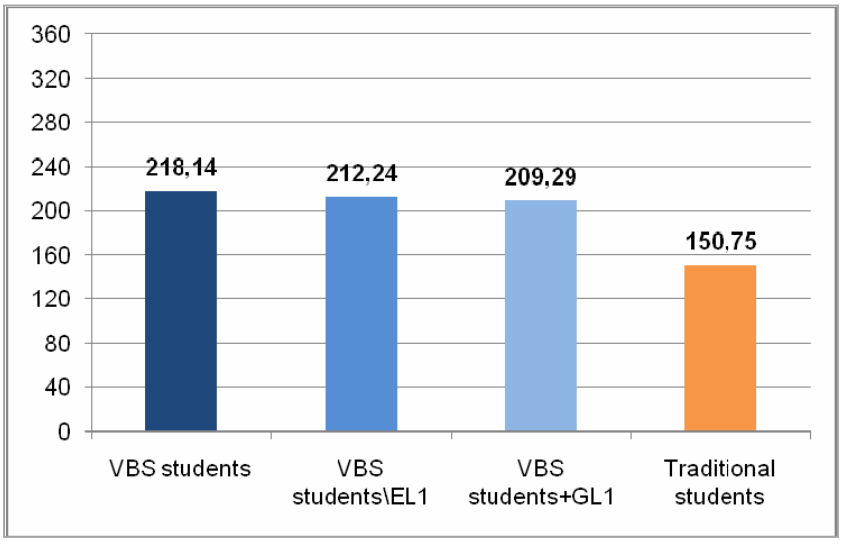


Figure 4:4 *Total mean score, group/factor: mother tongue*

As illustrated in the above bar chart, irrespective of any modifications with regard to native language, the bilingual students keep the lead with a mean score of 212.24 for the non-English subset and 209.29 for the subgroup of German speakers, compared to an average of 150.75 points achieved by their traditional peers. Furthermore, the validity of these numerical results is confirmed by the corresponding statistical analyses:

The F-tests, conducted in advance, substantiate that the difference in variance between the traditional group and the first as well as the second VBS subset is purely accidental, as is the variation in the dispersion measure between the original CLIL group and each one of its subdivisions. Consequently, a series of unpaired t-tests can be employed to evaluate the relation of the mean values from a statistical point of view.

4.3.1.1. English as L1[54]

As mentioned above, at first, a comparison is made between the average result of the conventional and the bilingual group, excluding all native speakers of English. The unpaired t-test results in a test value of approximately 4.54 (see Table 4:3), which exceeds the critical value, even at an alpha-level as low as 0.001. Hence, the difference between the two test groups is statistically highly significant.

Table 4:3 *Unpaired t-test, total mean score, group/factor: non-English mother tongue*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS\EL1 vs. Traditional n_1 = 17, n_2 = 12	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,703288423
				0,01	2,472659904
	61,48529412	13,55106775	4,53730254	0,001	3,421033621

The effect of English speakers' scores on the overall test result can also be assessed by contrasting the original VBS population with its non-English subset. In this particular case, the statistical evaluation reveals that the variation in means between the two groups is based on chance:

Table 4:4 *Unpaired t-test, total mean score, VBS/factor: non-English mother tongue*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS vs. VBS\EL1 n_1 = 21, n_2 = 17	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,688297694
				0,01	2,43449404
	5,907563025	13,4412589	0,439509652	0,001	3,332624256

Considering the relation between the test value and each respective t-value in Table 4:4, the presence or absence of students with English as their L1 appears to be irrelevant to the average VBS score and, eventually, to the test result as a whole. However, before drawing any hasty conclusions, the above findings must be specified more carefully.

[54] Admittedly, with a total of only four students, the subgroup of English natives is too small for any generally valid statistical analyses. Nevertheless, since these particular students play an essential role in the VBS programme, a special treatment of their test outcomes was considered indispensable. Evidently, all the related values and figures are simply restricted to the present research context and have to be taken with the utmost caution.

For this purpose, the mean value of the four English participants is related directly to their non-English classmates'. Not surprisingly, in absolute numbers, the native English subjects perform slightly better than their non-native peers with an average value of 243.25 points opposed to 212.24 points, as depicted in Figure 4:5.

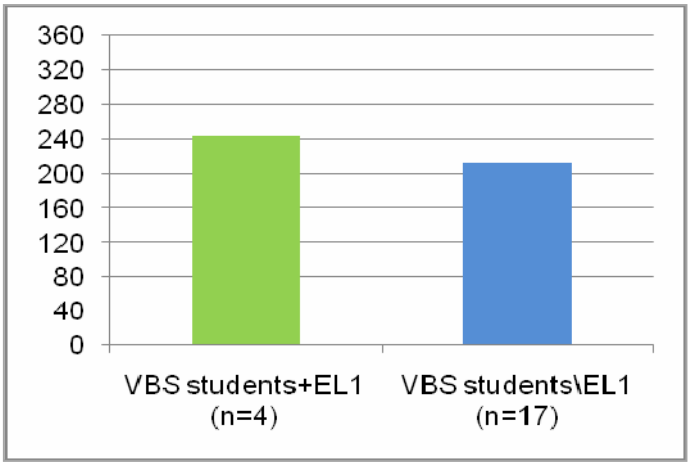


Figure 4:5 *Total mean score, VBS/factor: (non-) English mother tongue*

Nevertheless, this superiority cannot be confirmed by the outcomes of the corresponding t-test. As Table 4:5 illustrates, the test value remains below the t-value for all three significance levels 0.05, 0.01 and 0.001, meaning that the differences in mean scores between the group of English speakers and the non-English VBS subset are not statistically significant, but simply caused by variations among individual scores.

Table 4:5 *Unpaired t-test, total mean score, VBS/factor: (non-) English mother tongue*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS+EL1 vs. VBS\EL1 n_1 = 4, n_2 = 17	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,72913279
				0,01	2,53948319
	31,01470588	22,66173357	1,368593704	0,001	3,57940015

In conclusion, the preceding analyses have shown that the factor 'English mother tongue' has no decisive influence on the total result of the test sequence. Notwithstanding the exclusion of English native speakers, on average, the CLIL group scores considerably higher than its traditional counterpart (see Table 4:3).

Moreover, the participation of Anglophone students does not contribute substantially to increasing the total VBS mean (Table 4:4), since no statistical evidence has been found to confirm that these particular test subjects significantly outperform their non-English classmates (Table 4:5).[55]

4.3.1.2. German as L1

As a next step, the issue of mother tongue will be extended to the whole population of international students. Thus, in addition to the four English speakers, the overall scoring scheme will be analysed with regard to the remaining three students with non-German language backgrounds. Figure 4:4 (see above) shows that the exclusion of all international students still results in a total mean value of 209.29 for the VBS group compared to 150.75 for the traditional group. The following table illustrates that this difference is also statistically highly significant at the alpha-level of 0.001:

Table 4:6 *Unpaired t-test, total mean score, group/factor: German mother tongue*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS+GL1 vs. Traditional $n_1 = 14$, $n_2 = 12$	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,710882067
				0,01	2,492159469
	58,53571429	12,90568644	4,535652912	0,001	3,466777294

Furthermore, a reverse check, contrasting the subset of German speakers with the VBS group as a whole, corroborates that the omission of international students' scores does not substantially diminish the average result of the bilingual group. Nonetheless, in the same way as in the previous section, the situation of non-German CLIL students must be investigated in more detail:

A comparison of absolute numbers indicates that, with a mean score of 235.86, those participants with native languages other than German are slightly ahead of their 'typical Austrian' VBS colleagues, having achieved 209.29 points on average (see Figure 4:6).

[55] Apparently, in this connection, it is slightly problematical to speak of any 'statistical evidence', given the small number of representatives in the English-speaking subgroup. The above-mentioned conclusion is therefore only valid for the present test sample and does not reflect any general rule or tendency.

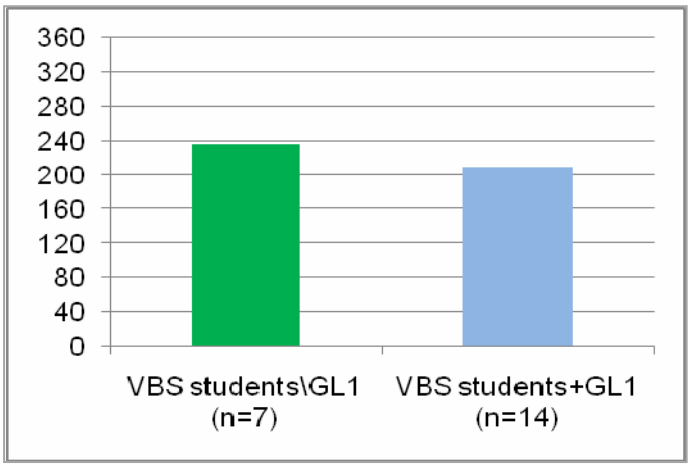


Figure 4:6 *Total mean score, VBS/factor: (non-) German mother tongue*

However, the related statistical analysis demonstrates that this variation is solely based on dissimilarities between individual subjects within the two groups and does not reflect significant 'real' differences (see Table 4:7). In particular, the supremacy in the mean value of the non-German VBS group can primarily be attributed to the fact that, among its seven participants, this subset also includes the absolute top performer of the whole test population.

Table 4:7 *Unpaired t-test, total mean score, VBS/factor: (non-) German mother tongue*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS\GL1 vs. VBS+GL1 $n_1 = 7$, $n_2 = 14$	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,729132792
				0,01	2,539483189
	26,57142857	18,82319227	1,411632426	0,001	3,579400148

Summing up, none of the above statistical evaluations has succeeded in validating the initial assumption that the VBS group's superior test performance relates directly to some students' multilingual backgrounds. Quite the contrary, it has been substantiated that neither the scores of English native speakers nor those of the entire international subset considerably raise the total VBS mean. Even though in detail, the average result of the English subgroup slightly exceeds that of its non-English counterpart, statistically, this disparity is simply ascribed to coincidental variations between individual subjects. The same is true for the whole population of international students, whose marginal predominance over their German-speaking classmates is mainly due to certain, exceptionally high, single values.

By and large, it turns out that, irrespective of their linguistic provenance, the VBS students outperform their traditional peers with regard to the total test results. However, before generalizing and attributing their impressive achievement exclusively to the CLIL method, the outcomes on the separate vocabulary tests, as well as the role of extracurricular English language input deserve closer attention. These aspects will be examined at later stages in my thesis. For the time being, a comparison will be made between the current research findings and the original study:

4.3.2. The original study

Since the present study is a replication of Sylvén (2004), it is of great interest to discover how the results under consideration relate to those obtained in the original context. In order to provide a sound basis for comparison, the above-mentioned total means have to be subjected to a few minor changes:

As mentioned earlier (see section 3.3.1.), for the present purposes, the majority of linguistic materials were adopted from Sylvén's test round III (compare Sylvén 2004: Appendix 3, 270-278). One additional vocabulary test, namely *cloze test 2*, was adopted from her second examination series (see Sylvén 2004: Appendix 2, 262-263). In the following, the results of this specific test type are therefore excluded from the overall scoring. Consequently, the maximum score for the present comparison will be 300 points (150 for the *self-report* test, 30 for the *multiple choice* test and 60 points each for the *words in context* test as well as for *cloze test 1*, see also Sylvén 2004: 76), and the average test score will be correspondingly smaller.

Moreover, the factor of students' mother tongue has to be taken into account. In a personal conversation, Sylvén mentioned that her study had only involved native speakers of Swedish (Sylvén Sept. 2007). Yet, certain passages in her thesis convey the impression that, similar to the Austrian research situation outlined above, the original test population comprised native Swedes as well as non-natives (cf., e.g., Sylvén 2004: 203-208), just as participants who were "bilingual from home" (Sylvén 2004: 41), possibly also multilingual, with English as one of their mother tongues.

Since there seem to be no clear specifications on this particular issue, each of the above possibilities will be considered in our comparison. Thus, in order to obtain a

comparative Austrian CLIL value, an average score is calculated from the total means of three specific subsets, namely the original VBS community, the non-English subgroup and the set of exclusively German speakers, as defined in the previous section. As for the traditional group, besides the exclusion of the *cloze test 2* scores, no further modifications are required, since it only includes native Austrians.

Table 4:8 shows the total mean values for the three VBS subgroups in the reduced test battery, as well as the resultant comparative value:

Table 4:8 *Total mean score, reduced test battery, VBS/factor: mother tongue, deduction of comparative value*

VBS^{r} subgroups	Total mean score	Comparative CLIL value
VBS^{r} total	188.29	185.34
VBS^{r}\EL1	184.88	
VBS^{r}+GL1	182.86	

Based on the above adaptations, Figure 4:7 finally reveals how the mean score results of the four test types in Sylvén's test round III relate to the corresponding findings in the present study:

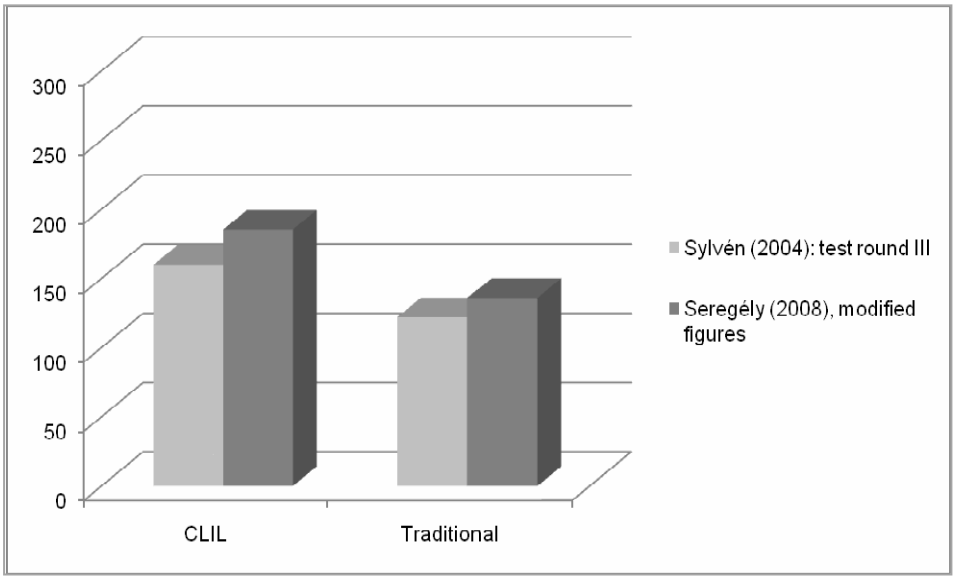


Figure 4:7 *Total mean score, reduced test battery, group/factor: research context*

As illustrated in the bar chart, the Austrian students are slightly ahead of their Swedish peers in both the CLIL group and the traditional group. While the

'bilingual' Swedes achieved an average of 160 points (Sylvén 2004: 95), the comparative value for their Austrian counterparts amounts to 185.34 points (see Table 4:8). With regard to the traditional groups, the mean scores are 122 for Sylvén's (2004: 95) subjects, compared to 136 for the students at the Viennese school.

Due to the fact that individuals' scores have been unavailable in the context of the Swedish study these differences cannot be statistically evaluated. However, taking into account that Sylvén's test population was considerably larger than the present one, with a total of 99 CLIL and 264 control subjects (compare Sylvén 2004: Abstract) contrasted to 21 and 12 respectively, it appears rather unlikely that any significant outcomes would be detected. What is more, the relatively low values for both mean differences (approximately 15 points for the CLIL groups and 14 points for the traditional groups) contradict rather than support the assumption of statistical relevance.

As for the Austrians' supremacy in average scores, one may only speculate about possible reasons: If we disregard, for the moment, the amount of *CLIL* and extracurricular English language input, the simplest explanation for the numerical differences lies in the scoring mode itself. As mentioned at the beginning of section 3.4, in the Swedish and the present study, two distinctive solution lists were consulted for the correction of students' responses. Even if these answer catalogues were fairly equal, which is quite probable, given that they both are based on common frameworks of reference, that is, the respective monolingual and bilingual dictionaries, as well as the *BNC* online, there is still the risk of individual bias. Indeed, it seems that, no matter how fixed the established assessment criteria, in the end, test correction remains a highly subjective issue, as one of the researchers might be more lenient than the other.

With regard to the present situation, this means that possibly some of the responses, classified as *partly correct* or *acceptable* according to the Austrian evaluation system, may have been rejected as *incorrect* within the Swedish context, thus resulting in a lower test score. Since the original solutions are not available for comparison, this hypothesis will never be entirely confirmed. Nonetheless, an analysis of the results on the separate test types will reveal whether the distribution

of answers among the four different categories *correct, partly correct, incorrect* and *no answer* considerably varies between Sylvén's study and the present replication. This comparison will be drawn in the course Chapter 5.

In addition to subjective variations according to the respective scoring mode, yet, another uncertainty factor must not be neglected, namely 'direct EFL input', in other words, linguistic contents being taught almost immediately **before** or even **in between** individual test procedures, in the context of students' English language lessons. In an ideal experimental situation, such influences are usually kept under strict control, more precisely speaking, ruled out in advance. However, this is not the case in the present study, as the experimental situation under scrutiny is 'real' or 'permanent', meaning that, within their educational environment, the participants constantly receive English language input through either the *traditional* method, that is, English as a separate subject at school, or both the *traditional* **and** the *CLIL* method (compare Sylvén 2004: 40-41).

As mentioned above, for the current discussion, we have excluded the CLIL aspect, as it appears extraordinarily difficult to identify each 'national' group's exact position on the CLIL continuum: For the Swedish test population, a selection of four schools was involved, each of them providing a distinctive level and amount of bilingual instruction. "[S]ome schools use CLIL throughout the school day while others use it only temporarily" (Sylvén 2004: 41, for more details see also the following pages in her thesis).

Similarly, this study's CLIL/VBS class is not utterly homogeneous: Some students were exposed to two languages from birth or have attended *Content and Language Integrated* programmes from early stages in their formal education onward, whereas others have only been exposed to English as a medium of instruction since their entry into the bilingual upper secondary at this particular school. In view of these inconsistencies, it appears rather problematic to attribute the superiority of the Austrian CLIL students exclusively to the fact that, at the time of the test

performance, they had 'officially' been taught with the CLIL method for one year longer than their Swedish peers.[56]

Besides, according to our previous speculation, the factor of direct EFL input seems to play a decisive role in the scoring of both the bilingual as well as the control subjects. So, it is not at all unlikely that, prior to the examination, by coincidence, certain lexical items included in the test battery were explicitly taught in one of the Austrian English language classes, while in the corresponding Swedish EFL lessons they were not, thus yielding a comparatively lower overall score. For pragmatic reasons, it is not feasible to verify this assumption. Nevertheless, it may at least partly account for the differences in achievement between the present study's participants and the original test subjects, especially pertaining to the learners taught in the traditional manner.

In connection with the respective two subgroups, our initial hypothesis was that the Swedes would outperform their Viennese peers owing to the fact that Swedish everyday life, and particularly the media, provides more opportunities for acquiring English vocabulary than Austrian mainstream culture.[57] However, as the above figure has shown, the converse proves to be true (see Figure 4:7). The exact details on the results according to students' leisure activities will be illustrated in Chapter 6. For the moment, it is sufficient to state that, although the Austrian control group lags behind its Swedish counterpart in the amount of extracurricular English language input, still, it is superior with regard to the total mean score of the four test types under consideration. This, in turn, seems to be a fairly persuasive argument for the impact of explicit EFL instruction.

Summing up, a comparison of total mean scores has revealed that the Austrian learners are slightly ahead of their Swedish peers, for both the CLIL group and the traditional group. Yet, these findings should be viewed cautiously, since apart from the shared test materials, research conditions in the two contexts were only similar to

[56] As a basic condition for the present replication it was assumed that, notwithstanding any previous experience, both CLIL groups had at least received some kind of bilingual training since their first year of upper secondary, which is grade 9 and 10 respectively, according to the Austrian and Swedish system of education (compare Sylvén 2004: 6; 44 and ESTIA 2003).

[57] This is mainly related to the different ways in which the two countries deal with TV series, films or movies produced in the English-speaking world: While in Sweden there seems to be a strong tendency for 'undubbed', if necessary, subtitled, original versions, the general standard for Austrian broadcasts are German synchronizations (compare also the comments in Sylvén 2004: 5; 69).

a limited extent. Moreover, as specific pieces of information are practically inaccessible, most explanations about the possible sources of the above-mentioned numerical differences are purely speculative. Thus, rather than praising the Austrians' superiority, on the whole, it seems more reasonable to conclude that, when split into CLIL and traditional group, the students attending the VBS school in Vienna show roughly analogous test results as their peers involved in Sylvén's study.

In the next sections, the comparison between the two studies will be further refined and included in the discussion of other relevant aspects, starting with gender-specific differences in the overall research findings:

4.3.3. Gender

In each of Sylvén's three test rounds, the male students performed better than their female peers in both the CLIL and the control group (compare Sylvén 2004: 77, 87 and 96 respectively). Figure 4:8 illustrates the related results for the present study:

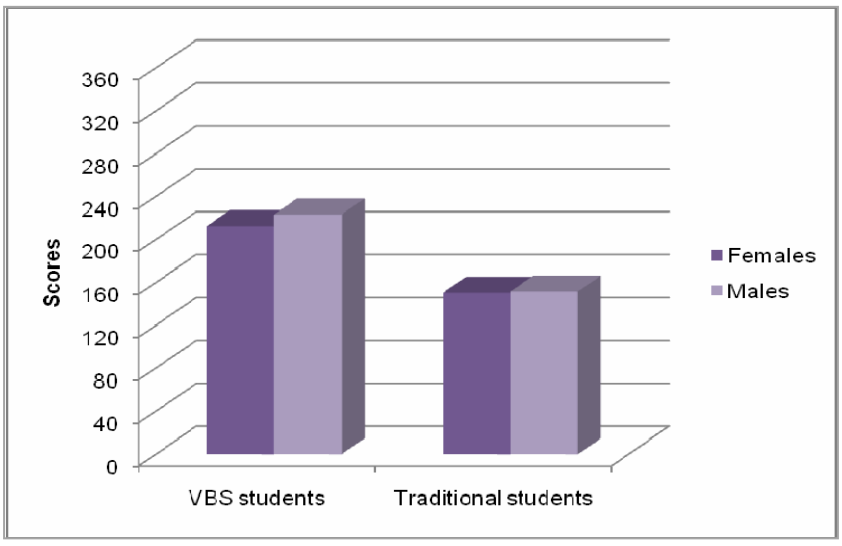


Figure 4:8 *Total mean score, group/factor: gender*

Similarly, the male participants score slightly higher than the females with total mean scores of 223.27 vs. 212.50 for the VBS class and 151.67 vs. 150.44 for the traditional class. Yet, while Sylvén (2004: 77, 87 and 96) discovered that the gender gap was larger for the control group than for the CLIL group, in the present context, the converse seems to apply. However, this situation cannot be interpreted as a

general rule, given the fact that this study's population size is disproportionately smaller than the original's, with a total number of 14 males (11 VBS, 3 traditional) and 19 females (10 VBS, 9 traditional) compared to those 201 and 162 respectively, participating in Sylvén's test sequence (see Sylvén 2004: 44).

Accordingly, an unpaired t-test with the categorization variable of gender does not yield significant outcomes for any of the two groups (see Table 4:9):

Table 4:9 *Unpaired t-test, total mean score, group/factor: gender*

Males vs. females	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	t-value, $t_{n_1+n_2-2;\ 1-\alpha}$		
				α = 0,05	α = 0,01	α = 0,001
VBS n_1 = 11, n_2 = 10	10,772727	18,511038	0,5819624	1,729133	2,539483	3,5794001
Traditional n_1 = 3, n_2 = 9	1,2222222	19,414644	0,0629536	1,812461	2,763769	4,1437005

Although in the VBS group the mean difference amounts to approximately 11 points, the test value remains below each of the t-values for the significance levels 0.05, 0.01 and 0.001. Thus, the gender variation is purely accidental, as is the case in the traditional sample. As for the latter, the insignificance of the test result has already been predictable from the minimal numerical difference between males' and females' average scores.

The issue of gender-related examination grades can also be explored from a more holistic perspective. Combining both test groups, the figures for the two sexes are as illustrated in Figure 4:9:

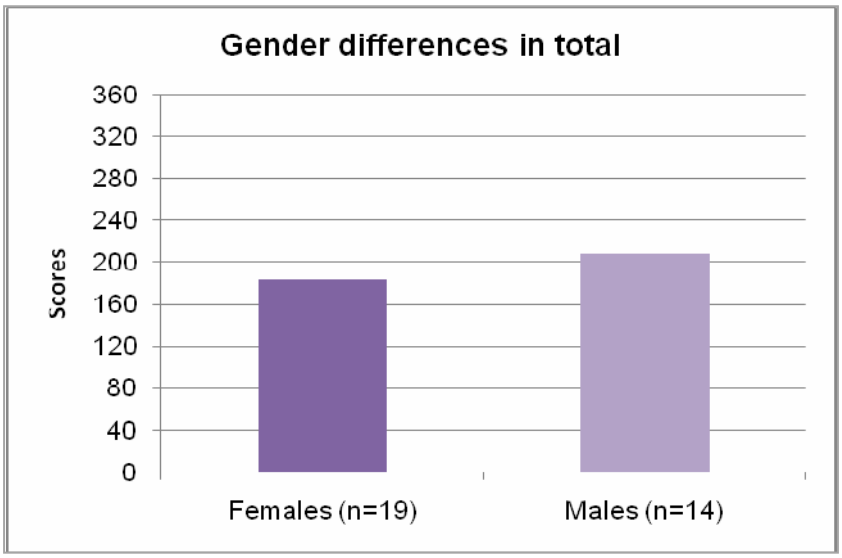


Figure 4:9 *Total mean score, factor: gender*

In the same way as in the two group-specific subsets, the male students also outperform their female peers if united into a single test population. The corresponding mean scores amount to 207.93 contrasted to 183.11 points. Due to the relatively low number of representatives of both genders, again, this difference cannot be proven to be statistically significant, as the test value is still smaller than the critical value, even at the alpha-level of 0.05 (see Table 4:10):

Table 4:10 *Unpaired t-test, total mean score, factor: gender*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
Males vs. females total n_1 = 14, n_2 = 19	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,69551874
				0,01	2,45282418
	24,82330827	17,08169615	1,453210972	0,001	3,37489928

Notwithstanding the lack of statistical corroboration, the male dominance in absolute scores is quite intriguing, especially in view of the common assumption that "[f]emale learners generally do better than male" (Ellis 1994: 202). The reasons for this apparent contradiction may be varied. As Sylvén points out,

> learning a language involves a multitude of elements of which only one is examined in the present study, namely vocabulary acquisition.
> (Sylvén 2004: 77)

Yet, it is surprising that, specifically in the context of the Swedish research paper and the present replication, the general assumption of female supremacy in language learning does not seem to be true.

The gender debate becomes even more controversial if students' scores are compared to their marks in English language, as recorded in the latest school report at the time of the test sequence.[58] As illustrated in Table 4:11 below, in the whole sample, the female participants are well ahead of their male counterparts with an average English mark of 2.42 contrasted to 3.29, whereas the preceding analysis of test scores has revealed exactly the opposite. The same paradoxical picture can be drawn when the population of 33 students is split into the two separate groups. Here, the corresponding figures are 3.09 vs. 2.30 and 4.00 vs. 2.56 for the male and female subjects within the VBS group and the traditional group, in that order. [59]

An unpaired t-test, based on the assumption of a normally distributed variable, shows that the gender difference in average marks is significant at the level of 0.01 in the entire test sample (test value = 2.82 > 2.45 = t-value), and highly significant in the VBS group, with a test value of approximately 3.60 compared to 3.58 for the t-value at the α-level 0.001. The result for the traditional group is indifferent, given the low amount of male representatives.

Table 4:11 *Total mean score vs. average English mark in digits per group and gender*

Sample	Gender	Total mean score	Average English mark
Whole population	male	207.93	3.29
	female	183.11	2.42
VBS	male	223.27	3.09
	female	212.50	2.30
Traditional	male	151.67	4.00
	female	150.44	2.56

The apparent discrepancy between the two sexes in test performance, on the one hand, and school performance, on the other hand, definitely requires closer attention.

[58] In this particular case, this was the half-yearly report at the beginning of February 2007.

[59] The figures are calculated from the students' responses to the question "Welche Noten hattest du im letzten Zeugnis in folgenden Gegenständen (Deutsch, Englisch, Mathematik)?" included in the last page of both student questionnaires, and checked against the information given in the school's register. The Austrian grading system is based on a five-point scale with "1 (Sehr gut)" being the best grade, followed by "2 (Gut)", "3 (Befriedigend)", "4 (Genügend)" and "5 (Nicht Genügend)" as the lowest grade, resulting in a fail.

In the following chapter, it will be shown whether there are significant variations between male and female students, also with regard to the individual test types in the present study. However, for a more detailed account of this complex issue, further research needs to be conducted.

5. Results on the separate tests

5.1. Self-report test

Figure 5:1 presents the mean score results of the *self-report* test (see section 3.4.1.) in the VBS group and the traditional group:

5.1.1. Total outcome

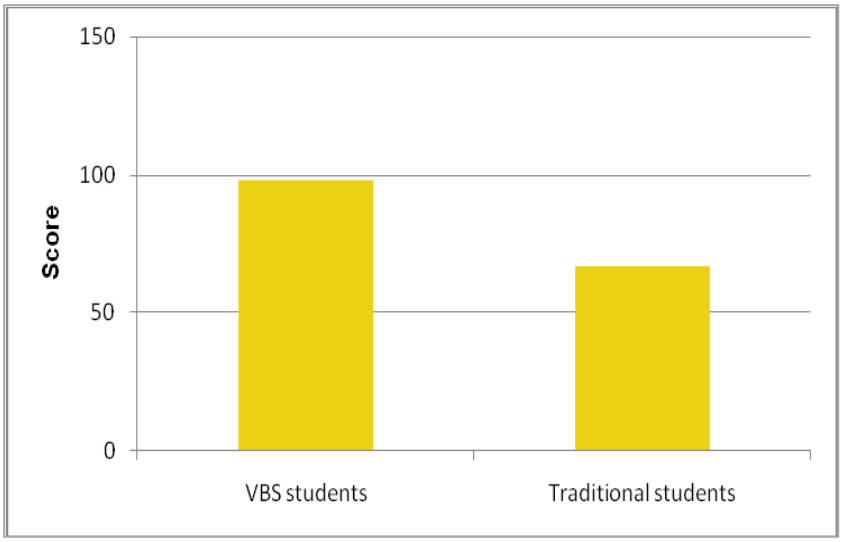


Figure 5:1 *Mean score self-report test*

Similar to the overall outcome accounted for in Chapter 4, in the *self-report* test, the VBS group is clearly in the lead with a mean score of 97.90 points compared to 66.42 points for the traditional group. In both sets, students' individual values closely adhere to the graph of the normal distribution function. Moreover, the paired F-test shows that the two standard deviations, respectively variances, are fairly equal except for the usual accidental variation:

Table 5:1 *Paired F-test: analysis of variance, results on the self-report test*

VBS vs. Traditional $n_1 = 21$, $n_2 = 12$	test value F_{pr}	α	F-value $F_{f_1,f_2;1-\alpha/2}$
	1,220884078	0,05	3,22614478
		0,01	4,85522025
		0,001	8,14150397

Based on these conditions, the difference in mean scores can also be evaluated in statistical terms. Table 5:2 illustrates the result of the corresponding unpaired t-test:

Table 5:2 *Unpaired t-test, mean score self-report test*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS vs. Traditional n_1 = 21, n_2 = 12	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,695518742
				0,01	2,45282418
	31,48809524	4,987541625	6,31334986	0,001	3,37489928

As the figures demonstrate, the test value considerably exceeds the critical value at the alpha-level 0.001, implying that the difference between the VBS group and the traditional group is highly significant. A more detailed assessment of students' responses will reveal that there are also remarkable differences in the answer patterns between these two groups. This analysis will be performed in section 5.1.4. As a next step, we look at the *self-report* results from the perspective of the students' mother tongue(s):

5.1.2. Results according to group and mother tongue

The relation presented in Table 5:2 remains unchanged if the values of the English native speakers are deducted from the VBS sample, the resultant mean score still being as high as 96.18 points. Furthermore, a reverse check, contrasting the original VBS assemblage with its non-English subset, corroborates that the exclusion of these particular four subjects has no substantial influence on the average test result. If the Anglophone VBS students are singled out as a separate group, it becomes apparent that, with an average score of 105.25, they are slightly ahead of their non-English classmates. Yet, this superiority proves to be statistically insignificant, as the outcomes of an unpaired t-test confirm (see Table 5:3):

Table 5:3

Unpaired t-test, mean score self-report test, VBS/factor: (non-) English mother tongue

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS+EL1 vs. VBS\EL1 n_1 = 4, n_2 = 17	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,72913279
				0,01	2,53948319
	9,073529412	7,852038857	1,155563488	0,001	3,57940015

As far as the factor German mother tongue is concerned, the situation appears fairly similar: Although the international students (for an exact definition of the term, cf., e.g. section 3.2.2.) actually score higher than their German-speaking colleagues, with a mean value of 105.14 compared to 94.29, there is no statistical evidence that this supremacy is also accountable for the supremacy of the entire VBS group over the traditional group. Table 5:4 illustrates that, even after excluding the seven multilingual students, the difference in the achievement between the CLIL class and its conventional counterpart is still highly significant, with a test value of approximately 5.31 contrasted with 3.47 for the critical t-value.

Table 5:4

Unpaired t-test, mean score self-report test, group/factor: German mother tongue

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS+GL1 vs. Traditional n_1 = 14, n_2 = 12	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,710882067
				0,01	2,492159469
	27,86904762	5,244756429	5,313697213	0,001	3,466777294

Summing up, in the same way as the overall findings (see 4.3.1.), the mean score results of the *self-report* test do not alter substantially, if, on the basis of their native languages, particular VBS subjects are included or excluded from the evaluation. For more transparency, the above-mentioned average scores of the different L1-specific subgroups are summarized in Table 5:5:

Table 5:5 *Summary, mean score self-report test, group/factor: mother tongue*

	Subgroups according to L1				
	Total	+EL1	\EL1	+GL1	\GL1
VBS	**97,90**	105,25	96,18	94,29	105,14
Traditional	**66,42**	-	-	= total	-

E = English, G = German, "+" = with, "\"= without

5.1.3. Results according to group, gender and research context

If the results of the *self-report* test are categorized according to the variable of gender and group, the following picture can be obtained:

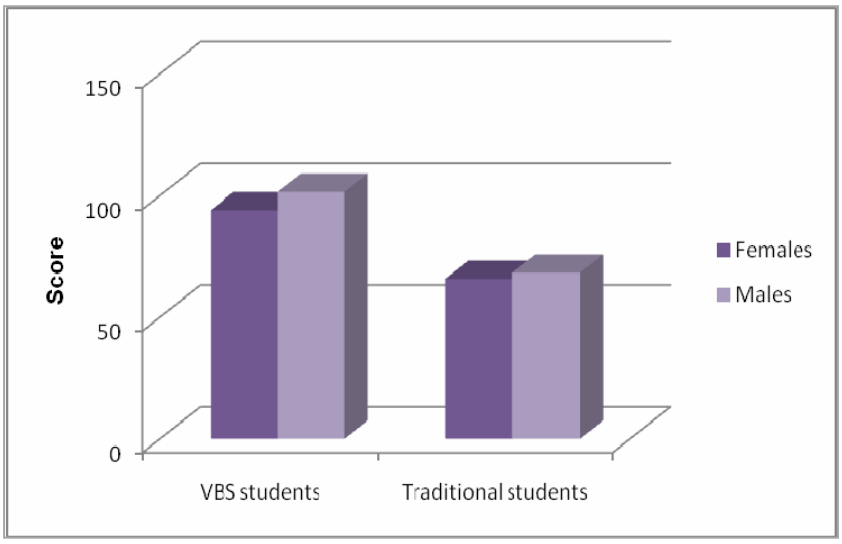


Figure 5:2 *Mean score self-report test, group/factor: gender*

Figure 5:2 illustrates that the male students outperform their female counterparts in both, the traditional group and the (original) VBS group, the corresponding mean scores amounting to 101.55 vs. 93.90 and 68.33 vs. 65.78 points respectively. As already noted in connection with the total results (see 4.3.3.), the gender difference is larger for the subjects being taught with the CLIL method. Yet, in either of the classes it is too small to assume statistical significance (At the α-level 0.05, the relation between the test value and the t-value is $1.24 < 1.73$ for the VBS group and $0.28 < 1.81$ for the traditional group).

In Sylvén's test round three, the male/female distinction for the *self-report* test yielded quite similar results, the only exception being that the gender gap in the Swedish control set proved to statistically valid (compare Sylvén 2004: 97-98). On the whole, a comparison of the two research contexts shows that, regardless of their gender, the Austrian traditional students are approximately on a par with their Swedish colleagues. However, concerning the CLIL groups, the VBS subjects are slightly ahead of the original test participants, with fairly equal differences between the means of both the male and the female subsets (see Sylvén 2004: 97). This variation in performance may again be traced to the fact that, according to the curriculum, the Viennese students have received CLIL instruction for a longer period of time than their Swedish peers. Alternatively, the VBS group's superiority may be connected with the modifications involved in adapting the original scoring mode to the requirements of the present study (see 3.4.1.). The following section attempts at providing further insights into this specific issue:

5.1.4. Distribution of answers among self-report categories

As illustrated in Chapter 3, in the *self-report* test, the participants were offered five different answer levels, according to which they had to estimate and, if necessary, also demonstrate their knowledge of each lexical item under examination. The definitions and instructions for these levels were as follows:

A. "I do not remember having seen this word before."
B. "I have seen this word before, but I do not know what it means."
C. "I have seen this word and think it means (*write a synonym or translation*)"
D. "I know this word. It means (*write a synonym or translation*)"
E. "I can use this word in a sentence (*write a synonym or translation in column D and then go to column E and write a full sentence which shows how the word is typically used*)"

Since quite a number of the present responses were difficult to classify on the basis of this system, a complex new scheme was developed (see Table 3:5), which, while still adhering to the original 1-5 point scale (compare Sylvén 2004: 52-53; 59), also takes into account twelve different types of 'deviating' or 'problematic' answers.

Corresponding to these modifications, in the marking of the *self-report* sheets, the students' responses were organized into the following (sub-) categories:

- **A** (includes answers of problem type *No. 1*)
- **B** (+ answers of problem type *No. 2*)
- **C correct** (+ *Problem No. 9*)
- **C incorrect** (+ *Problem No. 10*)
- **D correct** (+ *Problems No. 7* and *11*)
- **D incorrect**
- **E correct**
  (including the correct 'versions' of *Problems No. 3, 4, 5* and *6*, as well as problem type *No. 12*)
- **E acceptable**[60]
  (meaning that the sentence in column E is correct but ambiguous, since no further information is given, or the answer in column D contains some inaccurate grammar, i.e. *Problems No. 4* and *8*)
- **E incorrect**
  (comprises the incorrect 'versions' of *Problems No. 3, 4, 5* and *6*)

(For a detailed description of each problem type, see section 3.4.1.)

Table 5:6 illustrates the distribution of test answers in these categories or levels between the CLIL group and the traditional group in both the Austrian and the Swedish research setting:

Table 5:6

Distribution of answers (%) among self-report categories, group/research context

Answer level	*Seregély 2008*		*Sylvén 2004 TR III**	
	CLIL (VBS)	**Traditional**	**CLIL**	**Traditional**
A	9%	25%	10%	23%
B	22%	40%	32%	44%
C correct	2%	5%	4%	3%
C incorrect	4%	9%	10%	10%
D correct	5%	8%	4%	5%
D incorrect	3%	3%	3%	4%
E correct	39%	9%	27%	8%
E acceptable	3%	0%	0%	0%
E incorrect	13%	1%	10%	4%

*(compare Sylvén 2004: 186)

[60] This category is only used in the present research context.

If we first compare the proportions for the two Viennese test groups, the reasons for the overwhelming superiority of the CLIL learners over their traditional peers are immediately noticeable. Whereas answer level A, symbolizing total unfamiliarity with the item(s) under consideration (compare Table 3:4), accounts for only 9 percent of all *self-report* responses among the VBS students, in the traditional group, as many as 25 percent of the answers were categorized in the same way. On the contrary, option E, representing the highest level of lexical knowledge, was more frequently chosen in the bilingual class. Here, approximately 39 percent of all responses (compared to 9 percent in the traditional group) were classified as *E correct*, 13 percent (vs. 1 percent in the traditional group) were considered to be *(E) incorrect*, and 3 percent (0 percent) fell into the new category *E acceptable*. Level B, on the other hand, was especially popular in the traditional class, where 40 percent of the *self-report* responses (as opposed to 22 percent in the VBS class) were recorded in this second answer column.

Concerning level C, "I have seen this word and **think** it means ...", once again, the regular students were 'in the lead', with 5 percent of all answers having been categorized as *C correct* (2 percent in the VBS group) and 9 percent (4 percent) bearing the label *C incorrect*. These figures might indicate that, in accordance with the original assumption (compare, e.g. Sylvén 2004: 185), the traditional learners were indeed less self-confident about their knowledge of particular items under examination than their VBS peers were. However, if we compare the related proportions for the *actual* knowledge-level D ("I **know** this word. It means ..."), the differences between the two groups turn out to be less striking, but rather slightly contradictory. While in the traditional class, in total, 11 percent of the students' responses were organized into one of the D-categories (8 percent *D correct*, 3 percent *D incorrect*), the corresponding value for the VBS class amounts to only 8 percent (5 percent *D correct*, 3 percent *D incorrect*). At first glance, this seems to mean that "[t]he CLIL student is not necessarily more self-confident with regard to English vocabulary than the control student" (Sylvén 2004: 186). Yet, this conclusion proves to be completely absurd, as we have already shown that, instead of providing only a simple translation or synonym, "[t]he CLIL students who believe they know a word [most typically] go all the way to the E-level" (Sylvén 2004: 186)

in order to demonstrate their knowledge. Needless to say, this tendency is more representative of their (high) self-confidence than the outcomes in any other of the five *self-report* levels.

As regards the variation in the test results between the Austrian and the Swedish students (compare 5.1.3.), again, the figures in Table 5:6 are quite revealing. While, corresponding to the test scores, in both traditional groups, the distribution of answers among the five categories was almost equal (23 percent vs. 25 percent for answer level A in Sylvén's test round III and the present *self-report* test, in that order, 44 percent vs. 40 percent for level B, 13 percent vs. 14 percent for level C, 9 percent as opposed to 7 percent for level D, and 12 percent compared to 10 percent for level E in the Swedish and the Austrian setting respectively), in the two CLIL groups, proportional similarities were restricted to levels A (10 percent in Sylvén's test round III compared to 9 percent in the present context) and D (7 percent vs. 8 percent in the same order). Concerning categories B and C, the Swedes were well ahead of their Austrian peers with differences of 10 and approximately 7 percentage points respectively.

However, the most striking dissimilarity between the two groups can be noticed in connection with answer level E. Whereas in the present CLIL group, 54 percent of all *self-report* responses fell into this category, the corresponding figure in the original context was only 37 percent (see Sylvén 2004: 186). Even if we classified all *E acceptable* sentences as *incorrect* ones, the difference in the highest-score category (*E correct*) between the two CLIL samples would still amount to as many as 12 percentage points. It is therefore quite likely that the superiority of the Austrian CLIL learners over the Swedish ones mainly results from their strong preference for the most complex of all answer types.

5.2. Words in context test

In the five-part sequence, this particular test type (see also section 3.4.2.) immediately followed the *self-report* test. Figure 5:3 illustrates the average *words in context*-scores for the two test groups:

5.2.1. Total outcome

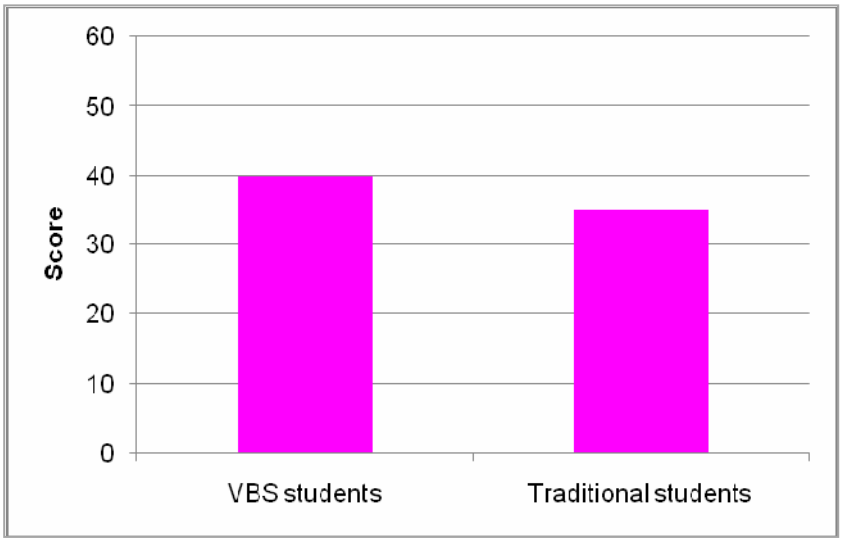


Figure 5:3 *Mean score words in context test*

Once more, the VBS group scores higher than its traditional counterpart. Yet, this time, the mean difference only amounts to an absolute value of 4.98 points. An unpaired t-test with the distinction CLIL/non-CLIL yields the following, rather unusual, result:

Table 5:7 *Unpaired t-test, mean score words in context test*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS vs. Traditional $n_1 = 21$, $n_2 = 12$	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,695518742
				0,01	2,45282418
	4,976190476	2,603757209	1,911157637	0,001	3,37489928

At the significance level 0.05, the test quantity t_{pr} exceeds the critical quantity with a value of 1.91 compared to approximately 1.70. For the levels 0.01 and 0.001, on the

other hand, the test value clearly remains below the t-value. Thus, the variation between the two groups is interpreted as *indifferent*:

> Indifferentes Resultat: Es kann weder eine Entscheidung für die Nullhypothese, noch eine Entscheidung für die Alternativhypothese gefällt werden. (DGQ 1993: A4, 12)

A possible explanation for an indifferent test outcome may be that the difference in mean scores between two data sets is simply brought about by considerable dissimilarities in the scoring pattern of **one** particular set (see also 4.2.). However, this does not seem to apply in the present context, since a paired F-test proves that the variance is fairly similar within the two groups (see Table 5:8):

Table 5:8 *Paired F-test: analysis of variance, results on the words in context test*

VBS vs. Traditional $n_1 = 12$, $n_2 = 21$	test value F_{pr}	α	F-value $F_{f_1,f_2;1-\alpha/2}$
	1,342442373	0,05	2,720861926
		0,01	3,755548863
		0,001	5,50343873

Consequently, the only way of decoding the indifferent result would be to perform an additional test run with an increased sample size (DGQ 1993: A4, 12). As this is organizationally infeasible, for the current discussion, it suffices to say that, with regard to the *words in context* test, the deviations between the two groups are statistically irrelevant. This relation appears quite astounding, if we consider that, hitherto, all analyses have confirmed the VBS students' overwhelming superiority over their traditional peers. Indeed, further evaluations will reveal that, within the entire sequence, the *words in context* test is the only test type allowing for approximately similar outcomes in both school classes.

5.2.2. Results according to group, gender and research context

The exceptional situation of the *words in context* test becomes even more obvious if the individual results are categorized according to the variable of group **and** gender:

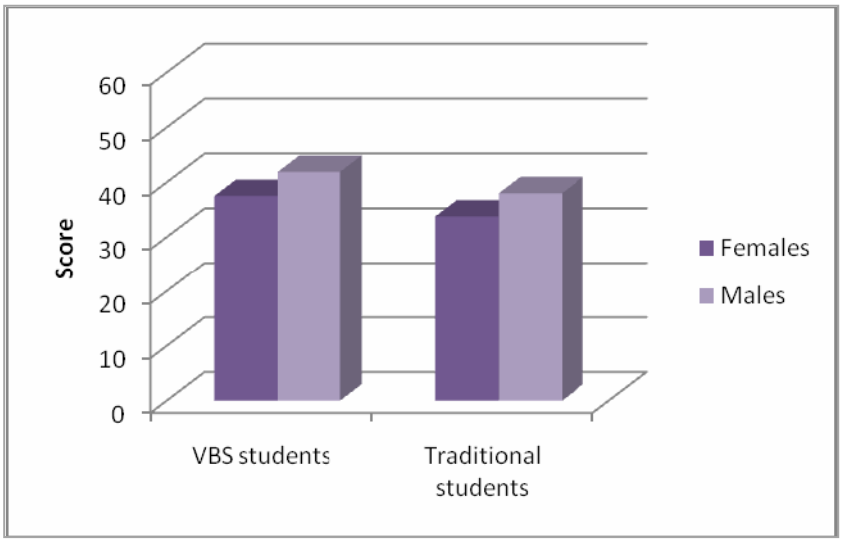


Figure 5:4 *Mean score words in context test, group/factor: gender*

As Figure 5:4 illustrates, with an average score of 41.91, the male VBS students occupy the top position in the ranking of the means. They are succeeded by the traditional males, who are approximately at the same score level as the VBS females (38 points contrasted to 37.5 points respectively). The traditional females rank last with the lowest mean value of 33.78 points. Again, neither of the internal gender gaps bears statistical significance.

In Sylvén's test round III, the results on the *words in context* test produced a slightly different picture: Even though, just as in the present context, in the two separate test groups the males scored higher than their respective female counterparts, in total, the CLIL subjects remained clearly ahead of their conventional peers (compare Sylvén 2004: 98). Actually, in the original study and the present replication the results for both the male and female CLIL students are fairly similar. On the contrary, the Austrian traditional students clearly outperform their Swedish colleagues, with mean score differences of approximately 7 and 9 points for the male and female subsets, in that order.

5.2.3. Results according to group and mother tongue

Returning to the variations within the Austrian context per se, the main reason for the score increase in the traditional group lies in the test design itself: As the name implies, the *words in context* test primarily focuses on "the students' ability to

understand words in the context of a longer text" (Sylvén 2004: 56). Additionally, to some extent, Henriksen's dimension three, the so-called "receptive-productive dimension" (Henriksen 1999: 306), is taken into account. Hence, for a full 2-point score, the meaning of the lexical item in question needs to be guessed correctly **and** be expressed in an adequate manner. In this respect, the productive aspect is not simply restricted to the target language English but also applies to the learners' linguistic competence in German, as the test instructions indicate:

> Read the article and explain the following words as they are used in the text. You can either translate the word into German, give a synonym in English, or explain the meaning in German or English. (see Appendix 3).

In the correction of the *self-report* test, it became apparent that the traditional students, unlike their VBS peers, mainly choose German translations instead of English synonyms as their answers in the performance categories C and D (for full details on the mode of answering see Chapter 7). This preference is even stronger in connection with the *words in context* test. Here, the option of using German, also for more complex explanations, proves highly beneficial for those participants who receive less EFL input or who, for different reasons, are less 'gifted' in the English language than their colleagues.

In contrast to the *self-report* test, which "measure[s] how well a person knows certain dimensions of a particular set of words at a given point in time" (Sylvén 2004: 54), the *words in context* test does not necessarily require any prior word knowledge on the students' part. Instead, for this particular test type, the meaning of each item under scrutiny may just as well be derived on the basis of five lexical cues, namely the co-text of the test word, the syntax of the accompanying sentence, the learners' knowledge of the world, their mother tongue(s), and possibly also other foreign languages (for a detailed account on the six sources of lexical inferencing, compare Haastrup 1991: 92-93; 239-244).

As regards the latter source, the traditional students have the advantage of attending the language-focused 'Gymnasium', meaning that they also study French and Latin as subjects at school, a supplementary linguistic background that may partially compensate for their deficient English vocabulary knowledge. The VBS programme, on the other hand, is only offered in the form of a so-called 'Realgymnasium' with a specific emphasis on natural sciences. Furthermore, the conventional group consists

entirely of native speakers of German, who definitely profit from the high degree of freedom allowed for defining the contextual meaning of the test words. Combining these two language-related factors with the above-mentioned extralinguistic knowledge sources, it seems no longer surprising that in the *words in context* test, the regular Austrian class is almost at the same mean score level as its VBS counterpart (see Figures:12 and 5:13 above).

The assumption that, to a certain extent, the present test type also reflects the participants' productive skills in German finds additional support if the VBS sample is divided in the usual L1-specific subgroups. As Table 5:9 illustrates, the average *words in context* score increases if the four English native speakers, and ultimately all subjects with non-German mother tongues, are excluded from the evaluation, whereas in the preceding analyses the converse has applied.

Table 5:9 *Summary, mean score words in context test, group/factor: mother tongue*

	Subgroups according to L1				
	Total	+EL1	\EL1	+GL1	\GL1
VBS	**39,81**	36	40,71	40,79	37,86
Traditional	**34,83**	-	-	= total	-

E = English, G = German, "+" = with, "\"= without

Actually, the highest mean score of 40.79 can be noticed in the subset of the German-speaking VBS students. By comparison, with a minimum of 36 points on average, the English subsample clearly remains below the VBS total. Undoubtedly, the numerical variations between the individual means are too small to be statistically significant. Nevertheless, selected answers reveal that a high level of German language competence also yields better results on the *words in context* test, as is the case in the following example:

(5) **Item:** *findings* (10)
Context: *While the jury is still out on that prediction, it would be hard to argue with the other findings of the research:*

Obviously, the German translation of the item *findings* seems to have presented some difficulties for the non-native students, as they came up with particularly creative word forms:

Answers:	'funden' [*sic!*]	[VBS.WIC.7m.EL1]
	'Findungen' [*sic!*]	[VBS.WIC.14m.EL1]
	'Fandungen' [*sic!*]	[VBS.WIC.18f.SL1] (S = Swedish)

Their German-speaking classmates, on the other hand, correctly noted the versions

'Ergebnisse' [VBS.WIC.4f; 5m; 10m; 17m; 19f; 20f.GL1]

and

'Resultate' [VBS.WIC.9f; 11m.GL1].

In the traditional group, the number of accurate German responses to test item (10) was even higher. Nine out of the twelve exclusively Austrian students decided on either 'Ergebnis(sse)/Funde' or 'Resultat(e)'. Only one student offered the slightly awkward translation 'herausgefundene Dinge', which was still categorized as *acceptable*.

In conclusion, the Austrian learners' performance on the *words in context* test not only depended on their lexical competence in English, but was also, albeit to an indeterminate degree, related to the factor (non-) German mother tongue. Since the present thesis is a replication, the instructions to this particular test type were similar to the original ones. Yet, Sylvén's (2004) study offers no details as to whether the participants' Swedish, just as their peers' German, language skills equally contributed to their average *words in context* scores. Therefore, the above-mentioned differences in achievement between the two traditional groups (see 5.2.2.) can hardly be accounted for.

5.2.4. Distribution of answers among words in context categories

As intimated above, in the correction of the *words in context* test, the students' answers were grouped into four different categories, viz. *correct, acceptable, incorrect* and *no answer*. Figure 5:5 illustrates the distribution of the test responses in these categories between the two groups. The figures are given in percentages of the maximum number of answers obtainable per group, that is, the respective number of students multiplied by the total amount of test items (i.e. 21 x 30 = 630 for the VBS and 360 for the traditional group):

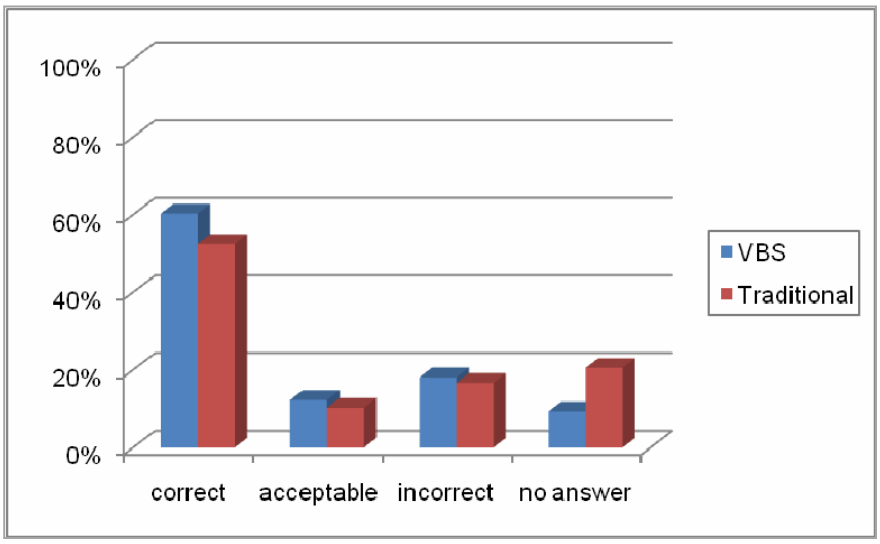


Figure 5:5 *Distribution of answers (%) among words in context categories*

As the first pair of columns demonstrates, around 60 percent of all answers in the VBS class were categorized as *correct*. For the traditional group, the corresponding figure amounts to 53 percent. With regard to the responses considered *acceptable* but not exactly "on target yield" (Sylvén 2004: 60), the values are 12 and 10 percent respectively for either of the two groups. In the category of *incorrect* answers, the VBS group attained approximately 18 percent compared to 17 percent for its traditional counterpart. The final pair of columns represents the number of occurrences where *no answer* was given for a particular test item. This specific response option was to a large extent avoided by the CLIL subjects, who came up with a value of barely more than 9 percent contrasted with almost 21 percent for their conventional peers.

The above figures provide additional support for what has already been discussed in connection with the mean score results, namely that, as far as the *words in context* test is concerned, the gap between the participants taught with the CLIL method, and those receiving regular EFL instruction, is only minor. Even though the traditional students tend to be less risk-taking, thus preferring *no answer*s to *incorrect* or only *acceptable* ones, the percentages for the two intermediate categories are fairly similar in both groups. Moreover, with a value of 7 percent, the degree of variation within the category *correct* is not considerably high, especially in comparison to the corresponding *cloze* test distributions (see 5.4.4. and 5.5.4.).

5.3. Multiple choice test

As explained in section 3.3.3., in the practical test administration, the *multiple choice* test was inserted between *cloze test 1* and *cloze test 2*. Figure 5:6 presents the mean score result for the two groups:

5.3.1. Total outcome

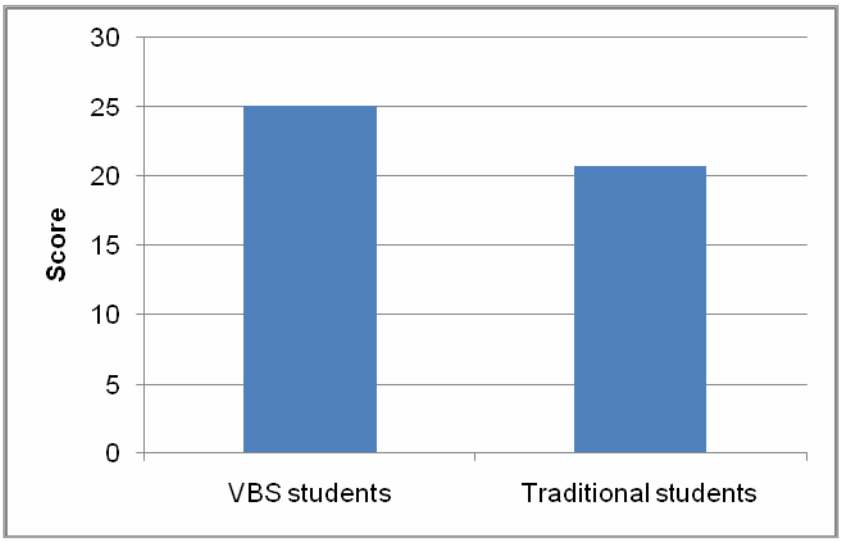


Figure 5:6 *Mean score multiple choice test*

Having attained an average score of 25.10 of a maximum of 30 points, the VBS students are well ahead of their traditional peers with a mean value of 20.67 points. The figures in Table 5:10 show that the difference between the two groups is statistically highly significant with a test value of approximately 3.81 contrasted with 3.37 for the critical value at the alpha-level 0.001:

Table 5:10 *Unpaired t-test, mean score multiple choice test*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS vs. Traditional n_1 = 21, n_2 = 12	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,695518742
				0,01	2,45282418
	4,428571429	1,163519691	3,806185202	0,001	3,37489928

5.3.2. Results according to group, gender and research context

In Sylvén's test round III, the average *multiple choice* test scores for both groups are considerably lower than those detected in the Austrian context (compare Sylvén 2004: 98). One possible explanation for these dissimilarities might be that, for the present purposes, the original test design was slightly altered to the students' advantage. As stated in section 3.4.3., two of the 30 items under examination were exempted from the underlying *multiple choice* convention permitting only **one** answer per sentence. For each of these particular test items, **two** alternative synonyms were accepted as correct. This modification, however, only accounts for approximately 1.24 points in the average VBS score and 1.33 in the comparative value of the traditional group, meaning that even 'under the original test conditions', the Austrian learners would have outperformed their Swedish peers in both groups. Given the closed format and the objectivity of the *multiple choice* test (compare Spolsky 1995: 179), in this connection, we can definitely speak of 'actual' or 'real' differences in mean results, whereas in the previous, non-psychometric tests, the variations between the subjects in Sylvén's and in my study could only be speculated upon.

Turning to gender differences, as the preceding sections have shown, in the *self-report* test as well as in the *words in context* test, the males achieved better results than the females in both groups. Figure 5:7 illustrates the related findings for the *multiple choice* test:

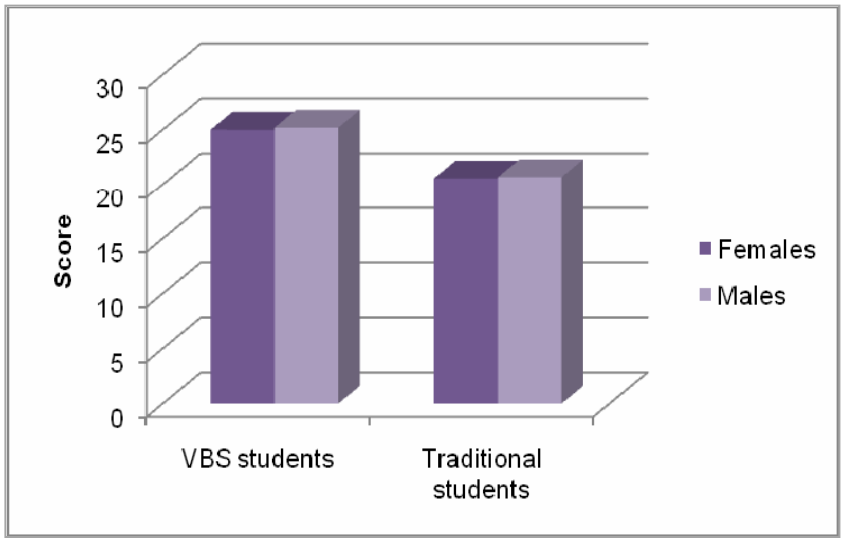


Figure 5:7 *Mean score multiple choice test, group/factor: gender*

As the two pairs of columns demonstrate, again, the male students are slightly ahead of their respective female counterparts in terms of average scores. Yet, this time, the mean differences are almost negligible, amounting to values of 0.18 and 0.11 for the VBS group and the traditional group, in that order. This outcome seems quite surprising, considering that a series of earlier studies (e.g. Bolger & Kellaghan 1990; Harding 1980; Hellakant 1994; Murphy 1982; Wood 1976, 1978) identified an overwhelming male superiority in examinations in the multiple choice test format. Notwithstanding this objection, it is in any case extremely doubtful whether the above figures represent an exception to a general tendency, since the sample under observation is too small to provide statistically valid results.

On the contrary, the outcomes of Sylvén's third *multiple choice* test round correspond quite closely to this 'common trend'. Thus, in both groups, the males are well ahead of their respective female counterparts, yet statistically so only in the group of traditional students (see Sylvén 2004: 99).

5.3.3. Results according to group and mother tongue

If the VBS sample is again split into the four L1-specific subgroups, we obtain the following individual mean score results:

Table 5:11 *Summary, mean score multiple choice test, group/factor: mother tongue*

	Subgroups according to L1				
	Total	+EL1	\EL1	+GL1	\GL1
VBS	**25,10**	26	24,88	25,14	25
Traditional	**20,67**	-	-	= total	-

E = English, G = German, "+" = with, "\"= without

As Table 5:11 illustrates, in contrast to the *words in context test*, now the subgroup of English natives takes the lead with an average of 26 test points compared to 24.88 for the non-English group. Yet, all in all, the L1-specific values neither deviate considerably from one another nor from the total VBS outcome. Moreover, even the difference between the smallest VBS 'sub-score' and the mean result of the traditional group is statistically significant at the level 0.001 (see Table 5:12):

Table 5:12

Unpaired t-test, mean score multiple choice test, group/factor: non-English mother tongue

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS\EL1 vs. Traditional n_1 = 17, n_2 = 12	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,703288423
				0,01	2,472659904
	4,215686275	1,21076851	3,481826822	0,001	3,421033621

Summing up, the *multiple choice* test clearly favoured the CLIL students and, among those, albeit to a lesser degree, especially the native speakers of English. The subsequent sections will reveal how the participants performed on the two *cloze* tests:

5.4. Cloze test 1

The first of the two gap-filling exercises is equivalent to the one used in Sylvén's third test round (compare Sylvén 2004: Appendix 3, 272-273). In the present five-part sequence, it was conducted in between the *words in context* and the *multiple choice* test (see also 3.3.3. and 3.4.4.). Figure 5:8 illustrates the two groups' mean score results on *cloze test 1*:

5.4.1. Total outcome

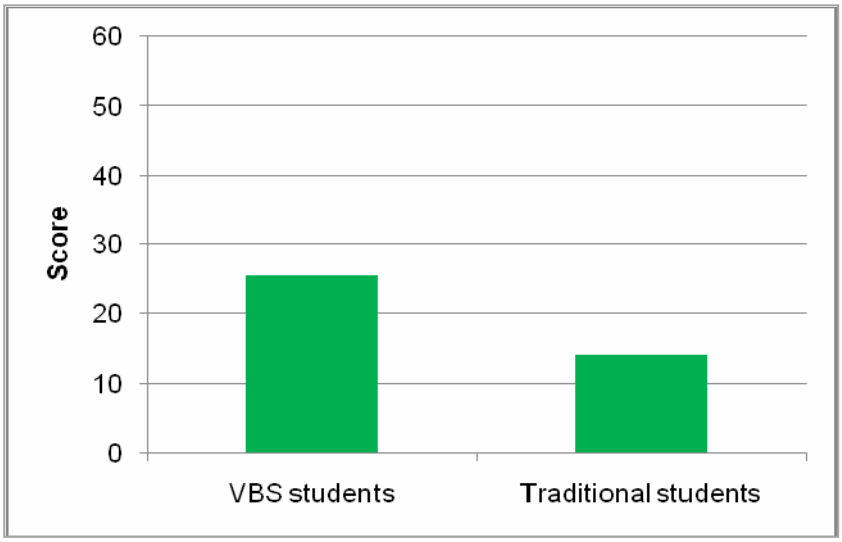


Figure 5:8 *Mean score cloze test 1*

As the bar chart shows, in comparison to the previous test types, the response rate for the first *cloze* test was relatively low. Out of a maximum total of 60, the VBS group gained no more than 25.48 average test points. The traditional group scored even considerably lower, with a mean value of only 14.08. Considering these figures, one is easily tempted to conclude that the difference in mean scores between the two groups is statistically significant. Indeed, an unpaired t-test with the distinction VBS/traditional yields the results illustrated in Table 5:13:

Table 5:13 *Unpaired t-test, mean score cloze test 1*

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS vs. Traditional n_1 = 21, n_2 = 12	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,695518742
				0,01	2,45282418
	11,39285714	3,38604836	3,364646908	0,001	3,37489928

The test value for the mean difference amounts to approximately 3.36 compared to 2.45 for the t-value at the significance level 0.01. Unfortunately, this relation proves to be statically invalid: As stated in the previous chapter (see 4.2.2.), one of the basic conditions for a reliable t-test is the similarity of the two standard deviations. While in the original context, this condition was merely assumed (see Sylvén 2004:71), in the present study, each respective sample pair has been subjected to a separate F-test. Hitherto, the corresponding results have always been favourable and have therefore only occasionally been included into the overall discussion. However, in connection with *cloze test 1*, the situation is more complex:

Table 5:14 *Paired F-test: analysis of variance, results on cloze test 1*

	test value F_{pr}	α	F-value $F_{f_1,f_2;1-\alpha/2}$
VBS vs. Traditional n_1 = 21, n_2 = 12	5,646271681	0,05	3,226144775
		0,01	4,855220247
		0,001	8,141503973

As Table 5:14 illustrates, for the first time, there is a significant difference in variance between the VBS sample and its traditional counterpart ($\alpha = 0.01$: test value F_{pr} > F-value). The unusually high dispersion of values within the VBS set is also

responsible for the differences in mean scores between the two test groups.[61] These interior variations become even more obvious if the participants' absolute *cloze test 1* scores are compiled and arranged in descending numerical order (see Figure 5:9):

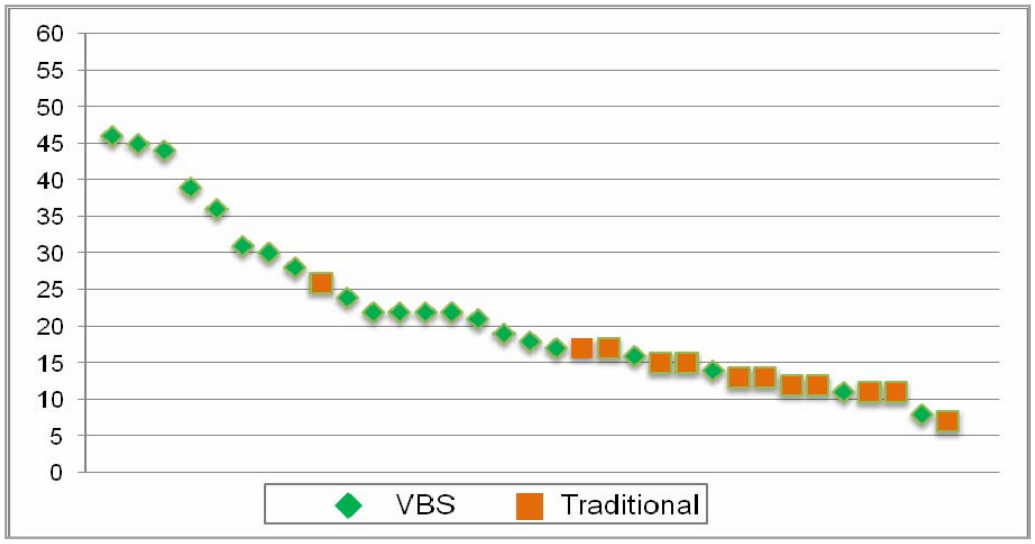


Figure 5:9 *Cloze test 1, ranking of individual test scores*

As the above figure shows, in the CLIL group, students' scores range from a maximum of 46 to a low value of 8 points. In the traditional group, on the other hand, apart from two exceptions (with 26 and 7 points respectively), all subjects attained values between 17 and 11. Even though, in sum, the VBS class performed better, four particular representatives did not even reach the top of the scale established by their conventional peers' values.

On the whole, the outcomes of *cloze test 1* reflect differences in the achievement of **individuals**, rather than differences between two specific test groups. In order to fully account for exceptionally high or unexpectedly low absolute scores, it would be necessary to carefully examine the students' extracurricular language influences. This is partly accomplished in Chapter 6. For the time being, the *cloze* test results will be analysed with regard to the factor mother tongue.

[61] Consequently, the above-mentioned t-test results are no longer relevant.

5.4.2. Results according to group and mother tongue

In the above ranking, three out of the five highest performing subjects (with scores between 46 and 36 points) are native speakers of English. The remaining two are, on the one hand, the absolute top scorer of the entire test sequence, a girl with Albanian as her first and English as her second language, and on the other hand, a boy who spent 5 years at an international, exclusively English-speaking, school abroad. This result is not particularly surprising, as it has often been claimed that especially receptive knowledge and correct use of complex lexical phrases and idioms "is [...] what most distinguishes advanced learners from intermediate ones" (Thornbury 2002: 116) and also "more frequently displays apparent native characteristics" (Arnaud & Savignon 1997: 157).

The fact that, in contrast to the previous test types, where the factor English L1 only played a subordinate role, *cloze test 1* almost **required** native (-like) language competence, is also mirrored in the mean score outcome of the respective VBS subgroup (see Table 5:15):

Table 5:15 *Summary, mean score cloze test 1, group/factor: mother tongue*

	Subgroups according to L1				
	Total	+EL1	\EL1	+GL1	\GL1
VBS	**25,48**	35,5	23,12	22,64	31,14
Traditional	**14,08**	-	-	= total	-

E = English, G = German, "+" = with, "\"= without

While the non-English subset came up with a mean value of 23.12, the four Anglophone VBS students attained as many as 35.5 average test points. Although in this particular case, the standard deviations are approximately equal (1.08 for the test value compared to 10.34 for the F-value at the level 0.001), the analysis of means still only yields an indifferent result, given the small size of the English subsample:

Table 5:16

Unpaired t-test, mean score cloze test 1, VBS/factor: (non-) English mother tongue

				α	t-value $t_{n_1+n_2-2;\ 1-\alpha}$
VBS+EL1 vs. VBS\EL1 $n_1 = 4$, $n_2 = 17$	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	0,05	1,72913279
				0,01	2,53948319
	12,38235294	5,668339835	2,184476108	0,001	3,57940015

In connection with the other VBS subgroups, it is impossible to make any general statements, because the differences between individual subjects seem to depend on a variety of factors, notwithstanding students' (non-German) native languages.

5.4.3. Results according to group, gender and research context

As illustrated above, *cloze test 1* is an exception to all the previous tests, as it clearly favours the English-speaking participants over their non-English peers. What is more, this particular test type also proves to be different in yet another respect: So far, the gender-related analyses have revealed that the male subjects, though at altering levels, consistently outperform their female counterparts in the areas tested. However, if the male/female distinction is applied to *cloze test 1*, we are presented with the following unexpected result:

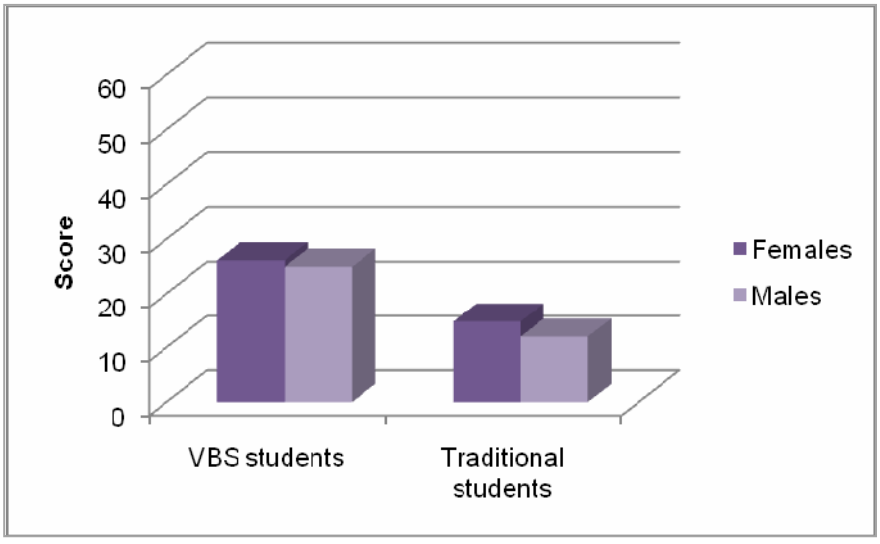


Figure 5:10 *Mean score cloze test 1, group/factor: gender*

In the VBS class, as well as in the traditional class, for the first time, the females scored above the males with mean differences of 1.19 and 2.78 points respectively. This relation only partly corresponds to the findings recorded in Sylvén's test round III: Here, the female CLIL students, too, outperformed their male colleagues by 1 point. Yet, in the control group, the males achieved to maintain their usual superiority (Sylvén 2004: 102-103).

In total, the average result for each of the two Austrian CLIL subgroups is about 8 points higher than in the corresponding original subsets. The Viennese traditional females are also slightly ahead of their Swedish peers. However, with regard to the traditional males, the converse holds true, as the Swedes attained approximately 4 points more than their Austrian counterparts (compare Sylvén 2004: 102, Figure 5:33).

Even though it is beyond the scope of the present study to fully account for these gender- and context-specific differences, an analysis of the distribution of students' answers in the specific *cloze* test categories will reveal some interesting underlying tendencies.

5.4.4. Distribution of answers among cloze test categories

Just as in the *words in context* test (see 5.2.4.), in the marking of the two *cloze* tests, a distinction was made between the four answer categories *correct*, *acceptable*, *incorrect* and *no answer*. Figure 5:11 illustrates the related results for the two test groups:

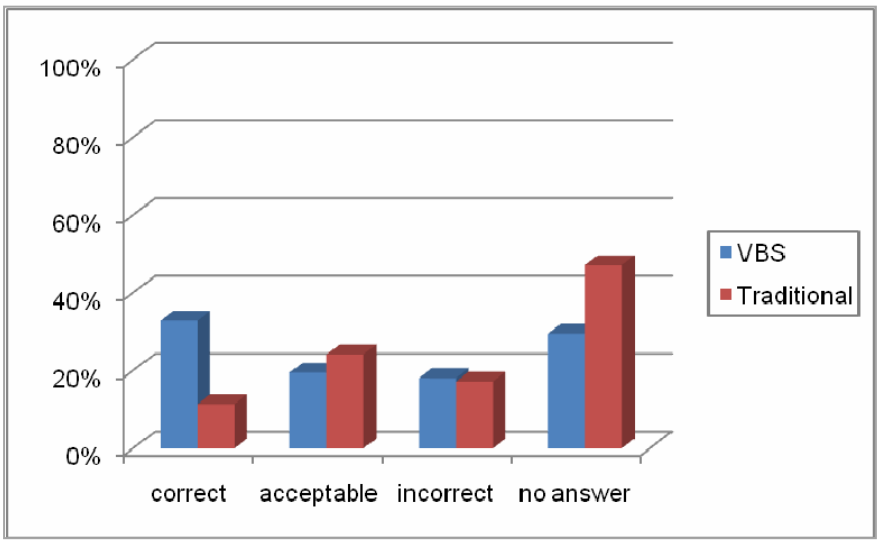


Figure 5:11 *Distribution of answers (%) among cloze test categories (CT1)*

While the open design of the *words in context* test partly compensated for deficient English language skills, the relative fixedness of the *cloze* test phrases required a sound knowledge of the English lexicon. Not surprisingly, there were fewer *correct* answers in *cloze test 1* than in the *words in context* test, namely 33 percent (60

percent in the *words in context* test) for the VBS class and 11 percent (53 percent) for the traditional class. The number of answers classified as *acceptable* was higher, with values of 20 percent (12 percent) and 24 percent (10 percent) for the VBS class and the traditional class, in that order.

As for the *incorrect* responses, the proportion was equal to that in the *words in context* test: 18 percent for the group of CLIL students and 17 percent for their conventional peers. In the category *no answer,* there were considerable variations, not only between the two test types, but especially also between the two test **groups**. In this particular response class, the traditional students were definitely 'in the lead' with more than 47 percent of 'zero' responses (21 percent for the *words in context* test) compared to 30 percent (9 percent) for their VBS counterpart.

In sum, the CLIL learners proved to be more expert at finding the **exact** items missing in the lexical units under consideration. The traditional students, on the other hand, seemed to be linguistically more creative, thus providing the higher amount of *acceptable* answers. Moreover, for both groups, yet especially so for the traditional group, one of the most preferred options was *no*t giving any *answer* at all, which may serve as an indication as to the complexity of the underlying language task (for the students' estimation of the *cloze* test type, compare also Sylvén 2004: 58).

In Sylvén's test round III, the distribution of *cloze* test answers was quite different (see Table 5:17):

Table 5:17

Distribution of answers among cloze test categories (CT1), group/factor: research context

Answer category	CLIL		Traditional	
	Seregély 2008	*Sylvén TR III**	*Seregély 2008*	*Sylvén TR III*
correct	33%	21%	11%	14%
acceptable	20%	14%	24%	16%
incorrect	18%	41%	17%	40%
no answer	30%	23%	47%	29%

*(compare Sylvén 2004: 102)

As regards the number of *correct* responses, the value for the CLIL group was 12 percent lower than in the Austrian context. In the traditional group, the situation was converse: Here, the Swedes outperformed their Viennese peers, although by only 3 percent. Within the category *acceptable*, both Swedish groups scored below their

respective Austrian counterparts. On the contrary, their amount of *incorrect* answers was considerably higher: 41 percent for the CLIL group and 40 percent for the control group. These variations can mainly be attributed to differences in the marking of the original, as opposed to the present, *cloze* test sheets. What is more, the diverging ratios between *incorrect* and *no answers* also indicate that, altogether, the Swedish participants were more risk-taking in their task-fulfilment than the Austrian students.

In sum, 53 percent of all answers in the Austrian CLIL group were rewarded with either 1 or 2 points. The corresponding original figure amounts to only 35 percent, which is the same value as for the Austrian traditional group. In the Swedish control group, the number of 0-point responses was considerably higher: 69 percent compared to approximately 30 percent for the two 'score-yielding' categories. Considering these dissimilar allocations, it becomes obvious that either of the two Viennese groups achieved a higher average *cloze* test result than its respective Swedish complement (for the 'original' *cloze* test values, compare Sylvén 2004: 103).

In a like manner, the categorization of test responses also serves to explain some of the above-mentioned 'internal' differences, existing between the male and female participants of the present study. Table 5:18 once more illustrates the distribution of the Austrian *cloze test 1* results, this time according to the additional factor of gender:

Table 5:18

Distribution of answers among cloze test categories (CT1), group/factor: gender

Answer category	VBS		Traditional	
	female	**male**	**female**	**male**
correct	34,33%	31,52%	11,48%	11,11%
acceptable	18,33%	20,61%	26,30%	17,78%
sum	**52,67%**	**52,12%**	**37,78%**	**28,89%**
incorrect	16,67%	19,39%	17,41%	16,67%
no answer	30,67%	28,48%	44,81%	54,44%
sum	**47,33%**	**47,88%**	**62,22%**	**71,11%**

As the figures reveal, in the category of *correct* answers, the traditional females are approximately on a par with their male colleagues. In the VBS group, the females are

slightly ahead of the males, a relation which is yet again reversed in the class of *acceptable* responses. In the traditional group, the females' superiority also remains valid for this second answer category. As for the lexical items classified as *incorrect*, the VBS males show the highest proportion, followed by the traditional females and, ultimately, also by the traditional males and the VBS females, who share exactly the same percentage figure.

As mentioned above, the category *no answer* enjoyed great popularity among the traditional subjects and, as Table 5:18 shows, in particular, among the male representatives of that very group. In the VBS class, on the other hand, it was rather the females who favoured the most 'convenient' of all answer options. Apparently, these preferences are symbolical, only for the (small) samples under consideration and do not reflect any underlying general tendencies.

If we finally add up the percentages of the categories *correct* and *acceptable*, for each individual subgroup, the gender-specific differences in the mean score outcomes of *cloze test 1* are perfectly understandable. Corresponding to their average test result, the traditional males attained the lowest proportion of 'score-yielding' answers (28.89 percent) compared to the VBS females who turned out to be the top scorers within the present test type.

5.5. Cloze test 2

Cloze test 2 eventually concluded the five-part test sequence. In the VBS class, it was conducted immediately after the *multiple choice* test. In the traditional class, however, it had to be postponed to a spare lesson on the subsequent day, when, unfortunately, one of the participants was missing from school. Nevertheless, since this was the only instance of absenteeism throughout the entire test series, the analysis of students' answers could be continued unimpeded.

5.5.1. Total outcome

Figure 5:12 illustrates the mean score result of the two groups on the final vocabulary test:

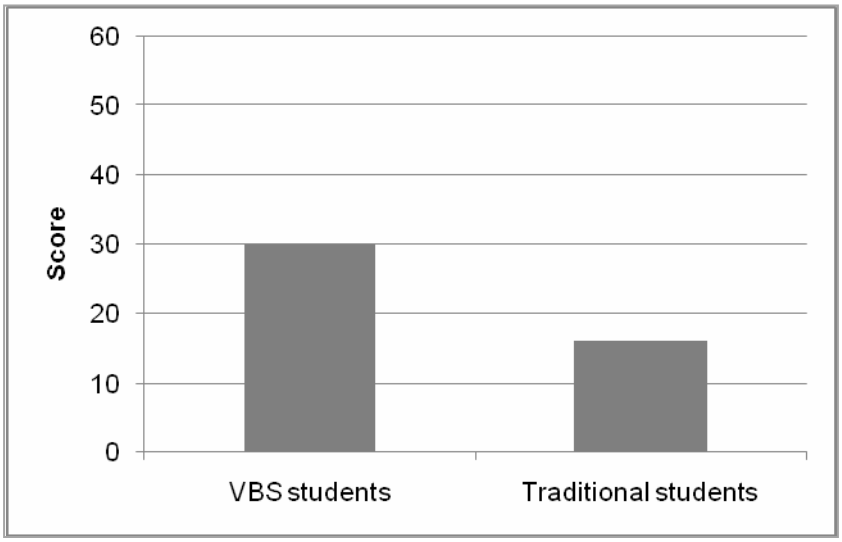


Figure 5:12 *Mean score cloze test 2*

As the figure reveals, the average values for both groups are slightly higher compared to those in *cloze test 1*, namely 29.86 for the VBS group and 16.09 for the traditional group. Yet again, this variation cannot be evaluated statistically, as there are significant differences in the standard deviations between the two samples (see Table 5:19):

Table 5:19 *Paired F-test: analysis of variance, results on cloze test 2*

VBS vs. Traditional n_1 = 21, n_2 = **11 (!)**	test value F_{pr}	α	F-value $F_{f_1,f_2;1-\alpha/2}$
	7,186469003	0,05	3,418543516
		0,01	5,274016749
		0,001	9,165072201

With a value of approximately 7.19, this time, the test quantity F_{pr} is even larger than for *cloze test 1* (see Table 5:14) and thus, clearly exceeds the F-values for both alpha-levels 0.05 and 0.01. This means that, among the VBS subjects, the *cloze test 2* scores are significantly more dispersed than among the traditional students, which becomes also noticeable in the following illustration:

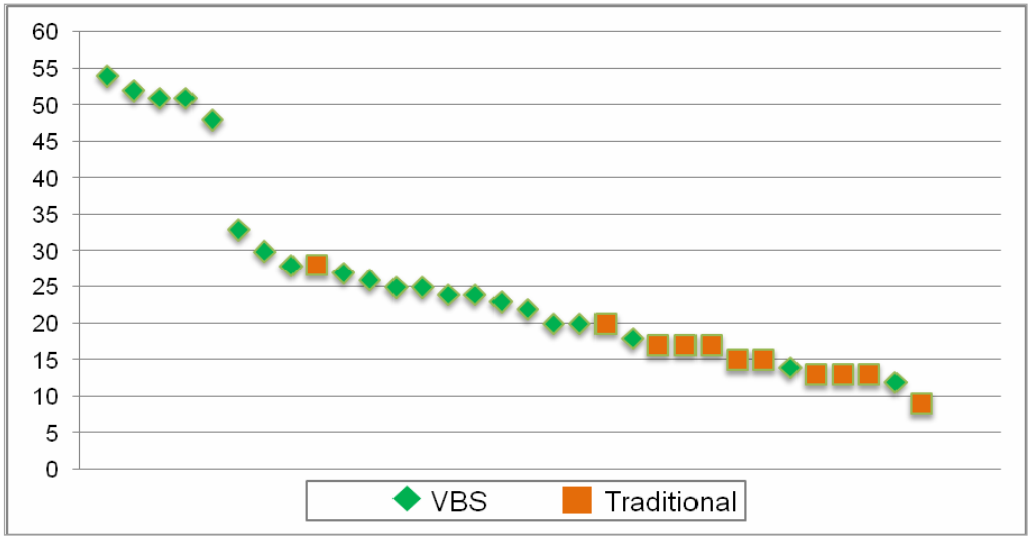


Figure 5:13 *Cloze test 2, ranking of individual test scores*

As Figure 5:13 illustrates, the scale of individual VBS results extends from an absolute value of 54 to 12 *cloze test 2* points, with a discontinuity between the score levels of 45 and 35. This range is four times intersected by one or more representative(s) of the traditional group, who attained values from 28 to 9, most frequently, however, from 17 to 13 points (compare also Figure 5:9). Just as in the first *cloze* test, on balance, the VBS group scored higher than its conventional counterpart. Still, three particular subjects remained clearly below the upper level of the orange scale. Interestingly, with regard to the results on *cloze test 2*, the top scorer of the traditional group (28 points compared to 26 in *cloze test 1*) is again among the 10 highest performing subjects of the entire population. As for the extreme variations within the VBS class, an analysis of the average test outcome according to students' L1s will probably provide more clarity:

5.5.2. Results according to group and mother tongue

In the previous test, the subgroup of English natives achieved the highest average score. Table 5:20 presents the L1-specific results for *cloze test 2*:

Table 5:20 *Summary, mean score cloze test 2, group/factor: mother tongue*

	Subgroups according to L1				
	Total	+EL1	\EL1	+GL1	\GL1
VBS	**29,86**	40,5	27,35	26,43	36,71
Traditional	**16,09**	-	-	= total	-

E = English, G = German, "+" = with, "\"= without

Once more, the Anglophone VBS students were in the lead with a mean value of 40.5 compared to 27.35 for their non-English classmates and 26.43 for the subset of Austrian natives. Yet, the differences in the mean scores between these subgroups are a bit misleading, since, in contrast to *cloze test 1*, in the test under consideration, only **two** English natives were among the five best performing subjects in the VBS class (and the population as a whole). Their remaining 'compatriots', on the other hand, obtained only mediocre test scores. Hence, the relatively high value for the VBS^{+EL1} mean predominantly results from the above-mentioned two, exceptionally high, individual scores, rather than displaying an inherent superiority of the entire English subgroup.

Similarly, the low mean value for the German-speaking VBS subset is slightly deceptive, as it does not take into account that three representatives of the very same group actually scored **above** their Anglophone classmates and thus, also joined in the select circle of *cloze test 2*-top performers. Summing up, the differences in the lexical achievement between individual (CLIL) learners proved to be more relevant to the present test result than the differences between the L1-specific subsets of the VBS group as a whole.

5.5.3. Results according to group, gender and research context

Compared with the Austrian results on *cloze test 2* (see 5.5.1.), the total mean values for both test groups were lower in the Swedish study, namely 18 points for the CLIL group and 14 points for the control group as a whole (Sylvén 2004: 93). This is scarcely surprising, if we consider that, originally, the test under consideration was performed in round II (compare Sylvén 2004: Appendix 2, 262-263), at the end of 10^{th} form (Sylvén 2004: 46), whereas the present *cloze test 2* was conducted in the second half of year 11. Moreover, only 34 percent of the Swedish students' answers fell into either of the two score-yielding categories *correct* (28 percent) or *acceptable* (6 percent). In the control group, the corresponding proportions were even lower, with 21 percent of all responses having been classified as *correct* and 5 percent bearing the mark *acceptable*. Section 5.5.4. will present the related distributions for the Austrian research context.

Concerning gender differences, it has been shown that *cloze test 1* ended with an unexpected outcome for the female participants in both the VBS group and the traditional group. Figure 5:14 illustrates the related mean score results on *cloze test 2*:

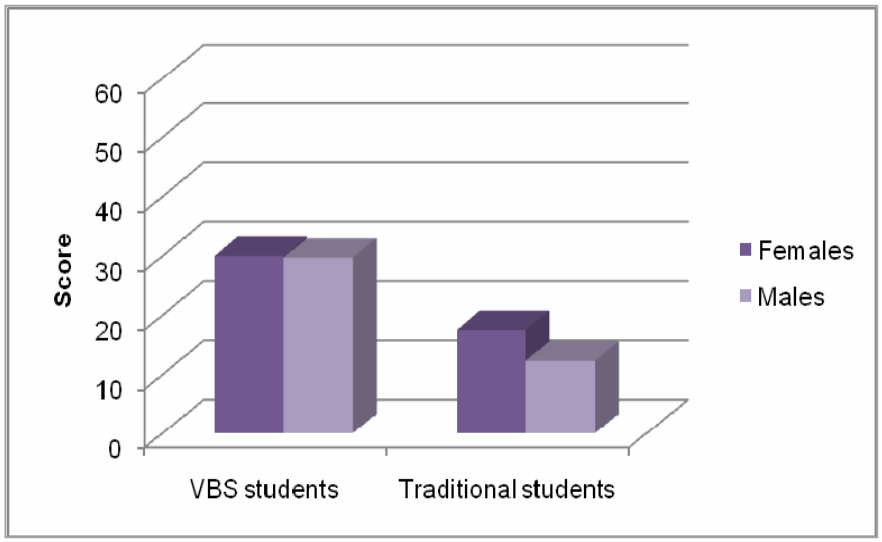


Figure 5:14 *Mean score cloze test 2, group/factor: gender*

Again the female VBS students outperformed their male peers, altough this time by only 0.27 points (1.19 points in *cloze test 1*). In the traditional group, the difference is more remarkable: Here, the females scored 17.5 points compared to approximately 12.33 for their male colleagues. Still, this variation is not statistically significant either ($\alpha = 0.05$: $t_{pr} = 1.69 < 1.83$ = t-value), given that the entire traditional group comprised no more than three male students.

Just as in the present context, in the original study, the female CLIL learners scored above their male peers and attained the highest average *cloze* test score. In the control group, however, the males achieved better results. Yet, neither of the gender-specific differences turned out to be statistically relevant (Sylvén 2004: 94).

5.5.4. Distribution of answers among cloze test categories

As the figure below illustrates, there were more *correct* answers in *cloze test 2* compared to the results in *cloze test 1*: 44 percent for the VBS group (33 percent in *cloze test 1*) and 17 percent (11 percent) for the traditional group.

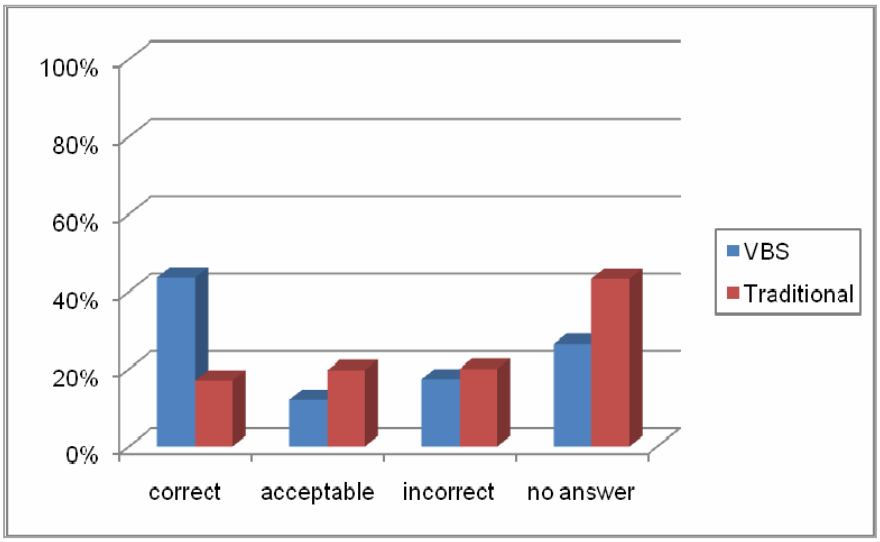


Figure 5:15 *Distribution of answers (%) among cloze test categories (CT2)*

The number of answers classified as *acceptable* was lower: 12 percent (20 percent) for the group of CLIL learners and 20 percent (24 percent) for their conventional peers. Again, the VBS students turned out to be more knowledgeable about the precise wordings of the lexical phrases under study, while their traditional colleagues tried to compensate lacking expertise by a higher degree of linguistic creativity. Unfortunately, their inventiveness did not always lead to favourable outcomes. Consequently, the proportion of *incorrect* answers was also higher within the traditional group: 20 percent (17 percent in *cloze test 1*) as opposed to 17 percent (18 percent) for the VBS group. Regarding the category *no answer*, the ratio between the two groups was even more striking: Whereas in the CLIL group, approximately 27 percent (30 percent) of all response gaps were left empty, the corresponding figure for the traditional group amounted to 43 percent (47 percent). Overall, for both groups, the total amount of score-yielding answers was slightly higher than in the first *cloze* test, which may be partly due to the fact that, originally, the test under consideration was designed for students at an age level one year inferior to the Austrian participants'. However, the most plausible explanation is that the students had simply become more familiar with the test format.

As for the differences in the mean score outcomes of the gender subgroups, the following table provides more detailed insights:

Table 5:21

Distribution of answers among cloze test categories (CT2), group/factor: gender

Answer category	VBS		Traditional	
	female	**male**	**female**	**male**
correct	45,67%	41,82%	18,75%	12,22%
acceptable	8,67%	15,15%	20,83%	16,67%
sum	**54,33%**	**56,97%**	**39,58%**	**28,89%**
incorrect	14,00%	20,30%	20,83%	17,78%
no answer	31,67%	21,82%	39,58%	53,33%
sum	**45,67%**	**42,12%**	**60,42%**	**71,11%**

In the subgroup of the VBS males, 41.82 percent of all answers were categorized as *correct*. The corresponding figure for the VBS females was approximately 4 percentage points higher, thus contributing to the variation in the average values between these two specific subgroups (compare Figure 5:14).

Indeed, it is exclusively the difference within the first response category that accounts for the subtle predominance of the VBS females over their male peers. The proportions in the remaining answers classes, on the other hand, are more likely to demonstrate the converse: While the VBS males recorded 15.15 percent *acceptable* answers, in the same category, their female colleagues attained only 8.67 percent. With regard to the *incorrect* responses, the VBS females scored below their male counterparts (14 percent vs. 20.3 percent), yet, their proportion of *no answer*s was almost 10 percentage points higher compared to that of their male classmates (31.67 percent compared to 21.82 percent). In the entire VBS sample, the total number of 0-point responses was higher among the females than among the males (45.67 percent contrasted with 42.12 percent). Nevertheless, the female CLIL learners achieved a slightly better average result, as they were ahead in the category yielding the highest score.

In the traditional class, the distribution of *cloze 2* answers produced a somewhat different picture: Just as in the group of VBS students, the females outperformed the males in the category of *correct* responses (18.75 percent vs.12.22 percent). Moreover, with a proportion of 20.83 percent compared to 16.67 percent for their male peers, the traditional females were also in the lead in connection with the answers classified as *acceptable*.

The distribution of *incorrect* responses was approximately similar: 20.83 percent for the traditional females and 17.78 percent for the male participants in the same test group. In a like manner as in the first *cloze* test, the option *no answer* was especially preferred by the male traditional students, who came up with a ratio of 53.33 percent (54.44 percent for *cloze test* 1), contrasted with their classmates' 39.58 percent, for this particular response category.

If we compare the two subgroups' total amount of score-yielding (as opposed to 0-point) answers, it is perfectly explicable that, in the test under consideration, the traditional females were ahead of their male peers by as many as 5.17 points on average.

5.5.5. Anchor items

To facilitate the process of test correction and more easily trace the participants' development throughout the course of the study, ten of the items included in the *cloze* test were so-called *anchor items* and, thus, identical in each of Sylvén's three test rounds (compare Sylvén 2004: 103). For the present research, two of the original *cloze* tests were utilized. However, since the entire five-part sequence was conducted within a span of scarcely one school week per class, in this context, statements about the students' progress seem barely justifiable.

Nevertheless, in order to preserve the original purpose of the recurrent items, the notion of advancement was re-created, as it were, on a miniature scale: As mentioned earlier, *cloze test 1* and *cloze test 2* were performed on two separate days, and in between, the *multiple choice* test was inserted. The main idea behind this specific way of structuring was to reduce the risk of careless mistakes due to exhaustion (see also Sylvén 2004: 47) and, in addition allow the students to refresh their knowledge, if necessary.

Figure 5:16 presents the outcomes of the ten anchor items in the two *cloze* tests according to the factor of group:

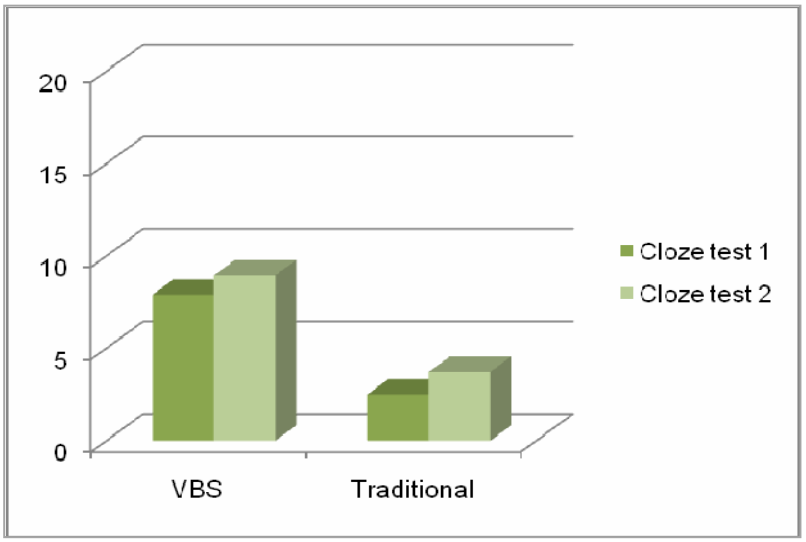


Figure 5:16 *Mean score results on anchor items, cloze test 1 and 2*

As the bar chart illustrates, both the VBS group and the traditional group managed to increase their mean score result by approximately 1 point each. Out of a maximum score of 20, on average, the CLIL learners attained 7.90 points in the first, and 8.95 points in the second *cloze* test. The traditional group scored considerably lower: 2.5 points on the anchor items in *cloze test* 1 and 3.7 points on the same objects in the second test round.

If the students' answers are grouped according to the usual *cloze* test categories, we obtain the following picture:

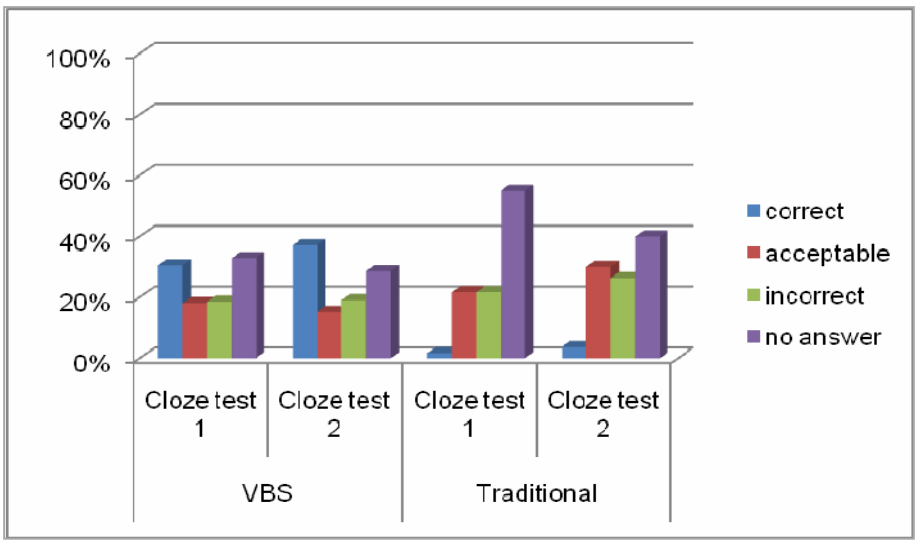


Figure 5:17 *Distribution of answers (%) among cloze test categories, anchor items, cloze test 1 and 2*

In the VBS group, the number of answers categorized as *correct* was higher in the second cloze test than in the first one, the corresponding proportions being 37 and 30 percent respectively. The same relation also applies to the traditional group, yet, in this specific connection, the figures are noticeably lower: approximately 2 percent for *cloze test 1* and 4 percent for *cloze test 2*.

As for the anchor items marked as *acceptable*, the proportion decreased in the group of VBS students (18 percent in *cloze test 1* and 15 percent in *cloze test 2*) and increased for their traditional peers (22 percent as opposed to 30 percent). In the *category* of incorrect responses, the CLIL learners maintained a constant level of 19 percent, while their traditional counterparts scored 4 percentage points higher in the second compared to the first *cloze* test (26 percent vs. 22 percent).

The increase in the two intermediate categories (*acceptable* and *incorrect*) also indicates that the conventional students became more courageous in the course of the test sequence. Correspondingly, their amount of *no answers* declined from 55 percent in *cloze test 1* to 40 percent in *cloze test 2*. The VBS group, too, recorded fewer unanswered anchor items in the second *cloze* test run than in the first one (33 percent vs. 29 percent in that order).

All in all, the total proportion of score-yielding responses was higher in *cloze test 2*, namely by 3 percent in the VBS class and by as many as 11 percentage points in the traditional class. Even though the above figures definitely hint at an improvement in the students' performance, it is impossible to determine whether this development was based on factors such as heightened task-awareness, increased attention, and possibly even explicit learning of the items under consideration, or whether it occurred simply by coincidence.

5.6. Summary of test results

As illustrated in the previous sections, the VBS group clearly outperforms the traditional group, not only with regard to the total test result, but also in connection with the five individual tests (test types). In the majority of the cases, the differences between the two test groups are statistically significant or even highly significant, the only exception being the *words in context* test, in which the traditional participants are approximately at the same mean score level as their CLIL-trained peers. This is mainly due to the fact that, to some extent, this particular test type also measures productive skills in German, thus allowing for higher scores, even among less-proficient learners of English.

This finding is further substantiated by the correlation existing between the outcome of the entire test sequence and the results on each of the five separate tests:[62]

Table 5:22 *Rank order correlation between scores on individual tests and total score*

N = 33 (32)[63]	SR-TOTAL	WIC-TOTAL	MC-TOTAL	CT1-TOTAL	CT2-TOTAL
Spearman r_s[64]	0,94752674	0,64346591	0,83372326	0,86798128	0,82239736
Interpretation	strong corr.	moderate corr.	strong corr.	strong corr.	strong corr.

As Table 5:22 illustrates, the correlation rate between the students' total test scores and their results on the *self-report* test, the *multiple choice* test, as well as *cloze* test *1* and *2,* respectively, ranges from approximately 0.82 to 0.95, symbolizing a particularly high degree of interrelation (For a detailed account on the interpretation of correlation indices, see Alderson, Clapham & Wall 1995: 78-79 and 184). On the other hand, the quantity defining the association between the total test outcome and the results on the *words in context* test amounts to 0.64, which is only a moderate correlation. Therefore, the *words in context* test is indeed less representative of the

[62] The concept of *correlation* defines "the extent to which two sets of results agree with each other" (Alderson, Clapham & Wall 1995: 77). It is expressed by means of the so-called *correlation coefficient/index* or *rate*, which can be calculated in various manners (Alderson, Clapham & Wall 1995: 80; 287). In the present case, the *(Spearman) rank order correlation* proves most suitable, since "there are only a small number of results to be correlated [and] the results are [listed] in ranks" (Alderson, Clapham & Wall 1995: 80) according to their numerical value.

[63] One traditional student was absent in *cloze test 2*.

[64] The Spearman rank order coefficient r_s is calculated according to the formula

$$r_s = 1 - \frac{6 \cdot \sum d^2}{N \cdot \left(N^2 - 1 \right)},$$

where d^2 is the squared difference between each students' two ranks on the separate tests under consideration, Σ the sum thereof and N the total number of students involved (compare Alderson, Clapham & Wall 1995: 278-279).

students' lexical performance in the English language than the remaining four subtests of the battery.

Furthermore, it has been shown that, on balance, the VBS students with English (or any other foreign language) as their mother tongue do not necessarily score above their German-speaking classmates in the vocabulary tests under consideration. Not surprisingly, the results on the two *cloze* tests, in particular those on *cloze test 1*, present an exception to this general tendency. Besides, the complexity of these two specific language tasks has enabled us, not only to differentiate English natives from non-natives, but also to filter out the highest performing students from the entire test population.

As far as the aspect of gender is concerned, just as in the Swedish context, in both Austrian test groups, the male participants outperform their respective female counterparts in almost all of the areas tested. Again, the only exception is the two *cloze* tests, in which the female representatives of both the VBS group and the traditional group have more *correct* answers than their male peers do. Moreover, it has been discovered that, in both *cloze* tests, the traditional males show a strong preference for the most 'convenient' of all answer options, that is, not giving any answer at all, which serves as a further explanation for their apparent inferiority to the remaining three gender-specific subgroups.

All things considered, it seems as if both the CLIL students and the traditional students involved in the present study have achieved slightly better total, as well as separate, test results compared to the subjects in Sylvén's third test round. Some of the reasons for the superiority of the Austrian participants have been already speculated upon in the previous sections. The following chapter will provide a more detailed analysis of this particular issue, by contrasting the students' outcomes according to a series of extralinguistic background factors.

6. Background factors

Prior to the five vocabulary tests, the students were asked to complete a questionnaire concerning reading, TV and Internet habits, time spent in English-speaking countries, parents' level of education, as well as the students' attitude towards different aspects of English language learning (compare also Sylvén 2004: 109). The outcomes of these question sheets reveal quite interesting tendencies among the two test groups:

6.1. Outcomes according to selected aspects

English TV and movies

While 75 percent of the traditional students noted down that they watched English TV programmes with German subtitles, no more than 52 percent of their VBS peers claimed to share this preference. Conversely, all of the bilingual students, compared to merely 42 percent of the traditional ones, watched English TV programmes without subtitles. With regard to English movies, the proportions are almost similar in the two groups: 29 percent of the CLIL learners and 25 percent of their conventional schoolmates stated that they went to see English films at the cinema on a monthly basis

PC games and the Internet

Likewise, role-plays, strategy games and other forms of (electronic) entertainment, including detailed information and instructions in English, enjoy almost equal popularity with the two groups: 48 percent in the VBS class and 42 percent in the traditional class claimed to spend part of their spare time playing these kinds of computer games. Concerning the Internet, 81 percent of the VBS students and 67 percent of the traditional students said that they surfed daily. However, with regard to English-speaking websites, the proportions of the two groups are more divergent: Whereas 71 percent of the CLIL learners claimed to visit such sites every day, only 17 percent of their control peers did the same.

Browsing through the World Wide Web turns out to be closely linked to yet another one of the students' preferred leisure activities, namely

Writing English texts:

As expected, it was mainly the bilingual subjects (81 percent of the same) who found considerable pleasure in composing letters, e-mails, notes, diary entries, online articles and even more literary pieces of writing, such as poems, stories and song lyrics, in the English language. In the traditional group, too, 42 percent of the students asserted that they wrote in English at least occasionally. Yet, their creative work was simply restricted to functional text types in the form of online game instructions, chat entries and e-mails.

Speaking English for pleasure

Concerning oral language production, the difference between the two groups is even more noticeable: 81 percent of the CLIL learners, however only 33 percent of the traditional learners, claimed to speak English in their leisure time. These figures are not in the least surprising, if we take into account that a considerable number of the CLIL participants in the present study come from bilingual or multilingual backgrounds and are thus quite likely to use the English language when talking to family members, relatives or friends.

Moreover, the specific, multicultural setting of the school allows for English-speaking conversations, even during breaks, spare lessons and other occasions outside of the CLIL or EFL classroom. As stated by the coordinator of the VBS project, this particular feature may also serve as an explanation for the Austrian students' superiority over their Swedish peers (Coord.VBS[65], e-mail: 9 June 2008).

The idea that speaking English for pleasure improves the students' lexical performance finds additional support in the outcomes of the traditional group: While the subjects claiming that they never spoke English in their leisure time attained a total mean score of 143.25 points, their classmates, who regularly talked to English-speaking relatives or friends, scored 22.5 points higher (see Figure 6:1).

[65] Name anonymized

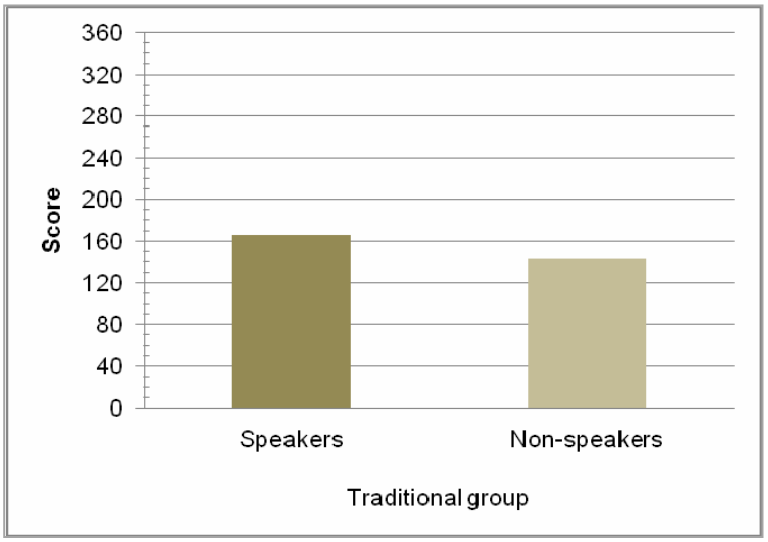


Figure 6:1 *Total mean score, traditional/factor: speaking English for pleasure*

Due to the low number of students, it is impossible to test whether the proportions of the 'speakers', as opposed to those of the 'non-speakers', are normally distributed. [66] An unpaired t-test therefore yields an insignificant result ($t_{pr} \approx 1.38 < 1.81$ = t-value at the level $\alpha = 0.05$). In the VBS group, there were no remarkable differences in the mean values between the 'speakers' and 'non-speakers': Both subgroups attained an average score of approximately 220 points each and were thus clearly ahead of the traditional students.

According to their conversational routines, the representatives of the two test groups also differentiated in their attitude towards

Speaking in front of an audience:

Whereas 86 percent of the students in the CLIL class felt 'excellent' or 'quite good' (the corresponding answer options were "sehr gut" and "gut") about speaking English in the presence of others, the corresponding figure for the traditional class was 33 percent. In connection with the German language, the traditional students were more confident: As many as 83 percent of them asserted that they experienced oral presentations as something extremely or fairly positive. Nonetheless, the

[66] Usually, the normality of a set of data is evaluated by means of a Chi^2-test, which is, however, only applicable if the sample under consideration comprises more than 50 items/subjects (DGQ 1993: A 3.4, 1).

comparative proportion for the VBS class was still higher: Here, 100 percent of the students claimed the same.

Another aspect reflecting the distinctive situation in the two groups is that of

Parents' education:

In the VBS group, 38 percent of the mothers and 67 percent of the fathers have a university or advanced college degree. In the conventional group, the matching proportions amount to 25 percent and 33 percent respectively. Similarly, in the original research context, the CLIL students were more likely to come from more highly educated family backgrounds than their control peers were (Sylvén 2004: 109). Yet, with regard to the total number of university graduates, the Swedish parents clearly outperform(ed) the Austrians in both groups under consideration.

Reading habits (in English)

Among the factors having most influence on the original test subjects was that of voluntary reading of English texts. In her study, Sylvén discovered that throughout the entire time-span of two school years reading had "roughly the same effects on the development of the students' English vocabulary as the CLIL method [did] *per se*" (Sylvén 2004: 117). Since the present work only comprises a single test round, it is impossible to determine the long-term effects of either the one or the other factor, or method in the widest sense. Nevertheless, it can be evaluated whether there are noticeable differences in the average test results between the readers and the non-readers among the Austrian participants.

As for the former category, in the VBS group, 14 percent of the students claimed to read English texts, other than compulsory homework-exercises, every day. 29 percent indicated that they did so on a weekly basis. 24 percent enjoyed the pleasures of Anglophone literature once a month, and another 14 percent occasionally browsed through English newspapers, books or magazines. The remaining 19 percent confessed that they never read anything in English, apart from their school texts. In the traditional group, the number of readers and non-readers was equally distributed: While one half of the students asserted that they read fictional and non-fictional works in English once in a while, the other half never did so in their leisure time.

Figure 6:2 shows the total mean values of the readers compared to the non-readers in both the VBS group and the traditional group:

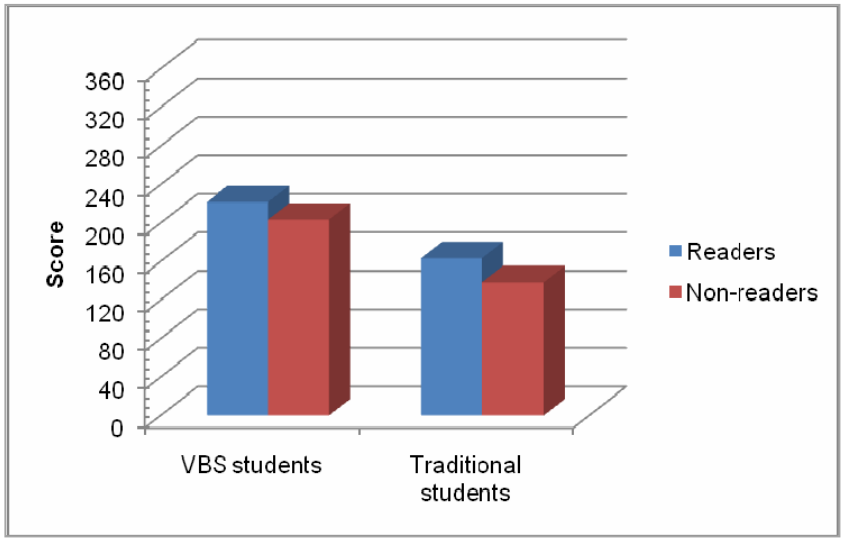


Figure 6:2 *Total mean score, group/factor: reading habits*

As expected, both the VBS readers and the traditional readers outperformed their respective non-reading peers by an average score of approximately 20 points each. A paired F-test shows that, within both test groups, the standard deviations of the subsamples are fairly equal, apart from accidental variations (see Table 6:1):

Table 6:1 *Paired F-test: analysis of variance, total results, group/factor: reading habits*

Readers vs. non-readers	test value F_{pr}	F-value, $F_{f_1,f_2;1-\alpha/2}$		
		α = 0,05	α = 0,01	α = 0,001
VBS n_{R+} = 17, n_{R-} = 4	1,222955705	4,076823063	6,30338459	10,3443301
Traditional n_{R+} = 6, n_{R-} = 6	3,490636282	7,146381829	14,9396055	39,7194468

The same statistical relation also holds true for the average test results of the readers versus non-readers. As Table 6:2 illustrates, neither in the VBS sample nor in the traditional sample, the differences in the total mean scores between these two subgroups prove to be statistically significant:

Table 6:2 *Unpaired t-test, total mean score, group/factor: reading habits*

Readers vs. non-readers	mean difference $\bar{x}_1 - \bar{x}_2$	s_d	test value t_{pr}	t-value, $t_{n_1+n_2-2;\ 1-\alpha}$		
				α = 0,05	α = 0,01	α = 0,001
VBS n_{R+} = 17, n_{R-} = 4	18,088235	23,387195	0,773425	1,729133	2,539483	3,5794001
Traditional n_{R+} = 6, n_{R-} = 6	24,833333	14,870739	1,669946	1,81246	2,763769	4,1437005

What is more, in contrast to Sylvén's (2004: 112-117) findings, the readers in the Austrian control group scored almost as many as 40 points below their non-reading peers in the CLIL group. All in all, as far as these momentary results can be considered reliable, this means that, within the present research context, the CLIL method still seemed to have a greater impact on the students' lexical performance than extracurricular reading habits did.

Unlike voluntary reading, one particular background factor turned out to be almost equally relevant to the outcomes of the Swedish and the Austrian test participants, namely

Time spent in English-speaking countries:

In the Austrian traditional group, all of the students indicated that they had participated in a two-week language course in Malta during the previous school year. In addition, five of the 12 conventional learners had already spent their holidays, however in sum no longer than 3 months, in countries such as England, Ireland, California and New Zealand. Only one student had completed an exchange programme and therefore stayed in an English-speaking country for one year.

In the VBS group, all except one student had spent some time in English-speaking countries. Excluding the subjects with English as their native language, two students had stayed abroad for more than one year, four between 3 and 12 months, and the remaining 11 participants had been in some corner of the English-speaking world for less than 3 months. In the following, a distinction will be made between the categories

- *+ T. in ESC*, including the students who had spent between 3 and 12, or more than 12 months in English-speaking countries, and

- – *T. in ESC*, comprising those subjects who had spent 0-3 months in English-speaking countries.

In either case, the students with English as their L1 are excluded from the evaluation.

Figure 6:3 presents the mean score results on the five individual vocabulary tests, as well as the total mean score for each of the above-mentioned sub-divisions in both the VBS group and the traditional group:

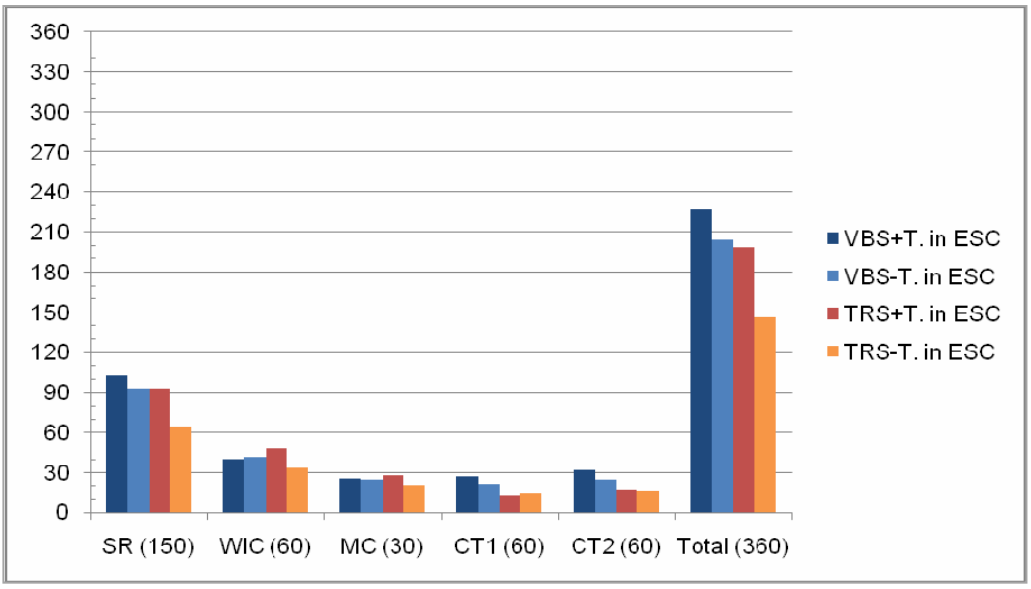


Figure 6:3 *Mean score results on individual tests and total test sequence, group\EL1/factor: time spent in English-speaking countries*

As the final array of columns shows, the highest total mean score was attained by the VBS students who had spent a longer period in English-speaking countries (*VBS+ T. in ESC*). Furthermore, the members of this specific subgroup also performed best on the *self-report* test (*SR*), as well as on both *cloze test 1* (*CT1*) and *cloze test 2* (*CT2*), indicating their high level of lexical competence. Concerning the *multiple choice* test (*MC*), which is yet another important marker of the participants' English vocabulary knowledge,[67] it seems as if the *VBS+ T. in ESC*-students were outperformed by their traditional peers with comparable international experience.

[67] As mentioned previously, these four particular tests are highly representative of the students' overall lexical performance, as their correlation with the total test result turns out to be exceptionally strong (see Table 5:22).

However, in this respect, the above figures are slightly misleading, as the superiority of the *TRS+T. in ESC*-subgroup basically results from the fact that the whole sample consists of only **one** student, whose individual test score accidentally happened to exceed the mean value of the VBS subgroup. Still, this particular student was also approximately at the same **absolute** score level as the majority of her VBS colleagues in both subsets. Only the traditional learners who had been abroad for less than 3 months obtained noticeably lower *multiple choice* test values.

Although it is difficult to judge the overall impact of time spent in English-speaking countries simply from the effects that this particular background factor has on a single test participant, the figures in the above bar chart are quite convincing. Not only did the *TRS+T. in ESC*-student score exceptionally highly on the *multiple choice* test, but her result on both the *self-report* test and the test sequence as a whole was (almost) on a par with the corresponding average outcomes of her *VBS- T. in ESC*-peers. In the *words in context* test, the traditional learner was even in the lead, this time, not only in comparison with the **mean** results of the other subgroups, but also as contrasted to (most of) her colleagues' **absolute** scores. Yet, in this particular connection, it is rather problematical to attribute the student's success only to her previous international experience, as it has been shown that the *words in context* test is less representative of the learners' lexical competence in English than any of the other tests under consideration is (for more details, compare section 5.2.2. and Table 5:22).

Concerning *cloze test 1* and *2*, both VBS subgroups were well ahead of their respective counterpart in the traditional group. Besides, the VBS students who had spent a longer period in English-speaking countries clearly outperformed their *VBS-T. in ESC*-classmates. Not surprisingly, the lowest test results were obtained by the conventional students who had stayed less than 3 months in countries with English as the national language.

In conclusion, it can be said that, even if with regard to some of the vocabulary tests, the traditional student who had spent one year in an English-speaking country scored approximately on a par with the VBS subjects who lacked such an experience, this was not true for the two *cloze* tests. Thus, it seems that, in connection with fulfilling these most complex lexical tasks, CLIL instruction may provide a more adequate

preparation than previous language input in English-speaking countries does (compare also Sylvén 2004: 226).

6.2. Summary

As illustrated on the preceding pages, the factors having most influence on the outcomes of the Austrian participants are speaking English for pleasure, extracurricular reading habits, and the amount of time the students have spent in countries with English as the medium of communication. While the former background factor proves to be of particular relevance only for the conventional students, the latter two also have a decisive impact on the results of the CLIL learners. In both test groups, on average, the participants who regularly read English texts on a voluntary basis attained higher overall test scores than their respective non-reading peers did. Yet, contrary to the findings in the original study, in the present research context, the differences between readers and non-readers seem to be less significant than the differences arising from the two teaching approaches per se. As a consequence, irrespective of their voluntary reading habits, the subjects taught in the traditional manner are inferior in lexical achievement and proficiency to the participants who receive constant English language input by means of CLIL instruction.

Concerning the factor of time spent in English-speaking countries, the situation is slightly different: As mentioned above, the traditional student who had spent one school year in an English-speaking country not only outperformed her classmates, but also managed to obtain approximately the same total test score as her *VBS- T. in ESC*-peers did on average.

As a general tendency in the evaluation of the students' questionnaires, it has been discovered that 'common' leisure activities, such as playing computer games and surfing the Internet, as well as the more expensive pastime-option of going to English movies, are almost equally relevant to the subjects in both test groups. However, as soon as the pursuit in question requires more complex aspects of English language usage (as in reading and writing English texts, speaking to natives, watching TV programmes without subtitles, and browsing through English websites), the VBS learners are clearly ahead of their traditional peers. This finding also

corresponds to Sylvén's description of the CLIL students' particular background situation:

> They seem to have established better and more active channels through which they receive English input in various forms [...]. [...] In turn, of course, this means that [the CLIL student] is likely to perform better on vocabulary tests than the student who lacks such well-established channels for input [...], as is often the case for the control student. (Sylvén 2004: 225)

In other words, those participants who already receive an increased amount of language input within the school setting are also more likely to use the English language for various purposes in their leisure time, which explains the huge differences in the test results between the VBS students and the traditional students in the present study.

Regarding the variations in the spare-time habits between the Austrian students and their Swedish colleagues, one can only speak of basic trends or tendencies, since, due to the different number of participants, the actual percentage figures in the two studies are not comparable. As intimated in the preceding discussion, extensive reading of English texts, other than compulsory school texts, proved to be extremely beneficial in the case of the Swedish control students (compare Sylvén 2004: 111-118). Besides, these particular subjects more often watched English movies than their Austrian counterparts in both the VBS group and the traditional group did (Sylvén 2004: 109). Time spent in English-speaking countries seemed to have an important influence on the results of all students, although within the Swedish context, its significance dramatically declined during the course of the study (Sylvén 2004: 110).

Given the fact that the present study was conducted almost 8 years after the distribution of the first original student questionnaires (see Sylvén 2004: 46-47), it is not at all surprising that the majority of the Austrian participants claimed to surf the Internet daily, while most of their Swedish peers indicated that they did so only once a week (Sylvén 2004: 109). On the other hand, English TV programmes, with and without subtitles, seemed to be more commonly used in the Swedish context, though in this connection, the information given by the Viennese participants is not entirely clear-cut (compare the results in 6.1. and Sylvén 2004: 109). As for the students' educational backgrounds, it has already been hinted that university graduates were more likely to be found among the parents of the Swedes than among those of the Austrians (compare Sylvén 2004: 109).

If we finally compare Sylvén's statement,

> a traditional student who receives a great deal of English input outside of school may score above a CLIL student who mainly gets English input in the classroom (Sylvén 2004: 225)

to the related findings in the present study, it becomes even more evident that extracurricular language input, of almost every kind, is not only of greater importance in the original research context, but also more easily accessible by the Swedish students than it is by the Austrians. Therefore, we may conclude, albeit with reservations, that the superior lexical achievement of both Viennese groups over their respective Swedish counterpart predominantly results from the higher-quality teaching in the corresponding CLIL or EFL classroom(s).

7. Further aspects

This chapter provides further insights into the learning and teaching situation in the two test groups under study by focusing on those aspects covered in the questionnaires that were not directly associated with the participants' performance on the five lexical tests, but seemed nonetheless relevant to the students' language acquisition in general. First, the emphasis will be placed on the learners' motivation, attitude and general satisfaction with their choice of school programme. Next, we will look at how the VBS group and the traditional group assessed their respective linguistic development in English in the course of their secondary education. The final section presents the outcomes of the teacher questionnaires and offers a detailed discussion on the teachers' perspective and roles within the VBS programme at the school under examination.

7.1. Motivation, attitude and general satisfaction

It has often been claimed that, as a "[n]atural [w]ay[...] of picking up [l]anguages" (*CLIL Compendium* 2001), *CLIL* helps to increase the students' motivation and create a more positive attitude towards learning (compare, e.g. *CLIL Compendium* 2001; Mewald 2004; Dalton-Puffer 2005; Darn 2006; Abendroth-Timmer 2007). As research has shown, this, in turn, also improves their foreign language proficiency (cf., e.g. Gardner & Lambert 1972; Gardner 1985). Thus, the relation between the learners' success and their motivation due to a stimulating environment seems to be mutual or, as Lightbown & Spada (1999: 56) point out,

> [W]e do not know whether it is the motivation that produces successful learning or successful learning that enhances motivation or whether both are affected by other factors.

What is, however, unquestionable, is that students who are required to use a particular foreign language in a variety of different situations and subjects will inevitably recognize "the communicative value of the [...] language and will therefore be motivated to acquire proficiency in it" (Lightbown & Spada 1999: 56). This principle is extensively exploited in the context of *Content and Language Integrated Learning*.

Even though the present study does not focus on these motivational aspects in detail, they were briefly touched upon in the questionnaires (see Appendix 2). For instance, in one of the first questions, the students had to explain their motives for choosing the VBS programme or, instead, the traditional programme. An additional question concerned their satisfaction with the respective choice. In this connection, the students had to evaluate their programme according to the five-point scale of the Austrian grading system. Moreover, they were asked about advantages and disadvantages of bilingual schooling, as opposed to traditional secondary education, and whether they would recommend their particular choice of school programme to future students.

Just as in the Swedish context, the most common reason for selecting the CLIL/VBS branch was the "linguistic focus [...] of the program" (Sylvén 2004: 211). As many as 38 percent of the VBS students indicated that they had chosen this particular school type, as they were especially interested in the English or, in the case of some of the 'international' students, also the German language and therefore wanted to acquire a high level of competence in it. 19 percent had seen the VBS programme at this particular school as a natural continuation of their bilingual lower secondary and/or primary education. These motives also correspond closely to the school's definition of the target group(s) for the VBS classes, that is,

> German speaking children who have a previous knowledge of English or a particular interest in learning English; children with English-speaking parents, who speak English as their mother tongue or as a language of communication and who have a previous knowledge of German [and] children who attended a bilingual primary school ("VBS Middle Schools" 2005: 2)

... or, with regard to the upper secondary, a bilingual middle school or any other language-focused lower secondary programme ("Bilinguale Schule": 22 August 2008; see also Chapter 3).

In addition, 10 percent of the CLIL students claimed that they had chosen the bilingual programme in order to prepare for higher studies abroad. 5 percent stressed the importance of advanced English language skills for their future working careers. In the original study, the proportion of CLIL learners citing future education and better job perspectives as the main motive for their choice of school programme was comparatively low (25 percent, see Sylvén 2004: 211). As Sylvén (2004: 211) notes,

these findings stand in an interesting contrast to what the Swedish, and likewise the Austrian, school authorities have established as the primary goals and reasons for CLIL instruction, namely "internationalization" (Nixon 2000: 41) and "the educational knowledge necessary to compete successfully in the international workplace" ("VBS Middle Schools" 2005: 1). Thus, in Gardner's & Lambert's (1972) terms, we may say that while the schools have an exclusively *instrumental motivation* for offering CLIL classes,[68] the students themselves seem to be more likely to attend those "for personal growth and cultural enrichment" (Lightbown & Spada 1999: 56), meaning that **their** motivation is predominantly *integrative*.

Evidently, this is not true for all of the participants in the present study. For instance, 14 percent of the VBS students indicated that their choice was mainly based on recommendations by their brothers and sisters or friends of the family. One student confessed that his parents had forced him to attend a bilingual class instead of a traditional one. Furthermore, 10 percent of the CLIL learners did not mention any specific motives, but noted that they had selected the programme 'by chance' or, as yet another one of their classmates claimed, 'out of youthful abandon'.

A possible explanation for this rather negative comment can be found in the traditional students' answers to the question "Hast du überlegt, den VBS-Zweig zu besuchen? – Ja, aber ich habe mich anders entscheiden, weil .../Nein, weil ...". Here, the most frequently cited counterargument concerned the high demands and excessive workload in the VBS classes. Moreover, two of the informants explained that they had decided against the bilingual programme, since they considered the CLIL learners 'snobbish' and 'arrogant', an opinion that had also been voiced by one of the Swedish subjects (see Sylvén 2004: 205). Another student criticized the VBS classes for being too heavily focused on English, while she wanted to learn also other foreign languages. One of her colleagues admitted that bilingual education would have been simply too difficult for her. The remaining students indicated that they had never been particularly interested in or not even thought about this specific option.

As regards the first complaint, long school days, difficult assignments and an excessive amount of work were also seen as the main drawback of the VBS programme by as many as 57 percent of the VBS learners themselves. Besides, three

[68] including the benefits that such programmes have for their overall reputation

of the CLIL students found specific subjects, such as Computer Science, "the worst thing they [could] think of" (Sylvén 2004: 204). One of their classmates complained about particular teachers, another about bad teaching methods, and a third one about boring contents. Additional criticism was directed at the realisation of the CLIL method itself. One student thought it was a major disadvantage that the VBS programme was only offered for particular classes, as he would prefer the whole school to be bilingual. Another one felt that there was too much teaching in English and too little in German, and one of his colleagues added that the teaching was mainly focused on British English, while he wanted to learn the American lexis and pronunciation.

Interestingly, exactly the same aspect was interpreted as an advantage by yet another VBS student. This ambiguity also applies to the factor of heavy workload and demands, as one of the CLIL learners pointed out:

> Die Schule erfordert viel, aber ich bin der Meinung dass das sehr gut ist weil man daraus mehr lernt. [QR.VBS.21f]

If we look at the question as to "what is the best thing" (Sylvén 2004: 204) about each respective programme in more detail, it becomes noticeable that, just as in the Swedish context (compare Sylvén 2004: 211), the answers of the groups were quite similar. More than half of the students in both the VBS class and the traditional class (once again) praised their programme's special emphasis on languages, meaning English in the first case and French with regard to the latter. Furthermore, one third of the traditional learners appreciated that they had fewer Mathematics and Science lessons. On the other hand, one of their VBS peers thought that the best thing about **his** programme was that he did not have any Latin, a subject which was, by the way, also strongly disliked by as many as eight out of the twelve students in the regular class.

Other factors that were mentioned in both groups concerned the positive classroom atmosphere as well as the sound general education provided by the respective school type in question. Besides, two of the CLIL learners explicitly cited the VBS characteristic of "a culturally diverse student population" (VBS Middle Schools" 2005: 1) as one of the greatest benefits of the programme, and another one

particularly praised the CLIL focus on fluency and extensive reading (Darn 2006) instead of "pressure for [grammatical] correctness" (Dalton-Puffer 2005: 174).

When asked about how they would rate the quality of their choice of programme on a five-point scale, 47 percent of the VBS and 33 percent of the traditional participants chose either of the Austrian marks '1' (= 'excellent, very good') or '2' (= 'good'). 38 percent of the CLIL and 50 percent of the control students considered their programme 'satisfactory' (= '3'), and three out of the 21 bilingual learners, as well as two of their conventional peers, assigned either a '4' or a '5'. Overall, just as in Sylvén's (2004: 205) study, the average 'satisfaction marks' for the two groups were approximately equal: 2.62 for the VBS group and 2.83 for the traditional group. Nonetheless, it was rather the CLIL learners (76 percent of them) than their traditonal colleagues (58 percent) who indicated that they would recommend their choice also to other students.

7.2. Self-assessment

Closely connected to the participants' evaluation of their school programmes was also the question on how they estimated their own progress in English in the course of their secondary education (for more details, see section 3.3.2. on the student questionnaires). As the distribution of the present *self-report* answers has shown (see section 5.1.4.), the CLIL learners turned out to be considerably more self-confident with regard to their knowledge of the items under examination than their traditional peers were. Moreover, the two groups also differed in their respective attitude towards speaking in front of an audience: All of the VBS students, compared to 83 percent of the traditional students, indicated that they felt 'excellent' or 'quite good' about speaking German in the presence of others. Concerning English, the corresponding figures were 86 percent for the VBS group and 33 percent for the traditional group (see also Chapter 6).

Table 7:1 below will reveal whether these differences also applied in connection with the learners' self-assessment of their development in six specific areas of English language competence. In order to facilitate the comparison with the related outcomes in the original study, the (mean) values for the students' ratings were transformed into Sylvén's five-point system "where 1 is *very bad* and 5 is *very good*" (Sylvén

2004: 211). Besides, the figures given by the four English natives were excluded from the present evaluation, since, not surprisingly, all of them indicated that their English language skills had deteriorated during their education in Austria. One student explicitly noted,

> In den meisten Bereichen ist mein Englisch schlechter geworden im Vergleich zu England – ich lerne gleichzeitig Deutsch. [QR.VBS.13f]

In these particular cases, it would probably have made more sense to ask about the learners' improvement. in German. This is, however, beyond the scope of the present study.

Table 7:1

Students' self-assessment of their progress in six areas of English language proficiency

Area	*Seregély 2008*		*Sylvén 2004 TR III**	
	CLIL (VBS)	**Traditional**	**CLIL**	**Traditional**
Speaking	4,3	3,8	3,8	2,9
Writing	4,4	4,0	3,8	2,9
Listening comprehension[69]	4,8	4,1	4,2	3,3
Reading comprehension[70]	4,6	4,0	4,1	3,3
Vocabulary	3,9	3,8	3,5	2,9
Grammar	3,9	3,8	3,4	2,6
Total mean	**4,3**	**3,9**	**3,8**	**2,9**

*(compare Sylvén 2004: 184)

As the above figures show, the Austrian CLIL students not only rated themselves considerably higher in all six areas than their traditional peers, but they also proved to be noticeably more self-assured than their Swedish colleagues were. In fact, the CLIL learners in the original study were approximately at the same level with regard to most of the aspects as the Viennese control students. The Swedish control students, on the other hand, rated themselves significantly lower than the remaining three subgroups.

The superior self-assessment of the two CLIL groups (compared to their respective traditional counterpart) exactly corresponds to what has been expected, given the "motivating, low-anxiety communicative atmosphere" (Dalton-Puffer 2005: 174)

[69] ~ "Understand spoken English" (Sylvén 2004: 70; 184)
[70] ~ "Understand written English" (Sylvén 2004: 70; 184)

that is said to be highly characteristic of CLIL classrooms. As regards the mean values for the Austrian and the Swedish context, it is quite likely that the variation predominantly results from the students' different interpretations of what constitutes their 'development during (upper) secondary education'. While in Sylvén's study, the corresponding question exclusively refers to the "first two years in upper secondary school" (Sylvén 2004: 70), in the present questionnaire, the focus is on the learners' progress in the course of their secondary education as a whole (compare section 3.3.2.). Needless to say, the longer the learning period is, the more opportunities there are for improvement. The figures mentioned in this specific respect therefore have to be taken with some caution. Still, the dissimilarity between the formulations in the Austrian and the Swedish question sheets is not particularly problematical, since our main interest in connection with self-assessment is on the distinction between CLIL learners and traditional learners in general, and not necessarily the comparison of the two research contexts.

Interestingly, though, in all of the four subgroups, the ranking of the individual aspects was basically the same (apart from those instances where two or more areas yielded the same score or mark). The greatest improvement was recorded in the category *listening comprehension* or "Understanding spoken English" (Sylvén 2004: 184), which gained an overall mean 'score' of 4.1,[71] followed by *reading comprehension*, which was rated with the grades 4.6, 4.1, 4.0 and 3.3 by the Austrian and the Swedish CLIL group, and the corresponding traditional groups in that order. *Writing* in English ranked third (with an overall mean score of 3.8), *speaking* fourth (3.7) and *vocabulary* development occupied the fifth place (3.5). With a total average mark of 3.4, English *grammar* turned out to be the area with regard to which all of the participants felt that they had achieved the least progress.

As for the outcomes of the traditional students, further research would be required in order to interpret their implications. However, concerning the CLIL learners, the above-mentioned ranking clearly reflects also the 'priorities' of the method itself, that is, "more language exposure and input" (Sylvén 2004: 4), as well as the basic rule that

> Fluency is more important than accuracy and errors are a natural part of language learning. Learners develop fluency in English by using English to

[71] i.e. the mean value of all four subgroup-means

> communicate for a variety of purposes [and r]eading is the essential skill. (Darn 2006)

The following section will show whether the CLIL students' estimation of their own linguistic development also corresponds to their teachers' views on this particular issue.

7.3. The role of the VBS teachers

In order to gain some information, not only about the students' attitudes concerning the VBS programme and its effects on their lexical skills, but also about the teaching perspective of *CLIL*, prior to the test sequence, questionnaires were distributed among all of the teachers in the present VBS class and the native speaker teachers at this particular school. Teachers involved in the traditional class were not asked to fill in the questionnaire.[72]

Just as in the original context, the return rate for the question sheets was relatively low (55 percent compared to 50 percent in the Swedish study, see Sylvén 2004: 199). As Sylvén notes,

> this leaves the impression that either the self-esteem among the CLIL teachers was at issue or there was a general lack of interest in taking part in the investigation. (Sylvén 2004: 199)

Neither of the two explanations seemed to apply in the case of the Viennese teachers, since the majority of them were quite eager to inform me about specific features of the VBS programme, and also wanted to learn more about the present study. The only possible reason why some of the teachers were still a bit careless about submitting their questionnaires was that they were too busy with their own projects, preparations for lessons, corrections and organisational matters, so that they simply forgot "to fulfill this unprioritized task" (Sylvén 2004: 1999).

Overall, 11 out of the 20 teachers who had received a questionnaire eventually handed in their sheets by the end of the deadline (beginning of May 2007). Seven of those were females, four males. Three of the teachers were native speakers of

[72] In the following, the term 'teachers' only covers the participants in the present study. Whenever a clear-cut distinction is made between the two subgroups, the terms 'Austrian/subject teacher' and 'native speaker/ English-speaking teacher' are used. Depending on the context, the expression 'VBS/CLIL teacher(s)' may either refer to the Austrian teachers or to the test population as a whole.

English,[73] while the remaining eight all had German as their mother tongue. Given the low number of actual 'participants', the outcomes of the questionnaires are not to be seen as representative of the whole population of VBS teachers, but only reflect the situation and opinions of certain individuals. For more transparency, in the following passages, the individual issues raised in the teacher questionnaires will be underlined in the text.

When asked about their general teaching experience, two of the VBS teachers as well as two of their English-speaking colleagues indicated that they had been working in their profession between five and ten years, three of the teachers and the third English native had been practising their job for 10-20 years, and the remaining three VBS teachers had a work experience of more than 20 years. With regard to the VBS programme, on average, the Austrian subjects had a teaching experience of 7.5 years, compared to approximately 6 years for their English-speaking colleagues.

Concerning formal education, only one out of the seven VBS teachers who filled in the questionnaire had a university degree in the English language. The remaining seven had completed their formal education in English with their graduation from upper secondary school, meaning that their language skills were approximately at the same level (or with regard to some of the CLIL students, possibly even below) as their students'. This outcome is well in line with Sylvén's (2004: 199) findings and also reflects what has been a much-debated issue in connection with CLIL teaching, namely the insufficient linguistic competence of some subject teachers (cf., e.g., Hall 1996; Nixon 2000; Mewald 2004; Sylvén 2004; Dalton-Puffer 2005; Caspari, Werner & Zydatiß 2007).[74] For a comprehensive account of the Austrian situation, see for instance Mewald (2004: 218 ff; 306 ff and 528-542). The present study will not discuss this aspect in more detail.

However, it should be pointed out that, unlike some of their Swedish peers (compare Sylvén 2004: 200), **none** of the Viennese students complained about their teachers' English proficiency. Even though this may also be merely coincidental, it seems quite likely that at this particular school, the language level of the VBS teachers who do not hold any university degree in English is nevertheless extraordinarily high. One

[73] In total, at the time of the test administration, the school employed four English natives.
[74] In this specific context, the term 'subject teachers' refers to those (non-native English) CLIL teachers who do not have any kind of university education in English.

reason for this assumption is that when the teachers were asked to evaluate their own competence in English (and other foreign languages, see questionnaire, Appendix 2) according to the four-part scale *beginner – intermediate – advanced – native speaker level*, all of them selected the third option. On the contrary, the same question in another Austrian study (Mewald 2004: 222-223), involving four different schools, yielded a considerable number of answers with regard to the first two levels. What is more, deficient English language skills, which were considered a major problem by almost all of the teachers in an earlier Swedish report (Hall 1996), were not mentioned at all in the open-ended question section at the end of the present teacher questionnaires.

Instead, more than half of the subject teachers in the present survey indicated that they had spent a longer period (at least 6 months) working or studying in English-speaking countries. One of the participants had even been to Great Britain and the United States for as many as 3 years in sum. Besides, all of the VBS teachers claimed to (have) receive(d) some additional training in English, either in the form of 'official' in-service courses and seminars for bilingual teaching, or organized and financed by themselves, as in the case of private tuition and further language classes abroad.

As regards the actual amount of CLIL used by each teacher, the figures varied between two and 20 lessons per week, of which an average of five were taught in teams together with the native speaker teachers. Concerning the benefits of team-teaching, a variety of factors were mentioned, such as:

→ enriched learning; varied ideas; content check [QR.NST.1m]

→ anregendere, interessantere Vorbereitung, neue Blickwinkel auf Thema, andere Art der Arbeit in der Klasse, […] mehr Zeit für einzelne Schüler … [QR.TVBS.2f]

→ Unterstützung; englischer Input [QR.TVBS.3m]

→ […] andere Kultur [QR.TVBS.1f]

→ language + subject matter support, split of work load [QR.TVBS.5f]

→ broadens horizon [QR.TVBS.7m

→ learn from each other, [...] kids get more attention [QR.TVBS.3f]

On the other hand, the only disadvantage of team-teaching turned out to be the

→ lack of time for preparation together – this is not built into the timetable [QR.NST.2m]

As for teaching methods and strategies, the native speaker teachers listed:

→ Cooperative learning, active learning, also old-school lecturing, group activities (e.g. create magazine, TV show, presentation for kids, films) [QR.NST.2m]

and

→ [...] discussions [QR.NS.3f]

However, the most essential feature seemed to be "a lot of interaction". [QR.NST.1m]

In order to obtain a clearer picture of the teaching situation in the VBS class under study, the teachers were asked to rate a set of statements about bilingual instruction (BI) according to their truth value. The corresponding answer options were: 'true', 'mainly true', 'partly true' and 'not true'. As for the statement "Bilingualer Unterricht bereichert den methodisch-didaktischen Bereich des Fachunterrichts.", 86 percent of the VBS teachers and 100 percent of the native speaker teachers opted for either of the first two categories. As expected, the sentence "There is sufficient in-service training for bilingual instruction." yielded predominantly negative answers (cf, e.g. Hall 1996 for similar findings in a Swedish research context): All of the native speakers as well as 71 percent of the Austrian CLIL teachers chose either 'partly true' or 'not true'. Still, two of the VBS teachers were quite content about their training situation and therefore selected the categories 'mainly true' and 'true' respectively.

With regard to the availability of teaching materials ("There are sufficient teaching materials for BI."), the participants' answers were almost equally distributed among the two 'intermediate' categories. All except two of the VBS teachers, who went for the option 'not true', chose either category 2 ('mainly true') or 3 ('partly true'). Based on the information gained from previous studies and personal conversations with some of the VBS teachers, it seems that this somewhat contradictory result has to be interpreted in the following way: While in general, it is true that published resources for bilingual teaching are rare, the same does not necessarily apply to

(topic-related) materials produced by the teachers themselves. For instance, Mewald points out,

> If teachers did not find appropriate materials they produced the materials themselves, frequently using internet resources. Material production was preferably achieved in teamwork or done by the NS [native speakers] (Mewald 2004: 324).

This was also a quite common practice in the school under consideration. On my first visit, the VBS coordinator actually explained that, due to a lack of appropriate printed resources, the teachers had been compiling a collection of self-made materials ever since the introduction of the bilingual programme. This assortment of worksheets, handouts, tests, learning games, as well as audio and video tapes was shared between the members of a specific subject group and continuously expanded and renewed. Needless to say, this particular form of material design requires an increased workload, more intensive planning and also a considerable amount of money on the part of the teachers' themselves (compare Mewald 2004: 226; 327). Thus, it is not in the least surprising that, when the participants were asked as to <u>which regulations could improve bilingual instruction/the VBS project in the future</u>, a lot of them mentioned aspects, such as:

→ mehr Ressourcen – vor allem finanziell! [QR.TVBS.3m]

→ mehr Native-Speaker, finanz. Unterstützung => Kauf von Literatur + Unterrichtsmaterialien [QR.TVBS.4f]

→ zusätzliche Werteinheiten für Planung, [...] bilinguale Materialien auf österr. Lehrplan abgestimmt [QR.TVBS.5f]

In this connection, another urgent request concerned the payment for the native speaker teachers. According to the official legislation, both the Austrian subject teachers and their English-speaking colleagues have equal rights to teach,

> wobei jede Lehrergruppe ihre dienstrechtlichen und besoldungsrechtlichen Bestimmungen beibehält. Englischsprachige Lehrer werden besoldungsmäßig AHS-Lehrern gleichgestellt, wenn sie die entsprechende Qualifikation nachweisen können ("Bilinguale Schule": 22 August 2008).

Unfortunately, very often this is not the case, since hitherto there do not seem to be any clear-cut, general guidelines as to what constitutes this specific professional 'qualification' (see also Mewald 2004: 306 ff.). Even though the natives involved in the present study have undoubtedly more to offer than their mere "nativeness"

(Norris 2001: 15), they do definitely not hold a university/teaching degree in every single subject they teach. For instance, one of the participants explicitly stated that he was giving lessons in Biology, Art, Film, Physics, Computer Science and Maths, while officially he 'only' had a university degree in General Pedagogy, and therefore earned considerably less than his Austrian colleagues with a comparable amount of teaching experience. That the problem of low salaries for a group of teachers who are virtually "indispensable to the project" (Mewald 2004: 239) is, indeed, quite serious, becomes also noticeable from the following statement by the school's VBS coordinator:

> [D]er ganze Schulzweig steht und fällt mit den Native Speaker Teachers […]. Diese Fachkräfte wären schon vorhanden, wenn nicht ihre Bezahlung ausgesprochen schlecht wäre. Immer wieder passierte es, dass sehr fähige, anfangs hoch motivierte englischsprachige Lehrer das Handtuch warfen und wegen der schlechten Bezahlung die Schule verließen oder bedauernd eine bereits gegebene Zusage zurückzogen. In der Praxis bedeutete das, so schnell wie möglich eine Ersatzkraft zu finden, die gleichwertig war - eine höchst schwierige Aufgabe angesichts des Hungerlohnes! Noch heute stehen wir LehrerInnen im VBS Zweig immer wieder vor der drohenden Tatsache, dass ein Native Speaker Teacher „unerwartet" kündigt. Zahlreiche Interventionen von Seiten der Direktion, eine Lösung für dieses Problem zu finden, sind bisher gescheitert. (Coord.VBS, unpublished article: 3) [75]

Evidently, it is beyond the scope of the present study to analyse this problem in more detail. To conclude the above discussion, for the present informants, there is only one possible solution to the situation of the native teachers, namely:

> Adequate pay [...] - or the project will ultimately fail!
>
> [QR.NST.2m]

Apart from the debate on the linguistic and professional requirements for subject teachers and natives in that order, another issue that has often been raised concerns the risk of deficiencies in the subject-content knowledge that the CLIL students acquire, compared to their traditional peers (see Sylvén 2004: 228). While a lot of studies conducted in this field (cf., e.g. Cummins & Swain 1986 for a review of bilingual programmes) have actually shown that the CLIL method has no negative effects on the learners' subject-content achievement, individual researchers (e.g. Washburn 1997) also provided counterevidence to this claim. In her study of content-

[75] Quote anonymized

based English language instruction at Lower Austrian middle schools, Mewald (2004: 530-531), for instance, discovered that

> [t]he concentration on the basics in the subject matter gave rise to worries in anticipation of the pupils' transition to higher schools. The fact that reduction in quantity in combination with the quality of tuition – including *emphasis on revision and the links between the subjects – would support the learners' understanding*, could not free the teachers completely from their bad conscience of possibly not having delivered enough content.

Similarly, one of the teachers involved in Sylvén's study asserted that the contents of her/his social science lessons suffered due to the fact that s/he had to teach them in English (Sylvén 2004: 229). In the present context, however, this view did not find any support. With regard to the statement "The students seem to miss out on subject knowledge due to bilingual instruction (BI).", only two out of the eight VBS teachers selected the option 'mainly true'. The remaining six, as well as all of the native speaker teachers, rejected this claim and chose either of the categories 'not true' or 'partly true'. Conversely, the statement "The students seem to improve their subject knowledge through BI." yielded four answers (one by a VBS teacher and the remaining three by the native speakers) categorized as 'true', additional four with the label 'mainly true', and three answers classified as 'partly true'. On balance, then, it seems that the Austrian CLIL teachers, unlike their Swedish colleague, were not particularly worried that bilingual tuition might have any harmful effects on their students' skills in Biology, Physics, History, Maths or any other subject.

What is more, concerning the impact of the CLIL method on the learners' English language competence, all except one VBS teacher (who went for the option 'mainly true') rated the statement "Die SchülerInnen verbessern durch bilingualen Unterricht ihre Fremdsprachenkompetenz deutlich." as 'true'. Likewise, each of the three natives was fully convinced that the "[s]tudents ha[d] improved their English clearly since [they had] been working with them" (for examples of the opposing view, see the discussion in Sylvén 2004: 232).

Table 7:2 shows the mean values for the teachers' assessment of the CLIL learners' development in the areas of speaking, writing, understanding spoken (= *listening comprehension*) and written (= *reading comprehension*) English, as well as English vocabulary and grammar. Again, the figures are given according to Sylvén's five-point scale with "1 is *very bad* and 5 *very good*" (Sylvén 2004: 201). In both the

Swedish and the Austrian sample, the ratings reflect the responses of the entire population of participants. For the individual results of the four Swedish schools, see Sylvén (2004: 201). As for the present context, the 'marks' awarded by the Austrian subject teachers and the English natives were basically the same.

Table 7:2 *Teachers' assessment of CLIL students' progress in six areas of English language proficiency*

Area	***Seregély 2008*** *(All)*[76]	***Sylvén 2004**** *(All*[77]*)*
Speaking	4,9	3,7
Writing	4,0	3,4
Listening comprehension	4,8	4,5
Reading comprehension	4,8	4,2
Vocabulary	4,4	3,9
Grammar	3,8	3,1
Total mean	**4,4**	**3,7**

*(compare Sylvén 2004: 201)

Just as the Austrian CLIL students themselves (compare Table 7:1 in the previous section), their teachers assigned considerably higher 'grades' in all six areas than their Swedish colleagues had done. (The corresponding total mean values are 4.3 and 4.4 compared to 3.8 and 3.7 for the Austrian students and teachers and their respective Swedish counterparts in that order.) As mentioned earlier, this variation can mainly be attributed to the fact that, while in the original version of the two questionnaires, the emphasis was placed on the students' linguistic development during the first two years of upper secondary school, in the present replication, the focus was on the learners' bilingual secondary education as a whole.

If we compare the figures in Table 7:1 and 7:2, it becomes obvious that, similar to the students, both teacher groups rated the receptive skills *listening* and *reading comprehension* particularly high: 4.8 and 4.5 for the understanding of spoken English, and 4.8 and 4.2 for the understanding of written English in the Austrian and the Swedish (teacher) sample respectively. On the contrary, *grammar* yielded average 'marks' of 3.8, in the present, and 3.1, in the original research context, and

[76] Total sample of Austrian teachers and native speaker teachers
[77] CLIL teachers of all four schools

was therefore perceived as the area where the learners had improved the least. As Sylvén (2004: 201) points out,

> These results seem reasonable since the CLIL student needs first of all to be able to understand spoken and written English. The fact that grammar comes last is probably a result of it not being focused on in the CLIL classroom but rather left for the English language class.

In connection with *grammar*, the teachers' assessment was exactly in line with how the students themselves estimated their own development. On the other hand, the progress in the areas of *writing* and English *vocabulary* was rated differently by the two 'opposite members' of the school community. While both the Austrian and the Swedish students felt that *CLIL* had been more beneficial to their written language production, their respective teachers prioritized the method's positive effects on the acquisition of vocabulary (see Tables 7:1 and 7:2 and the discussion in section 7.1.).

Concerning *speaking*, the Viennese and the Swedish CLIL students, as well as the teachers of the latter, all awarded rank 4 (the corresponding grades were 4.3, 3.8 and 3.7 in the same order). In contrast, the Austrian teachers asserted that oral communication was the area where the students had improved the most (average mark: 4.9). This estimation also corresponds to what they previously defined as one of the characteristic features of their particular school setting, namely that it fosters "a lot of interaction" [QR.NST.1m] in the bilingual lessons and English-speaking conversations among the students during breaks (Coord.VBS[78], e-mail: 9 June 2008). While it may therefore be the case that the learners' *speaking* skills progressed indeed, to a larger extent in the present VBS context than they did in the Swedish CLIL classes, without any further information, the 'interior' differences in the ratings between the Austrian teachers and students can hardly be accounted for.

Overall, the correlation (see section 5.6. for an explanation of this statistical concept) between the rankings of the six linguistic areas is considerably stronger for the Swedish teacher-student pair (0.81) than for the corresponding two Austrian groups (0.49). With regard to the evaluation of the Austrian vs. Swedish teachers, the coefficient amounts to a value of 0.64, which is a fairly moderate degree of interrelation. However, as Table 7:1 has shown, the strongest association exists in the

[78] Name anonymized

average 'improvement marks' between the two CLIL (student) groups themselves. Here, the correlation rate is as high as 0.97.

Closely connected to the teachers' evaluation of the students' linguistic development is yet another factor that was covered in the questionnaires, namely the treatment of language points. As expected, in this connection, there were considerable differences in the responses between the VBS teachers and the native speakers: While all of the natives assured that they explicitly focused on linguistic errors in the subject lessons (i.e. answer option 'true'), only one of their Austrian colleagues did the same. (Not surprisingly, this was the teacher who had English as a second subject.) As for the other VBS instructors, two teachers each claimed that they often (i.e. 'mainly true') or sometimes ('partly true') corrected the students' language mistakes. The remaining three indicated that they never ('not true') paid any attention to linguistically incorrect utterances. This result makes sense, since the majority of the participants agreed that it was predominantly the native teachers' task to "provide the English language input" [QR.TVBS.4f] and act as "[t]he monitor who corrects language, especially pronunciation" (Mewald 2004: 239).[79]

Interestingly, with reference to the statement "I regularly conduct vocabulary tests in the subject lessons.", none of the informants decided on option 1 ('true'), and only one teacher each selected 'mainly true' and 'partly true' respectively. All the other participants (i.e. the three natives as well as six subject teachers) chose the alternative 'not true'. Hence, it might be speculated that, in contrast to traditional foreign language classes, in the present CLIL group, vocabulary acquisition happened indeed more or less incidentally. However, as classroom procedures and learning processes are difficult to assess, even if the persons involved are monitored during lessons,

> [n]eedless to say, it is virtually impossible to say anything about these factors if classroom observations are not part of the investigation. (Sylvén 2004: 199)

Further research, possibly also in the form of interviews, would therefore be necessary to correctly interpret the implications of the above-mentioned result.

As a final point in the present discussion, once again, I would like to draw the attention to the aspect of motivation. In the previous sections, it has been shown that

[79] For more details on the amount and quality of corrective feedback in Austrian CLIL classrooms, compare Dalton-Puffer (2005: Chapter 9).

the CLIL students not only had a more positive attitude towards their choice of educational programme, but that their general affection for English even motivated them to use the language in their leisure time to a much larger extent than their traditional peers did. This finding is also substantiated by the teachers' reactions to the statement "BI increases the students' motivation." In this connection, all of the native speakers and three of their Austrian colleagues chose the answer category 'true'. The remaining five VBS instructors selected the option 'mainly true'.

Besides, *Content and Language Integrated Learning* turned out to be equally motivating for the teachers themselves, since as many as six of them (3 natives and 3 subject teachers) rated the statement "BI increases my own motivation." as 'true', and five went for the category 'mainly true'. This overwhelmingly positive attitude becomes also obvious from certain answers in the open-ended section at the end of the questionnaire:

What are your personal motives for working as a native speaker teacher of VBS?

Besides loving teaching and the students, I teach at a school with excellent resources and good people [QR.NST.1m]

What are your personal aims in your work with bilingual instruction?

To both culturally and educationally enrich children's lives.
Drive them to excellence. Have fun. [QR.NST.1m]

Finally, the teachers' hopes and perspectives for the future of *CLIL* (bilingual instruction) were clear:

→ Hoffentlich Standard an österreichischen Schulen! [QR.TVSBS.1f]

→ Stark zunehmen und immer „normaler" werden … [QR.TVBS.2f]

→ Ich hoffe, dass immer mehr Schulen diesen Schulversuch praktizieren! [QR.TVBS.4f]

→ at some point it will be the standard way of teaching in secondary/upper schools [QR.TVBS.7m]

…

→ It will spread. – The advantages are obvious I think. [QR.NST.1m]

8. Conclusion

The present thesis is a replication of a Swedish study (Sylvén 2004) which investigated the long-term effects of *Content and Language Integrated Learning* on students' incidental vocabulary acquisition. In contrast to the original, comprising three separate test rounds, which were conducted during a period of two full school years (see, e.g. Sylvén 204: 221), the present work is based on a single test sequence, carried out in February/March 2007. Thus, my focus has not so much been on developmental aspects but on students' lexical knowledge and performance at one specific point in time. In this final chapter, I will summarize the main findings of my investigation, also in comparison with the related original results. For this purpose, I will return to my starting point and provide answers to the questions posed in the introduction to the present work. In this connection, I will also offer some suggestions for future research.

8.1. The study

Similar to the original (compare Sylvén 2004: 221), the central concern of the present study was to investigate the following issue:

- ◊ Do CLIL students have a larger and more complex English vocabulary than traditional students?[80]

More specifically, in the present thesis, I have looked at the following research questions:

- ◊ To what extent is students' lexical performance dependent on/reflected by the test type and format used?
- ◊ Do leisure activities, such as reading, writing, watching TV/movies and playing computer games in English, have an impact on students' lexical proficiency?
- ◊ To what extent does time spent in English-speaking countries contribute to students' lexical proficiency?

[80] In Sylvén's longitudinal study the corresponding question was as to whether CLIL students *acquired* a larger and more complex English vocabulary (Sylvén 2004: 4; 221). For the above-mentioned reason, in the present work, we could only focus at their level of lexical knowledge at the time of the test performance. Any statements about the actual process of acquisition (and learning) are therefore more or less speculative.

- ◊ Are there gender differences with regard to test performance and lexical proficiency?
- ◊ To what extent do students' native languages affect their performance with regard to English vocabulary?
- ◊ What is the key to acquiring a high-level lexical competence, i.e. what are the characteristic features of the highest performing students in the present test population?

Apart from these topic-related questions, I have also dealt with two, more general aspects in connection with *CLIL*:

- ◊ Are there differences in motivation, attitude and self-assessment between CLIL students and traditional students?
- ◊ How do teachers estimate the success of their school's bilingual programme/the CLIL method in general?

As mentioned above, in order to answer these questions, an empirical study was carried out. The participants were 33 students of grade eleven (i.e. third year of upper secondary school) at a Viennese grammar school. Of these, 21 were in the CLIL group (10 females, 11 males) or VBS group[81], and 12 (9 females, 3 males) in the traditional group. The CLIL students attended the natural science programme ('Realgymnasium'), the traditional students the upper secondary section with a special focus on languages, in this particular case Latin and French ('Gymnasium'). All students participated in the entire test sequence, except for one girl, who was missing during the last lexical test.

Overall, the sequence consisted of a battery of five individual tests (four different types of lexical tests), a *self-report* test, a *words in context* test, a *multiple choice* test and two *cloze* tests.

In the *self-report* test the students were asked to indicate their level of knowledge of each lexical item according to a five-point scale, where level A signifies total unfamiliarity with the word under examination ("I do not remember having seen this word before."), and level E implies that the word is known so well that it can be used

[81] The abbreviation VBS stands for *Vienna Bilingual Schooling*, which is how the bilingual programme at this particular school is called.

with semantic and grammatical correctness in a sentence (to be shown on the test sheet).

In the *words in context* test, the learners had to derive the meaning of a set of 30 lexical items from their use in a newspaper article. The respective answer could either be given in the form of an English synonym, a German translation, or an explanation in English or German.

The *multiple choice* test consisted of 30 sentences, each with one particular lexical item underlined. The students were asked to identify the best synonym for this specific item among five possible alternatives.

Finally, the two *cloze* tests, measured the knowledge of fixed lexical phrases and idiomatic expressions. In each of the 30 sentences under consideration, only **one** particular word was missing which the learners were asked to fill in.

In addition to the five lexical tests, questionnaires were distributed among all students as well as the teachers and native speaker teachers of the VBS group. These question sheets focused on sociolinguistic aspects, such as the participants' native languages, time spent in English-speaking countries and, in the case of the students, their parents' level of education. Besides, both the students and the teachers were asked about personal aims and motives, and their general perception of the CLIL programme. Furthermore, the teacher questionnaires covered issues such as professional background (formal education, teaching expertise), classroom procedures, and materials and methods used in connection with bilingual instruction. On the other hand, the student sheets were particularly concerned with 'extracurricular' sources of English language input, thus including questions on the students' reading, TV, movie and Internet habits, as well as on other leisure activities involving the use of English.

The test results were analysed, not only according to VBS and traditional group, but also with regard to the present vs. the original research context, the students' mother tongue and gender, and some of the background factors mentioned above. Moreover, the answers in the *self-report*, *words in context* test and the *two cloze* tests were classified on the basis of their respective quality (e.g. *correct*, *acceptable*, *incorrect*). Due to the limited scope of the present study, an explicit error analysis was not

carried out. For a detailed account on possible sources for erroneous answers, consult Chapter 6 in the original study (Sylvén 2004: 119-178).

8.2. Main findings and suggestions for future research

After all tests have been corrected and scored, and all questionnaires analysed, our research questions can be answered in the following way:

◊ *Do CLIL students have a larger and more complex English vocabulary than traditional students?*

If we look at the figures in Table 8:1, it becomes obvious that the answer to this question is a resounding yes:

Table 8:1 *Total mean score (round numbers), factor: group (and research context)*

	CLIL	Traditional
Seregély (2008), total	218	151
Sylvén (2004), TR III	160	122
Seregély (2008), comparative values[82]	185	136

In the present as well as the original study (see also Sylvén 2004: 225), the CLIL students significantly outperformed their traditional peers. What is more, in both the CLIL group and the traditional group, the Austrian participants scored slightly above their respective Swedish counterparts, involved in the comparative test round III.

In Sylvén's study it has been shown that, on average, CLIL students also **acquire** a larger English vocabulary than their traditional peers. This finding is based on the two groups' improvement in mean scores during the individual test rounds and in particular, during the entire period of two school years (compares the figures in Sylvén 2004: 225). However, in this connection, it should be noted that the Swedish CLIL students were already ahead in test round I, at the beginning of year 10, that is, "before they had an extra English input in the form of CLIL" (Sylvén 2004: 223). This result reflects one of the major problems common to all studies on the effects of *CLIL*:

[82] calculated from the total mean score of the reduced battery (excluding *cloze* test 2); in addition, with regard to the CLIL group the scores of three different subgroups are taken into account: the original (total) VBS group, the non-English subgroup and the subset of exclusively German speakers (for more details, see Chapter 4).

> [T]here is (normally) a requirement of a certain level of proficiency before entering a CLIL class. Therefore, students who enjoy, and are interested in, English are more likely to find the CLIL method appealing than those who are less interested in English. (Sylvén 2004: 180)

Apparently, this situation also applies to the present school setting, where the CLIL/VBS classes typically consist not only of linguistically gifted Austrian students, but also of native speakers of English with a particular interest in German (see Chapter 3). Thus, it might be speculated that the Viennese CLIL students, just as their Swedish colleagues, already had a more profound lexical knowledge than the traditional students before their first, actual experience with the method itself.

Another difficulty in analysing the test outcomes has been mentioned in both Chapters 4 and 6, viz. the fact that the VBS group's superior lexical achievement cannot only be traced to the CLIL input alone. Background factors such as voluntary reading or speaking and writing English for pleasure also play an essential role, as will be discussed in detail in connection with one of the subsequent questions below. Moreover, the Austrian CLIL students also receive EFL input in the traditional way, that is, in the English language classroom. Hence, with regard to the lexical items under examination, one can never be absolutely certain as to which of them were incidentally **acquired**, and which consciously **learned**.

Notwithstanding these objections, the answer to the above question remains the same: The average CLIL student undoubtedly **has** a larger and more complex English vocabulary than her/his traditional peer. However, it seems virtually impossible to determine which parts of her/his lexicon were learned, or acquired within, or outside the school context, and which ones already existed before bilingual instruction actually took place. Further research into this direction would therefore be most welcome.

◊ *To what extent is students' lexical performance dependent on/reflected by the test type and format used?*

Table 8:2 shows that the VBS group was in the lead with regard to all five lexical tests.

Table 8:2 *Mean score, individual tests (maximum test score)*

	SR (150)	WIC (60)	MC (30)	CT1 (60)	CT2 (60)
VBS	97,90	39,81	25,10	25,48	29,86
Traditional	66,42	34,83	20,67	14,08	16,09

As illustrated in Chapter 5, in the majority of the cases, the differences between the two test groups are statistically significant or even highly significant. The only exception to this overall tendency is the *words in context* test. Here, the traditional students, too, managed to attain relatively high scores. This is mainly due to the fact that this particular test type measures not only the students' ability to derive the meaning of words from their context, but also to express it in an adequate manner. Since the respective answer could be given in English **or** German, even less proficient learners of English have the chance to achieve favourable results.

Indeed, a comparison of the correlation rates between each individual test type and the overall outcome (see section 5.6.) has shown that the *words in context* test is less representative of the students' lexical proficiency in English than any of the other tests in the battery. Nonetheless, in an immersion setting such as the present one, it may still serve as a useful tool: If we restricted the answer mode to German explanations and translations, we could test the linguistic improvement and productive skills of the second VBS target group, that is, the English natives.

The importance of test type and format also becomes noticeable from the results on the two *cloze* tests. In this connection, I did not only detect the usual differences between VBS and traditional students, but also considerable variations among individual subjects **within** each of the two groups, in particular the VBS group. Since this specific test type presents the most difficult task, it has been concluded that, among the entire test population, the *cloze* top scorers are the students with the highest level of lexical proficiency. The 'secret(s)' behind their success will be revealed later on.

Another test format that clearly indicates the state of the learners' lexical knowledge is the five-part *self-report* scale. As illustrated in section 5.1.4., here, the CLIL students frequently opted for the most complex answer mode, thus providing a full sentence, which showed how the lexical item under study was typically used. On the other hand, their traditional peers were more likely to give a simple German translation or English synonym, when they thought or were certain that they knew the meaning of a particular word.

Overall then, the total outcome discussed in the context of the previous question is also reflected in the results on the separate tests. Yet, certain tasks turned out to be more representative of the students' lexical competence than others.

◊ *Do leisure activities, such as reading, writing, watching TV/movies and playing computer games in English, have an impact on students' lexical proficiency?*

If this question refers to the two groups in general, then the answer is definitely yes. As illustrated in Chapter 6, the background factors having the greatest influence on the outcomes of the present test participants are speaking English for pleasure and voluntary reading of English texts. While it has been proven that the former factor is of particular relevance only for the traditional learners (see section 6.2.), the latter also has a decisive impact on the results of the VBS students. In both test groups, on average, the participants who indicated that they regularly read English texts other than compulsory homework-exercises attained higher overall scores than their respective non-reading peers did. However, the differences between readers and non-readers, just as those between 'speakers' and 'non-speakers', 'writers' and non-writers' etc. seem to be less significant than the differences between the two test groups themselves. In other words, irrespective of their extracurricular usage of and exposure to the English language, the subjects taught in the traditional manner are inferior in lexical achievement and proficiency to their colleagues who receive constant English language input by means of *CLIL*.

This finding stands in contrast to the results in the original study, as Sylvén (2004: 225) has discovered that

> a traditional student who receives a great deal of English input outside of school may score above a CLIL student who mainly gets English input in the classroom.

At any rate, in this respect, the two research contexts are hardly comparable, since the present one only comprises a single test round, whereas the original focuses in detail on the long-term effects of voluntary reading habits vs. *CLIL* (compare Sylvén 2004: 111-118).

However, what has been found in both studies is that, as a general tendency, CLIL students more often read and write English texts for pleasure, watch English movies and TV programmes, and visit English-speaking websites, than their traditional peers. These factors, taken together with the higher amount of classroom input, fully account for the CLIL groups'[83] outstanding results.

◊ *To what extent does time spent in English-speaking countries contribute to students' lexical proficiency?*

In Chapter 6, it has been shown that the highest overall and individual test scores were attained by the VBS students who had spent more than three months in English-speaking countries. What is more, the traditional student who had stayed abroad for one school year not only outperformed her classmates but also managed to score approximately on a par with the VBS subjects who lacked such an experience. Even though these outcomes seem fairly convincing, the low number of informants and the restricted time span of the present investigation make it impossible to draw any far-reaching conclusions regarding the impact of time spent in English-speaking countries as opposed to that of *CLIL*.

◊ *Are there gender differences with regard to test performance and lexical proficiency?*

Since the original study has revealed that there are major differences in performance between the genders (compare Sylvén 2004: 227), this aspect has also been one of my main research concerns. As the related analyses have shown (see Chapters 4 and 5), just as in the Swedish context, in both Austrian test groups, the males outperform the females in almost all of the areas tested, the only exception being the two *cloze* tests, where the female students have more *correct* answers than their male peers do.

[83] both the Swedish and the Austrian

The gender-specific variations are more striking in the VBS group, but then again, these figures are also more representative, given that males are under-represented in the traditional group.

Although the overwhelming superiority of the male students is well in line with the findings of previous empirical studies (compare the discussion in Sylvén 2004: 215-218), it is nonetheless quite surprising, especially if we consider that in the Austrian context, the female students in both test groups are clearly ahead of the males with regard to their average English marks (for more details, see 4.3.3.). The relation between the format of (linguistic) tests, or other types of examination, and gender would therefore be an interesting area for further research, also in view of the educational standards that are currently established (or to be established) for the teaching and learning at Austrian schools.

◊ *To what extent do students' native languages affect their performance with regard to English vocabulary?*

While the present traditional group consists exclusively of native speakers of German, the VBS group also comprises four English-speaking students and one participant each with Albanian, Croatian and Swedish as their respective mother tongue. My initial assumption was that the English natives, and possibly also the other 'international students', would perform better than their 'typical Austrian' classmates. This hypothesis could only be confirmed to a certain extent. Even though some of the 'international' participants are indeed among the top scorers in the overall test sequence, as well as with regard to some of the lexical tests, the same applies neither to the English, nor to the non-German, subset **as a whole**. However, in this connection, it should once again be noted that this result may just as well be merely accidental, as the individual VBS subsamples are too small to represent any general rule.

Corresponding to the findings discussed above (see question 2), in the *words in context* test, on average, the German-speaking CLIL students achieved slightly better results than their peers with English or any other of the aforementioned foreign languages as their L1. Not surprisingly, concerning the two *cloze* tests, the situation

was more or less converse. Here, the English natives were clearly in the lead in terms of mean scores.

◊ *What is the key to acquiring a high-level lexical competence, i.e. what are the characteristic features of the highest performing students in the present test population?*

As mentioned above, my definition of 'top scorer' refers to those students who attained above-average values in *cloze test 1* and *2*.[84] This makes sense, if we consider that this test type is the most difficult one, as it measures native-like lexical skills (see also the description in Chapter 4). What is more, the results on the two *cloze* tests have turned out to be not only fairly similar (The corresponding correlation rate amounts to 0.74.), but also highly representative of the test outcome as a whole. (The value for the correlation between the individual test result and the overall outcome is 0.86 for *cloze test 1* and 0.82 for *cloze test 2*, meaning that in both cases, the correlation is considerably strong.)

In connection with the previous question, I have already indicated that one determining factor behind a high *cloze* test value is the student's mother tongue. Not surprisingly, in both *cloze* test *1* and *2*, three out of the ten highest performing subjects are native speakers of English (in *cloze test 2* even all four English-speaking students are among the top ten). This result appears self-explanatory; what is more interesting is to discover whether there are common denominators for the non-English participants who achieved excellent *cloze* test scores.

If we take a closer look at the top ten positions in both *cloze* tests, it becomes noticeable that, disregarding the Anglophone participants, they are occupied by five individuals altogether, three females and two males, four CLIL learners and one student of the traditional group. One of the top scorers has Albanian as her first language, the remaining four are native speakers of German. All five students have been to English-speaking countries, but only two of them for more than one year. (Of the latter, one has spent five years at an international, exclusively English-speaking, school abroad.)

[84] Here, the critical values are the mean scores of the entire student population, that is, 21.33 for *cloze test 1* and 25.13 for *cloze test 2*.

Three informants regularly use the English language when talking to relatives or friends. All of the high achievers read English fiction, magazines or newspapers for pleasure at least once a month. One of them even does so every day. All except the traditional student watch English TV programmes on a monthly basis, or occasionally. The same holds true for movies; yet, with regard to this specific question, also the traditional student indicates that she goes to see English films at the cinema once in a while.

PC games only seem to be of interest for the two male students,[85] while English-speaking websites are visited daily (n = 2), weekly (n = 2) or monthly (n = 2) by both the males and the females in this 'exclusive' group. Correspondingly, three of the best performers voluntarily write English texts, in particular e-mails and chat entries, every day (n = 2) or at least once a week (n = 1).

All except one of the *cloze* test top scorers come from a family background where either one or both of the parents has/have a university education. This finding is well in line with the related result in the original study (see Sylvén 2004: 204-207), but nevertheless difficult to explain:

> [W]hether this [i.e. the fact that students from more highly educated family backgrounds outperform their peers whose parents have little or no education above compulsory, vocational or upper secondary school] is due to well-educated parents being more focused on their children's schoolwork or to other factors is beyond the scope of this study. (Sylvén 2004: 227)

Finally, all except one of the students under consideration have above-average English (and German)[86] marks, i.e. a 'Sehr gut' or 'Gut'. However, this close correlation between scores and marks does not apply to the student population in general, as we have already seen in connection with the gender-discussion that school performance and test performance do not necessarily coincide.

Overall then, in the present research context, the typical (*cloze* test) top scorer is either an English native or a participant who

[85] This is a general tendency among **all** participants, not only the top scorers. In the original study, it has also been discovered that "the involvement in computer games and role plays [is] typical of male students" (Sylvén 2004: 226), one possible factor accounting for the gender-related differences in the overall result.

[86] Although good German marks seem to be irrelevant in the case of the two *cloze* tests, it has been shown that they may be beneficial to the students' performance on the *words in context* test, to a lesser extent probably also the *self-report* test (including German translations as one possible answer mode), and thus the test outcome as a whole.

- comes from an educated family background
- has not necessarily spent more than three months in English-speaking countries
- regularly reads English texts on a voluntary basis
- frequently surfs on English-speaking websites and
- has an above-average mark in English (and German) as a separate subject.

Even though the above description is not explicitly linked to one of the two groups, in the overall sequence, the highest performing students are more often than not to be found in the CLIL/VBS class. Indeed, only two traditional learners attained total test values that were approximately at the same level as their CLIL-trained peers' (see also Chapter 4), namely the one mentioned in the above discussion, and her colleague, who has stayed in an English-speaking country for one year (see question 4). However, with regard to the two *cloze* tests, the latter student scored even **below** the average values calculated for the traditional group (14.08 in *cloze test 1* and 16.09 in *cloze test 2*). Although this may be a chance result, it closely corresponds to what has been indicated above: The factor of time spent in English-speaking countries undoubtedly has a positive impact on the students' vocabulary knowledge and test performance; yet, it does not seem to be particularly relevant for completing the most difficult of the tasks.

◊ *Are there differences in motivation, attitude and self-assessment between CLIL students and traditional students?*

If we simply look at the students' answers to the questions about advantages and disadvantages of their upper secondary programmes (see Chapter 7), it is difficult to discern which of the two groups is generally more motivated than the other one. Both the CLIL students and the traditional students praise the linguistic focus (English-German vs. French) as well as subject-specific aspects of their respective programme. In the same way, both complain about certain contents, methods, teachers, and - in the case of the VBS group - also about the excessive amount of workload. Likewise, the average 'satisfaction marks' are approximately equal for the two groups: 2.62 for the VBS group and 2.83 for the traditional group.

However, if we take into account that the CLIL students more frequently use the English language for various purposes in their leisure time (see question 3), it becomes obvious that, although their *extrinsic motivation*, in connection with the educational setting, is not necessarily higher, they are undoubtedly more *intrinsically* motivated to acquiring the target language than their traditional peers. Besides, the outcomes of the questionnaires have shown that the VBS students have a more positive attitude towards speaking in front of an audience in English **and** German, and also rate their own linguistic development in the course of secondary education considerably higher than the average control student does. This seems to be another reason for their remarkable performance, as Sylvén (2004: 226) points out,

> Not surprisingly, a positive attitude toward the language and a high level of motivation are helpful for the language acquisition process at large, whereas students whose attitudes are less positive and who lack motivation will have greater difficulties improving their lexical proficiency.

◊ *How do teachers estimate the success of their school's bilingual programme/the CLIL method in general?*

As illustrated in Chapter 7 (Table 7:2), the VBS teachers are even more convinced about their students' linguistic progress achieved by way of bilingual instruction than the students themselves. While the teachers' personal aims and motives for implementing *Content and Language Integrated Learning* are varied, almost all of them express the wish that the method will soon become the standard way of teaching at Austrian schools. Yet, the actual realisation of this objective seems to be crucially dependent on a series of external factors, above all, financial support for teacher training, material and human resources (especially native speaker teachers), and the establishment of clear-cut, national curricula for *CLIL*. Even though many attempts have been made into this direction, the majority of those by individual teachers or schools, in Austria as well as in other European countries, "CLIL is still far from being a consolidated and fully articulated educational model" (Dalton-Puffer 2007: 1).

The only way to solving this problem is to develop a comprehensive theoretical framework that takes into account the findings of previous studies on the benefits and possible weaknesses of the method and prepares the ground for further research. In other words, the introduction of *CLIL* into mainstream teaching will only be achieved

through the joint effort of educationalists, schools, researchers (applied linguists as well as subject specialists) and the national (and international) authorities concerned.

Evidently, in today's complex and rapidly progressing world, the development of educational standards, not only for bilingual instruction, but for (language) teaching in general, will forever be an ongoing process. Nonetheless, I hope that the present thesis has provided at least some new insights and encouraged the VBS teachers to proceed and extend their valuable work.

References

Abendroth-Timmer, Dagmar. 2007. *Akzeptanz und Motivation: empirische Ansätze zur Erforschung des unterrichtlichen Einsatzes von bilingualen und mehrsprachigen Modulen*. Frankfurt: Peter Lang.

Aitchinson, Jean. 1994. *Words in the mind*. Oxford: Blackwell.

Alderson, J. Charles; Clapham, Caroline; Wall, Diane. 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Arnaud, Pierre J. L.; Savignon, Sandra J. 1997. "Rare words, complex lexical units and the advanced learner". In Coady, James; Huckin, Thomas (eds.). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 157-173.

Baker, C.; Prys Jones, S. 1998. *Encyclopedia of bilingualism and bilingual education*. Clevedon: Multilingual Matters.

Bolger, Niall; Kellaghan, Thomas. 1990. "Method of measurement and gender differences in scholastic achievement". *Journal of educational measurement* 27:2, 165-174.

Boyd Zimmerman, Cheryl. 1997. "Historical trends in second language vocabulary instruction". In Coady, James; Huckin, Thomas (eds.). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 5-19.

Breitsprecher, Roland et al. 1999. *PONS Globalwörterbuch Teil 1 Englisch-Deutsch*. Wien: ÖBV Pädagogischer Verlag.

Brinton, Donna. M.; Snow, Margeruite A.; Wesche, Marjorie B. 1989. *Content-based second language instruction*. New York: Newbury House.

British National Corpus. <http://www.natcorp.ox.ac.uk/index.xml>.

Brooks, Nelson. 1964. *Language and language learning*. (2nd edition). New York: Harcourt, Brace & World.

Carter, Ronald. 1987. *Vocabulary – applied linguistic perspectives*. London: Allen & Unwin.

Carter, Ronald; McCarthy, Michael. 1988. *Vocabulary in language teaching*. London: Longman.

Caspari, Daniela; Werner, Bettina; Zydatiß, Wolfgang. 2007. "Bili macht Schule – und Schule macht Bili …: Herausforderungen im Bilingualen Unterricht in Berlin. Dokumentation einer Podiumsdiskussion". In Caspari, Daniela et al. (eds.). *Bilingualer Unterricht macht Schule: Beiträge aus der Praxisforschung*. Frankfurt: Peter Lang, 221-228.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: The MIT Press.

CLIL-Compendium. 2001. <http://www.clilcompendium.com/clilcontacts.htm> (5 September 2008).

Coady, James; Huckin, Thomas (eds.). 1997. *Second language vocabulary acquisition*. Cambridge: Cambridge University Press.

Collins COBUILD English dictionary. 1995. London: HarperCollins.

Crichton, Hazel. "The Vienna Bilingual Schools Programme". 2006. *Scottish languages review 14*. <http://www.scilt.stir.ac.uk/Downloads/slr/issue_14/SLR14Crichton%20final.pdf> (14 January 2008).

Cruse, Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Cummins, Jim; Swain, Merrill. 1986. *Bilingualism and education*. New York: Longman.

Daller, Helmut; Milton, James; Treffers-Daller, Jeanine (eds.). 2007. *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.

Dalton-Puffer, Christiane. 2005. *Immersion classrooms as language learning environment: applied linguistics perspectives.* Habilitationsschrift. University of Vienna.

Dalton-Puffer, Christiane. 2007. "Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe". In Delanoy, Werner; Volkmann, Laurenz (eds.). *Further perspectives for English Language Teaching*. Heidelberg: Carl Winter, 1-15.

Darn, Steve. 2006. "Content and Language Integrated Learning". *British Council BBC: Teaching English*. <http://www.teachingenglish.org.uk/think/articles/content-language-integrated-learning> (5 September 2008).

Deutsche Gesellschaft für Qualität e.V. 1993. *Gesamtlehrgang Teil II/A*. (2. Ausgabe). Frankfurt: Deutsche Gesellschaft für Qualität e.V.

Duden Rechtschreibung der deutschen Sprache. 1996. Mannheim: Bibliographisches Institut & F. A. Brockhaus AG.

Ellis, Nick C. 1994. “Consciousness in second language learning: psychological perspectives on the role of conscious processes in vocabulary acquisition”. In Hulstjin, Jan H.; Schmidt, R. (eds.). *Consciousness in second language learning. AILA Review* 11: 37-56.

Ellis, Nick C. 1997a. "Implicit and explicit language learning: an overview”. In Ellis, Nick C. (ed.). *Implicit and explicit learning of languages*. London: Academic Press, 1-32.

Ellis, Nick C. 1997b. “Vocabulary acquisition: the implicit ins and outs of explicit cognitive mediation”. In Ellis, Nick C. (ed.). *Implicit and explicit learning of languages*. London: Academic Press, 211-282.

Ellis, Rod. 1990. *Instructed second language acquisition*. Oxford: Basil Blackwell.

Ellis, Rod. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, Rod; He, Xien. 1999. “The roles of modified input and output in the incidental acquisition of word meanings”. *Studies in second language acquisition* 21, 285-301.

ESTIA Sweden. 2003. “The Swedish education system”. <http://www.estia.educ.goteborg.se/sv-estia/edu/edu_sys.html> (19 May 2008).

European Commission. 2003. “Promoting language learning and linguistic diversity: an action plan 2004 – 2006”. *COM(2003) 449*. <http://ec.europa.eu/education/doc/official/keydoc/actlang/act_lang_en.pdf> (4 October 2008).

European Commission. 2007. “Language teaching: CLIL (Content and Language Integrated Learning)”. *Education and training*. <http://ec.europa.eu/education/policies/lang/teach/clil_en.html> (4 October 2008).

European Community. 1984. *The Council Document 7706/84*.

Eurydice. 2006. *Content and language Integrated learning (CLIL) at school in Europe: Austria, national description – 2004/05*. <http://eacea.ec.europa.eu/ressources/eurydice/pdf/070DN/070_AT_EN.pdf> (4 October 2008).

Finocchiaro, Mary; Brumfit, Christopher. 1983. *The functional-notional approach: from theory to practice*. New York: Oxford University Press.

Gardner, Robert C.; Lambert, Wallace E. 1972. *Attitudes and motivation in second-language learning* Rowley: Newbury House.

Gardner, Robert C. 1985. *Social psychology and second language learning: the role of attitudes and motivation*. London: Edward Arnold.

Gierlinger Erwin. 2002. "Fremdsprache als Arbeitssprache in Österreich: Situation und Perspektive". *Comenius 2.1. project*. <http://www.pze.at/memo/download/05bestat.pdf> (22 August 2008), 1-18.

Grabe, William; Stoller, Fredricka L. 1997. "Content-based instruction: research foundations". In Stryker, Stephen B; Leaver, Betty L (eds.). 1997. *Content-based instruction in foreign language education: models and methods*. Washington D.C.: Georgetown University Press, 5-21.

GRg 23/VBS. 2005. "Vienna Bilingual Schooling Middle Schools – Sekundarstufe: Taking a world view". *VBS-Folder Unterstufe des GRg 23*. <http://www.grg23vbs.ac.at/fileadmin/media/infomat/folder/15E.pdf>(22 August 2008), 1-2.

GRg 23/VBS. 2008."Bilinguale Schule – Bilingual Schooling". *Gymnasium Draschestraße Vienna Bilingual Schooling*. <http://www.grg23vbs.ac.at/home.html> (22 August 2008).

GRg 23/VBS. 2008. "Unterrichtsorganisation". *Gymnasium Draschestraße Vienna Bilingual Schooling*. <http://www.grg23vbs.ac.at/unterrrichtsinfo.html> (23 August 2008).

Haastrup, Kirsten. 1991. *Lexical interferencing procedures or talking about words*. Tübingen: Gunter Narr Verlag.

Haastrup, Kirsten; Henriksen, Birgit. 1998. "Vocabulary acquisition: from partial to precise understanding". In Haastrup, Kirsten; Viberg, Åke (eds.). *Perspectives in lexical acquisition in a second language*. Lund: Lund University Press, 97-123.

Haastrup, Kirsten; Henriksen, Birgit. 2000. "Vocabulary acquisition: acquiring depth of knowledge through network building". *International journal of applied linguistics* 10:2, 221-240.

Hall, Per. 1996. *TCFL Teaching content through a foreign language*. Mittöskolan.

Harding, J. 1980. "Sex differences in performance in science examinations". In Deem, R. (ed.). *Schooling for women's' work*. London: Routledge & Kegan Paul.

Hartiala, Aini-Kristiina. 2000. *Acquisition of teaching expertise in content and language integrated learning*. Turku: Turun Yliopisto.

Hedge, Tricia. 2002. *Teaching and learning in the foreign language classroom*. (3rd impression). Oxford: Oxford University Press.

Hellakant, Jan. 1994. “Are multiple-choice tests unfair to girls?”. *System* 22:3, 349 352.

Henriksen, Birgit. 1999. “Three dimensions of vocabulary development“. *Studies in second language acquisition* 21:2, 303-317.

Howatt, Anthony P. R. 1984. *A history of English language teaching*. Oxford: Oxford University Press.

Hymes, Dell. 1972. “On communicative competence”. In Pride, John B.; Holmes, J. (eds.). *Sociolinguistics*. New York: Penguin, 269-293.

Johnson, Keith. 2001. *An introduction to foreign language learning and teaching*. Essex: Pearson Education Limited.

Kelly, Louis G. 1969. *25 centuries of language teaching*. Rowley: Newbury House.

Kemmeter, Luise. 1997. *Multilingual gestütztes Vokabellernen im gymnasialen Englischunterricht*. Frankfurt: Peter Lang.

Krahnke, K. 1987. *Approaches to syllabus design for foreign language teaching*. New York: Prentice Hall.

Krashen, Stephen D.; Terrell, Tracy D. 1983. *The natural approach: language acquisition in the classroom*. Oxford: Pergamon.

Krashen, Stephen D. 1985. *The input hypothesis: issue and implications*. Harlow: Longman.

Krashen, Stephen D. 1989. “We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis”. *The modern language journal* 73: iv, 440-464.

Krashen, Stephen D. 1993a. “The case for free voluntary reading”. *Canadian Modern Language Review* 50:1, 72-82.

Krashen, Stephen D. 1993b.*The power of reading*. Inglewood: Libraries Unlimited.

Lambert, Wallace. 1977. “Cognitive consequences of bilingual schooling”. *Linguistics* 187:3, 13-33.

Laufer, Batia. 1997. “The lexical plight in second language reading: words you don’t know, words you think you know, and words you can’t guess”. In Coady, James; Huckin, Thomas (eds.). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 20-34.

Laufer, Batia; Hulstjin, Jan H. 2001. “Incidental vocabulary acquisition in a second language: the construct of task-induced involvement”. *Applied Linguistics* 22: 11, 1-26.

LEO Deutsch-Englisches Wörterbuch: Ein Online-Service der LEO GmbH. <http://dict.leo.org/>.

Lewis, M. 1993. *The lexical approach: the state of EFL and a way forward*. Hove: Language Teaching Publications.

Lexico Publishing Group LLC. 2008. *Thesaurus.com: Roget's new millennium thesaurus*. < http://thesaurus.reference.com/browse/authorities> (13 March 2008).

Lightbown, Patsy M.; Spada, Nina. 1999. *How languages are learned*. (2nd edition). Oxford: Oxford University Press.

Linnarud, Moira. 1986. *Lexis in composition*. Malmö: Liber.

Ljung, Magnus. 1990. *A study of TEFL vocabulary*. Stockholm: Almqvist & Wiksell.

Long, Michael H.; Richards, Jack C. 1999. "Series editors' preface". In Singleton, David. *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press, xi-xii.

Long, Michael H.; Richards, Jack C. 2001. "Series editors' preface". In Nation, I.S.P. *Learning vocabulary in another language*. Cambridge: Cambridge University Press, xiii.

Macmillan English dictionary for advanced learners. 2002: Oxford: Macmillan Education.

Meara, Paul. 1996a. "The dimensions of lexical competence". In Brown, G.; Malmkjær, K.; William, J. (eds.). *Performance and competence in second language acquisition*. Tampere: Publications de l'Association Finlandaise de Linguistique Appliquée (AFinLA), 35-35.

Meara, Paul. 1996b. "The vocabulary knowledge framework". *The lognostics virtual library*. <http://www.lognostics.co.uk/vlibrary/meara1996c.pdf> (18 April 2008).

Mewald, Claudia. 2004. *Paradise lost and found: a case study of content based foreign language education in Lower Austria.* Unpublished PhD thesis. Norwich: Department of Education and Professional Development, University of East Anglia.

Mobärg, Mats. 1997. "Acquiring, teaching and testing vocabulary". *International journal of applied linguistics* 7:2, 201-222.

Murphy, R.J.L. 1982. "Sex differences in objective performance". *British journal of educational psychology* 52, 213-219.

Nation, Paul; Coady, James. 1988. "Vocabulary and reading". In Carter, Ronald; McCarthy, Michael (eds.). *Vocabulary and language teaching.* London: Longman, 91-110.

Nation, Paul. 1990. *Teaching and learning vocabulary*. Boston: Heinle & Heinle.

Nation, I.S.P. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nikula, Tarja. 2002. "Teacher talk reflecting pragmatic awareness: a look at EFL and content-based classroom settings". *Pragmatics* 12:4, 447-467.

Nixon, John. 2000. *Content and language integrated learning and teaching in Sweden*. Stockholm: Swedish National Agency for Education.

Norris, S. 2001. "Der Mythos des Native Speakers. Native Speaker sind ein gutes Werbemittel, doch profitieren auch die Schüler?". *Buch & Co*, 73 A: 3/2001.

Oxford advanced learner's dictionary of current English. 2000. (6th edition). Oxford: Oxford University Press.

Oxford collocation dictionary for students of English. 2002. Oxford: Oxford University Press.

Oxford dictionary of English idioms. 1993. Oxford: Oxford University Press.

Paribakht, T. Sima; Wesche, Marjorie. 1997. "Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition". In Coady, James; Huckin, Thomas (eds.). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 174-200.

Pickett, G. D. 1978. "The foreign language learning process: an occasional paper". *ETIC Publications*. London: The British Council.

Poisel, Eva. 2003?. "Die Bilinguale Oberstufe am GRG23 VBS Draschestraße- ein gelungenes Beispiel für eine attraktive neue Form der Oberstufe." Unpublished article.[87]

"Protokoll der Generalversammlung des Vereins 'Freunde der VBS' vom 7.12.2006". <http://www.grg23vbs.ac.at/fileadmin/media/shared/vbsverein/2006_protokoll_der_generalversammlung.pdf> (22 August 2008).

Richards, Jack C. 1976. "The role of vocabulary teaching". *TESOL Quarterly* 10: 1, 77-89.

Richards, Jack C.; Rodgers, Theodore S. 2001. *Approaches and methods in language teaching*. (2nd edition). Cambridge: Cambridge University Press.

[87] Unfortunately, the exact details of the publication are unknown (year of publication is either 2003 or 2004).

Rietveld, Toni; van Hout, Roeland. 1993. *Statistical techniques for the study of language and language behaviour*. Berlin: Mouton de Gruyter.

Rivers, Wilga M. 1964. *The psychologist and the foreign language teacher*. Chicago: University of Chicago Press.

Rivers, Wilga M. 1981. Teaching foreign language skills. (2nd edition). Chicago: University of Chicago Press.

Rivers, Wilga M. 1983. *Communicating naturally in a second language: theory and practice in language teaching*. New York: Cambridge University Press.

Rojahn, Christian; Bollinger, Susan. 2005. *Großes Handbuch English idioms*. München: Compact Verlag.

Schmitt, Norbert. 1999. "Current trends in second language vocabulary". *TESOL research newsletter* 6:2, 3-5. <https://www.nottingham.ac.uk/english/research/cral/doku.php?id=people:schmitt#journal_articles> (15 September 2008).

Schmitt, Norbert. 2000. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Simensen, Aud Marit. 1998. *Teaching a foreign language*. Bergen: Fagbokforlaget.

Singleton, David. 1999. *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Spolsky, Bernard. 1995. *Measured words*. Oxford: Oxford University Press.

Stadtschulrat für Wien. 2008a. "Vienna Bilingual Schooling". *Wien.at: Webservice der Stadt Wien.* <http://www.wien.gv.at/bildung/stadtschulrat/schulsystem/vbs.html> (22 August 2008).

Stadtschulrat für Wien. 2008b. "Vienna Bilingual Schooling (VBS)". *Bilingualität Englisch*. <http://stadtschulrat.echonet.at/bilingualitaet/catid18/> (22 August 2008).

Stern, Hans H. 1983. *Fundamental concepts of language teaching*. Oxford: Oxford University Press.

Swales, John; Feak Christine. 1994. *Academic writing for graduate students: a course for non-native speakers of English*. Ann Arbor: University of Michigan Press.

Sylvén, Liss Kerstin. 2004. *Teaching in English or English teaching?: on the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition.* Unpublished PhD thesis. Göteborg: Department of English, Göteborg University.

Thorén, Birger. 1976. *10000 ord för tio års engelska* [10000 words for 10 years of English]. Lund: Liber Låromedel.

Thornbury, Scott. 2004. *How to teach vocabulary*. (3rd impression). Essex: Pearson Education Limited.

Timischl, Wolfgang. 1996. *Qualitätssicherung: Statistische Methoden*. (2. Auflage). München: Carl Hanser Verlag.

Washburn, Linda. 1997. *English immersion in Sweden: a case study of Röllingby high school 1987-1989*. Unpublished PhD thesis. Stockholm: Department of English, Stockholm University.

Wesche, Marjorie; Paribakht, T. Sima. 1999. "Introduction to the special issue on incidental L2 vocabulary acquisition". *Studies in second language acquisition* 21, 175-180.

Widdowson, H.G. 1978. Teaching language as communication. Oxford: Oxford University Press.

Widdowson, H. G. 1996. *Linguistics*. Oxford: Oxford University Press.

Wilkins, David Arthur. 1972. *Linguistics in language teaching*. London: Edward Arnold.

Wilkins, Arthur. 1974. *Second language learning and teaching*. London: Edward Arnold.

Willis, Dave. 1990. *The lexical syllabus*. Collins COBUILD.

Wode, Henning. 1999. "Incidental vocabulary acquisition in the foreign language classroom". *Studies in second language acquisition* 21, 243-258.

Wolfsberger, Judith. 2007. *Frei geschrieben: Mut, Freiheit & Strategie für wissenschaftliche Abschlussarbeiten*. Wien: Böhlau Verlag.

Wood, R. 1976. "Sex differences in mathematics attainment at GCE Ordinary level". *Educational studies* 2, 141-160.

Wood, R. 1978. "Sex differences in answers to English language comprehension items". *Educational studies* 4, 157-165.

Woolard, George. 2000. *Lessons with laughter: photocopiable lessons for different levels*. Hove: Language Teaching Publications.

Wright, Daniel B. 1997. *Understanding statistics*. London: SAGE Publications.

List of Appendices

Appendix 1

Information sheets:

1. Information for students
2. Information for parents

INFORMATION
FÜR SCHÜLERINNEN UND SCHÜLER

Liebe Schülerin! Lieber Schüler!

Ich heiße Eva Maria Seregély und studiere an der Universität Wien Lehramt Englisch und Mathematik. Im Rahmen meiner Diplomarbeit möchte ich die Sprachkompetenz von Schülerinnen und Schülern der 7. Klasse AHS erheben. Zu diesem Zweck werde ich in deiner Klasse zuerst Fragebögen austeilen, welche deine Lese-, Fernseh- und Internetgewohnheiten bzw. etwaige Auslandsaufenthalte betreffen. In einem nächsten Schritt werdet ihr von mir verschiedene schriftliche Testaufgaben erhalten. Ich bitte ich dich, sowohl die Fragebögen als auch die Testaufgaben sorgfältig und nach bestem Wissen zu beantworten. Die Resultate dieser Aufgaben werden nur für meine Forschungsarbeit verwendet und haben keinerlei Einfluss auf deine Schulnoten.

Durch deine Mitarbeit leistest du einen wichtigen Beitrag für eine wissenschaftliche Untersuchung.

Herzlichen Dank für deine Unterstützung!

Eva Maria Seregély

INFORMATION FÜR ELTERN

Liebe Eltern!

Ich heiße Eva Maria Seregely und studiere an der Universität Wien Lehramt Englisch und Mathematik. Im Rahmen meiner Diplomarbeit möchte ich die Sprachkompetenz von Schülerinnen und Schülern der 7. Klasse AHS erheben. Zu diesem Zweck werde ich in der Klasse Ihrer Tochter/Ihres Sohnes Fragebögen austeilen, welche ihre/seine Lese-, Fernseh- und Internetgewohnheiten bzw. etwaige Auslandsaufenthalte betreffen. In einem nächsten Schritt wird Ihre Tochter/Ihr Sohn von mir verschiedene schriftliche Testaufgaben erhalten. Die Resultate dieser Aufgaben werden nur für meine Forschungsarbeit verwendet und haben keinerlei Einfluss auf die Schulnoten Ihres Kindes.

Ich bitte Sie, Ihr Einverständnis für diese Untersuchung zu geben.

Mit freundlichen Grüßen,

Eva Maria Seregély

Einverständniserklärung

Ich bin mit der Teilnahme meiner Tochter/meines Sohnes _____
_____ an einer Wortschatzuntersuchung einverstanden.

(Name, Klasse)

Wien, am _____ Unterschrift: _____

Appendix 2

Questionnaires:

1. Questionnaire – Traditional student
2. Questionnaire – VBS student
3. Questionnaire – Teacher
4. Questionnaire – Native speaker teacher

Bitte die Kästchen frei lassen!

NAME: _____ ☐

KLASSE: _____

1. Was ist deine Muttersprache? _____ ☐

2. Hast du überlegt, den VBS-Zweig zu besuchen?

   ☐ Ja, aber ich habe mich anders entschieden, weil _____

   _____ ☐

   ☐ Nein, weil _____

3. Warst du schon in einem englischsprachigen Land? ☐ Ja ☐

   ☐ Nein

   Wenn ja, in welchem Land/welchen Ländern? _____ ☐

   _____

   Wie lange warst du insgesamt im englischsprachigen Ausland?

   ☐ kürzer als 3 Monate

   ☐ 3 - 12 Monate ☐

   ☐ länger als 1 Jahr

4. Sprichst du außerhalb der Schule mit jemandem Englisch? ☐ Ja ☐

   ☐ Nein

   Wenn ja, mit wem (Mutter, Vater, Cousin/e, Freunde, Nachbarn …)?

   _____ ☐

5. Liest du in deiner Freizeit englischsprachige Literatur (Hausaufgaben ausgenommen)? ☐ Ja ☐

   ☐ Nein

   Wenn ja, wie oft? ☐ täglich

   ☐ mehrmals in der Woche ☐

   ☐ mehrmals im Monat

   ☐ gelegentlich

Welche Art von Literatur liest du? (Mehrfachnennungen möglich)

- ☐ Unterhaltungsliteratur
  (Krimis, Romane, Tagebücher etc.)
- ☐ Sachliteratur
  (Fachzeitschriften, Ratgeber, Gebrauchsanweisungen, Reiseführer, etc.)
- ☐ Wochenzeitschriften und Magazine
- ☐ Tageszeitungen
- ☐ Sonstiges: _____

6. Wie oft siehst du englischsprachige Fernsehsendungen mit deutschen Untertiteln?

   ☐ täglich
   ☐ mehrmals in der Woche
   ☐ mehrmals im Monat
   ☐ gelegentlich
   ☐ nie

7. Wie oft siehst du englischsprachige Fernsehsendungen ohne deutsche Untertitel?

   ☐ täglich
   ☐ mehrmals in der Woche
   ☐ mehrmals im Monat
   ☐ gelegentlich
   ☐ nie

8. Wie oft siehst du dir englischsprachige Filme (im Kino, auf DVD oder Video) an?

   ☐ mehrmals in der Woche
   ☐ mehrmals im Monat
   ☐ gelegentlich
   ☐ nie

9. Spielst du PC-, Tele-, Video- oder Rollenspiele oder sonstige Spiele, deren Anweisungen in englischer Sprache sind? ☐ Ja ☐ Nein

Wenn ja, wie oft? ☐ täglich
☐ mehrmals in der Woche
☐ mehrmals im Monat
☐ gelegentlich

Welche Art von Spielen? _____

10. Schreibst du in deiner Freizeit auf Englisch (Hausaufgaben ausgenommen)? ☐ Ja ☐ Nein

Wenn ja, wie oft? ☐ täglich
☐ mehrmals in der Woche
☐ mehrmals im Monat
☐ gelegentlich

In welchem Zusammenhang? ☐ Brief
☐ E-Mail
☐ Chat
☐ Bewerbungsschreiben
☐ Sonstiges: _____
_____

Wie oft surfst du im Internet? ☐ täglich
☐ mehrmals in der Woche
☐ mehrmals im Monat
☐ gelegentlich

… auf englischsprachigen Seiten? ☐ täglich
☐ mehrmals in der Woche
☐ mehrmals im Monat
☐ gelegentlich

11. Welche Ausbildung haben deine Eltern abgeschlossen?

(Bitte Zutreffendes ankreuzen!)

	Mutter	Vater
Pflichtschule		
Lehre		
Fachschule oder Berufsbildende Mittlere Schule (ohne Matura)		
Allgemeinbildende oder Berufsbildende Höhere Schule (mit Matura)		
Kolleg, Akademie		
Universitäts- oder Fachhochschulstudium		

12. Wie fühlst du dich, wenn du vor einer Gruppe von Leuten stehst und **Deutsch** sprichst? (Bitte Zutreffendes ankreuzen!)

☐ Sehr gut ☐ Gut ☐ Weniger gut ☐ Es ist mir unangenehm/peinlich.

13. Wie fühlst du dich, wenn du vor einer Gruppe von Leuten stehst und **Englisch** sprichst? (Bitte Zutreffendes ankreuzen!)

☐ Sehr gut ☐ Gut ☐ Weniger gut ☐ Es ist mir unangenehm/peinlich.

14. Wie würdest du die Verbesserung deiner Englischkenntnisse in folgenden Bereichen seit deinem Eintritt ins Gymnasium beurteilen? Bewerte nach Schulnoten!

Sprechen	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Schreiben	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Hörverständnis	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Leseverständnis	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Wortschatz/Vokabeln	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Grammatik	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

15. Welche Noten hattest du im letzten Zeugnis in folgenden Gegenständen?

    Deutsch: _____Englisch: _____ Mathematik: _____

16. Wie bist du mit der Wahl deines Schultyps/Unterrichtsschwerpunktes zufrieden?

Bewerte nach Schulnoten! ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

17. Warum hast du dich für diesen speziellen Schultyp/Unterrichts-schwerpunkt entschieden?

_____

_____

_____

18. Würdest du ihn anderen weiterempfehlen?

☐ Ja ☐ Nein

19. Was gefällt dir an diesem Schultyp/ Unterrichtsschwerpunkt am besten?

_____

_____

_____

20. Was gefällt dir dabei weniger gut?

_____

_____

_____

☺ HERZLICHEN DANK FÜR DEINE MITARBEIT!

Bitte die Kästchen frei lassen!

NAME: _____ ☐

KLASSE: _____

1. Was ist deine Muttersprache? _____ ☐

2. Warum hast du dich für den VBS-Zweig entschieden?

   _____

   _____ ☐

3. Warst du schon in einem englischsprachigen Land? ☐ Ja ☐ Nein ☐

   Wenn ja, in welchem Land/welchen Ländern? _____

   _____ ☐

   Wie lange warst du insgesamt im englischsprachigen Ausland?

   ☐ kürzer als 3 Monate
   ☐ 3 - 12 Monate
   ☐ länger als 1 Jahr ☐

4. Sprichst du außerhalb der Schule mit jemandem Englisch? ☐ Ja ☐ Nein ☐

   Wenn ja, mit wem (Mutter, Vater, Cousin/e, Freunde, Nachbarn …)?

   _____ ☐

5. Liest du in deiner Freizeit englischsprachige Literatur (Hausaufgaben ausgenommen)? ☐ Ja ☐ Nein ☐

   Wenn ja, wie oft?
   ☐ täglich
   ☐ mehrmals in der Woche ☐
   ☐ mehrmals im Monat
   ☐ gelegentlich

Welche Art von Literatur liest du? (Mehrfachnennungen möglich)

- ☐ Unterhaltungsliteratur
  (Krimis, Romane, Tagebücher etc.)
- ☐ Sachliteratur
  (Fachzeitschriften, Ratgeber, Gebrauchsanweisungen, Reiseführer, etc.)
- ☐ Wochenzeitschriften und Magazine
- ☐ Tageszeitungen
- ☐ Sonstiges: _____

6. Wie oft siehst du englischsprachige Fernsehsendungen mit deutschen Untertiteln?

   ☐ täglich
   ☐ mehrmals in der Woche
   ☐ mehrmals im Monat
   ☐ gelegentlich
   ☐ nie

7. Wie oft siehst du englischsprachige Fernsehsendungen ohne deutsche Untertitel?

   ☐ täglich
   ☐ mehrmals in der Woche
   ☐ mehrmals im Monat
   ☐ gelegentlich
   ☐ nie

8. Wie oft siehst du dir englischsprachige Filme (im Kino, auf DVD oder Video) an?

   ☐ mehrmals in der Woche
   ☐ mehrmals im Monat
   ☐ gelegentlich
   ☐ nie

9. Spielst du PC-, Tele-, Video- oder Rollenspiele oder sonstige Spiele, deren Anweisungen in englischer Sprache sind? ☐ Ja ☐ Nein

Wenn ja, wie oft? ☐ täglich
☐ mehrmals in der Woche
☐ mehrmals im Monat
☐ gelegentlich

Welche Art von Spielen? _____

10. Schreibst du in deiner Freizeit auf Englisch (Hausaufgaben ausgenommen)? ☐ Ja ☐ Nein

Wenn ja, wie oft? ☐ täglich
☐ mehrmals in der Woche
☐ mehrmals im Monat
☐ gelegentlich

In welchem Zusammenhang? ☐ Brief
☐ E-Mail
☐ Chat
☐ Bewerbungsschreiben
☐ Sonstiges: _____
_____

Wie oft surfst du im Internet? ☐ täglich
☐ mehrmals in der Woche
☐ mehrmals im Monat
☐ gelegentlich

… auf englischsprachigen Seiten? ☐ täglich
☐ mehrmals in der Woche
☐ mehrmals im Monat
☐ gelegentlich

11. Welche Ausbildung haben deine Eltern abgeschlossen?

(Bitte Zutreffendes ankreuzen!)

	Mutter	Vater
Pflichtschule		
Lehre		
Fachschule oder Berufsbildende Mittlere Schule (ohne Matura)		
Allgemeinbildende oder Berufsbildende Höhere Schule (mit Matura)		
Kolleg, Akademie		
Universitäts- oder Fachhochschulstudium		

12. Wie fühlst du dich, wenn du vor einer Gruppe von Leuten stehst und **Deutsch** sprichst? (Bitte Zutreffendes ankreuzen!)

☐ Sehr gut ☐ Gut ☐ Weniger gut ☐ Es ist mir unangenehm/peinlich.

13. Wie fühlst du dich, wenn du vor einer Gruppe von Leuten stehst und **Englisch** sprichst? (Bitte Zutreffendes ankreuzen!)

☐ Sehr gut ☐ Gut ☐ Weniger gut ☐ Es ist mir unangenehm/peinlich.

14. Wie würdest du die Verbesserung deiner Englischkenntnisse in folgenden Bereichen seit deinem Eintritt ins Gymnasium beurteilen? Bewerte nach Schulnoten!

Sprechen	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Schreiben	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Hörverständnis	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Leseverständnis	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Wortschatz/Vokabeln	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Grammatik	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

15. Welche Noten hattest du im letzten Zeugnis in folgenden Gegenständen?

    Deutsch: _____Englisch: _____ Mathematik: _____

16. Wie bist du mit der Wahl deines Schultyps/Unterrichtsschwerpunktes zufrieden?

Bewerte nach Schulnoten! ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

17. Warum hast du dich für diesen speziellen Schultyp/Unterrichts-schwerpunkt entschieden?

_____

_____

_____

18. Würdest du ihn anderen weiterempfehlen?

☐ Ja ☐ Nein

19. Was gefällt dir an diesem Schultyp/ Unterrichtsschwerpunkt am besten?

_____

_____

_____

20. Was gefällt dir dabei weniger gut?

_____

_____

_____

☺ HERZLICHEN DANK FÜR DEINE MITARBEIT!

Bitte die Kästchen frei lassen!

Geschlecht: □ weiblich □ männlich

Ich unterrichte bilingual. ☐ ja ☐ nein

Falls Sie nicht bilingual unterrichten, werden Sie einige Antworten auslassen. Should you prefer to answer in English, please feel free to do so!

1. Wie viele Jahre unterrichten Sie? Bitte kreuzen Sie an:

☐ unter 10 Jahre ☐ 10 - 20 Jahre ☐ über 20 Jahre

2. Welche Gegenstände unterrichten Sie? _____

_____

3. Was ist Ihre Muttersprache? _____

4. Welche Fremdsprachen sprechen Sie und auf welchem Niveau?
(beginner - intermediate – advanced – native speaker level)

_____

_____

5. Welche formelle Ausbildung haben Sie in Englisch absolviert?

☐ AHS/BHS
☐ Lehramt Englisch
☐ Sonstiges: _____

6. Haben Sie eine zusätzliche Fortbildung für bilingualen Unterricht erhalten? ☐ Ja ☐ Nein

Wenn ja, in welchem Ausmaß erfolgte diese Fortbildung?

ca. _____Einheiten

In welcher Form erfolgte diese Fortbildung (inkl. private Fortbildung)?

_____

_____

7. Haben Sie im Ausland gearbeitet oder studiert? ☐ Ja ☐ Nein

   Wenn ja, in welchem Land/welchen Ländern? _____

   _____

   Wie lange? _____

8. Haben Sie sich längere Zeit in einem englischsprachigen Land aufgehalten? ☐ Ja ☐ Nein

   Wenn ja, in welchem Land/welchen Ländern? _____

   _____

   Wie lange? _____

9. Wie lange unterrichten Sie bereits bilingual?

   _____

10. In wie vielen Unterrichtsstunden pro Klasse/Woche unterrichten Sie bilingual? _____

11. In welchen Gegenständen unterrichten Sie bilingual?

    _____

12. Unterrichten Sie im Team mit einem Native Speaker?

    ☐ Ja ☐ Nein

    Wenn ja, wie oft? ca. _____Einheiten pro Woche/Monat

    Welche Vorteile haben Sie durch Teamarbeit und durch die Zusammenarbeit mit dem Native Speaker? _____

    _____

    _____

    Welche Aufgaben erfüllt der Native Speaker im Schulzweig VBS?

    _____

    _____

Fragen 13 – 22:

	Bitte Zutreffendes ankreuzen! (BU = bilingualer Unterricht)	trifft zu	trifft eher zu	trifft eher nicht zu	trifft nicht zu
13.	BU bereichert den methodisch-didaktischen Bereich des Fachunterrichts.				
14.	Es gibt genügend Fortbildungsangebote für BU.				
15.	Es gibt genügend geeignete Materialien für BU.				
16.	Die SchülerInnen haben durch BU Defizite im Fachbereich.				
17.	BU verbessert die Leistungen in den Fachgegenständen.				
18.	BU hat eine positive Auswirkung auf die Motivation der SchülerInnen.				
19.	BU hat eine positive Auswirkung auf meine persönliche Motivation.				
20.	Ich korrigiere Sprachfehler im Fachgegenstand.				
21.	Ich führe im Fachgegenstand regelmäßig Wortschatztests durch.				
22.	Die SchülerInnen verbessern durch BU ihre Fremdsprachenkompetenz deutlich.				

23. Wie würden Sie die Verbesserung der Fremdsprachenkompetenz der SchülerInnen durch bilingualen Unterricht in folgenden Bereichen beurteilen? Bewerten Sie nach Schulnoten!

Sprechen	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Schreiben	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Hörverständnis	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Leseverständnis	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Wortschatz/Vokabeln	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Grammatik	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

24. Welche Motive führten Sie persönlich zum Einsatz von bilingualem Unterricht? _____

_____ ☐

_____

25. Welche Ziele verfolgen Sie persönlich mit bilingualem Unterricht?

_____

_____ ☐

_____

26. Welches Niveau sollten SchülerInnen am Ende der bilingualen Oberstufe Ihrer Meinung nach in der Fremdsprache erreicht haben? _ ☐

_____

_____

27. Nennen Sie weitere Erfahrungen, Verbesserungsvorschläge, Vorteile oder Nachteile von bilingualem Unterricht/des VBS-Projekts!

_____

_____ ☐

_____

28. Welche Maßnahmen könnten Ihrer Einschätzung nach bilingualen Unterricht/das VBS-Projekt in der Zukunft fördern/verbessern?

_____

_____ ☐

_____

29. Wie wird sich bilingualer Unterricht Ihrer Einschätzung nach in der Zukunft entwickeln? _____

_____ ☐

_____

☺ HERZLICHEN DANK FÜR IHRE MITARBEIT!

QUESTIONNAIRE – Native Speakers

Please leave the boxes empty!

Gender: □ female □ male

Nationality/-ies: _____

1. How long have you been teaching?

☐ less than 5 years ☐ 5 - 10 years ☐ more than 10 years

2. What kind of training do you have? Do you hold any teaching degrees?

_____

3. Which subjects do you teach? _____

4. Have you received any additional training for bilingual instruction?

☐ Yes ☐ No

If yes, how much training have you had? approx._____units

What kind of training have you had (including private courses, self-study etc.)? _____

_____

5. Which foreign languages do you speak and at what level? (beginner - intermediate – advanced – native speaker level)

_____

_____

6. Have you worked or studied abroad? ☐ Yes ☐ No

If yes, in which country/-ies? _____

_____

How long? _____

7. How long have you been teaching bilingual lessons (in Austria)?

_____

8. In how many classes do you teach bilingual lessons?

_____

9. How many hours per week do you use bilingual instruction?

_____

10. In which subjects do you use bilingual instruction?

_____

11. Do you team-teach with Austrian subject teachers?

☐ Yes ☐ No

If yes, how often? approx. _____units per week/month

What is your collaboration with the subject teachers and/or English teachers like (advantages/disadvantages)? _____

_____

_____

12. Which tasks do you have within the VBS project?

_____

_____

_____

13. What are the main teaching methods/strategies you are using in your bilingual lessons? _____

_____

_____

Questions 14 – 23:

	Please mark the true statements! (BI = bilingual instruction)	true	mainly true	partly true	not true	
14.	BI makes the subject lessons richer.					☐
15.	There is sufficient in-service training for native speakers teaching BI.					☐
16.	There are sufficient teaching materials for BI.					☐
17.	The students seem to miss out on subject knowledge due to BI.					☐
18.	The students seem to improve their subject knowledge through BI.					☐
19.	BI increases the students' motivation.					☐
20.	BI increases my own motivation.					☐
21.	I correct language mistakes in the subject lessons.					☐
22.	I regularly conduct vocabulary tests in the subject lessons.					☐
23.	Students have improved their English clearly since I have been working with them.					☐

24. In terms of foreign language competence, how would you estimate the students' improvement through bilingual instruction in the following areas/skills? Please use the Austrian grading system for your estimation!

Speaking	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	
Writing	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	
Listening comprehension	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐
Reading comprehension	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	
Vocabulary	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	
Grammar	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	

25. What are your personal motives for working as a native speaker teacher of VBS? _____

_____ ☐

_____

26. What are your personal aims in your work with bilingual instruction?

_____

_____ ☐

_____

27. In your opinion, what level of English should the students reach at the end of the bilingual Upper Secondary? _____ ☐

_____

_____

28. Would you like to add any further comments, suggestions for improvements, advantages or disadvantages concerning bilingual instruction/the VBS project? _____

_____ ☐

_____

_____

29. Which activities and/or regulations could improve bilingual instruction/the VBS project in the future? _____

_____ ☐

_____

30. In your opinion, how will bilingual instruction develop in the future?

_____

_____ ☐

_____

☺ THANK YOU VERY MUCH FOR YOUR CO-OPERATION!

Appendix 3

Lexical tests:

1. Self-report test
2. Words in context test (article + 2 test sheets)
3. Multiple choice test
4. Cloze test I
5. Cloze test II

Die im Folgenden dargestellten Testaufgaben basieren auf einer Untersuchung der Wortschatzkompetenz von schwedischen Schülerinnen und Schülern, die im Jahr 2004 von der Universität Göteborg durchgeführt wurde. Mein besonderer Dank gilt Frau Dr. Liss Kerstin Sylvén, die ihre Unterlagen freundlicherweise für die vorliegende Replikationsstudie zur Verfügung gestellt hat.

NAME: _____

CLASS: _____

1. SELF REPORT

Use the column which best corresponds to your knowledge of the word. The columns are the following:

A. I do not remember having seen this word before.
B. I have seen this word before, but I do not know what it means.
C. I have seen this word and think it means (*write a synonym or translation*)
D. I know this word. It means (*write a synonym or translation*)
E. I can use this word in a sentence (*write a synonym or translation in column D and then go to column E and write a full sentence which shows how the word is typically used*)

	A	B	C	D	E	Do not use this column!
1. Adjacent						
2. To adjust						
3. An attire						
4. A bargain						
5. A calamity						
6. A clue						
7. To commence						
8. Crucial						
9. To deviate						
10. To disguise						

Do not use this column!

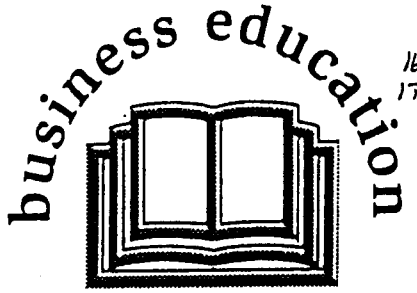
	A	B	C	D	E	
11. An editorial						
12. To emerge						
13. To exaggerate						
14. A fraud						
15. Furthermore						
16. Hostile						
17. To illuminate						
18. In conclusion						
19. Incidentally						
20. Ingenuity						
21. To merge						
22. A mortgage						
23. Notwithstanding						
24. A poll						
25. To precede						
26. A prediction						
27. A rate						
28. A refuge						
29. To rely						
30. Repulsive						

AN AMERICAN SUCCESS STORY

Increasingly, MBAs are the ticket to the corporate suite.*

There is no denying the success of the MBA. The World Resource Institute estimates that there are currently about 100,000 students graduating each year with a master of business administration degree — an astounding quarter of all postgraduate degrees granted in the United States. The Yale School of Management calculates that applications to business schools were up by an average of 2.4 percent last year.

The popularity of the MBA is closely linked to the business cycle. According to the 12th edition of "Which MBA?" (published by the Economist Intelligence Unit and Pearson Education), by the late 1980s, MBA programs were "basking in the success engendered by the strong economic performance of that decade." But by the early '90s, interest was flagging as the worldwide recession took its toll on recruitment.



Golden asset

Released last October, before the talk of the U.S. slowdown really took hold, the report said, "The beginning of the 21st century appears to be another golden age for MBA students." While the jury is still out on that prediction, it would be hard to argue with the other findings of the research: "The degree itself has also become more credible and more important."

A master's in business is increasingly a prerequisite for many top management jobs. According to "Which MBA?" 21 percent of global chief executive officers have earned one, a figure that rises to 39 percent of CEOs in North America.

A school's overall reputation continues to be the most important criterion, both for those seeking enrollment and those who will eventually hire them. Other important criteria include program content, location and quality of teaching.

Among the business programs with the best reputations in recent years have been Harvard Business School and the Wharton School of the University of Pennsylvania — despite their dissimilar styles. Harvard adheres to the case study as its sole teaching method, while Wharton is not tied to any single teaching practice. Wharton instructors break down their teaching into separate concepts and skills, imparted face-to-face or via technology.

Teaching methods aside, the two schools do have a number of philosophical similarities. Neither has joined the rush to form comprehensive alliances with other business schools and profit-making educational organizations. Wharton recently announced, however, that, like Harvard, it will soon establish a presence on the U.S. West Coast. Harvard set up its California Research Center in Menlo Park in 1997. Wharton West, announced last December, will also offer student programs and faculty research opportunities on the opposite side of North America.

Harvard is one of a number of schools that have made sweeping changes in their curricula in recent years. The graduating class of 2000 was the last required to take general management, once considered the school's signature course. Instead, students must now enroll in a class called "The Entrepreneurial Manager." Faculty members are quick to point out that this new course does not ignore the study of large companies, but rather encourages Harvard-trained managers to be more entrepreneurial, regardless of company size.

Canadian business schools are also raising their profiles. The Richard Ivey School of Business at the University of Western Ontario was recently ranked among the top five non-U.S. schools.

E-investments

E-business education is another area that business schools have been developing over the past few years. Last year, Stanford announced the creation of a $20 million Center for Electronic Business and Commerce as part of the California institution's aim to be the world leader in e-commerce research. MIT's Sloan School of Management has introduced a "Management Track in E-Business" concentration, and Thunderbird has formulated a new curriculum for its Master of International Management, with a heavy emphasis on e-commerce.

No one knows for sure if these investments in e-commerce are entirely well-founded, but many insist that the current economic conditions should be put into perspective.

Says Trent Anderson, vice president of education for Kaplan Inc., a career-services company: "The Internet has had a profound effect on business, even in light of the current dot-com downturn. The investments that U.S. schools made in revamping their curricula are invaluable for ensuring that their graduates are prepared to succeed in any business. Thus, it may not be necessary to offer a class on Internet marketing, but it will be essential to teach students about e-mail marketing as part of a larger discussion of marketing strategy."

Staying on top of technology is another priority for business schools. Duke University's Fuqua School of Business has established the Next Generation Client Computing project to identify and influence the design of computing products that could enhance its learning environment. ●

* MBA = Master of Business Administration degree

NAME: _____

CLASS: _____

2. WORDS IN CONTEXT

Read the article and explain the following words as they are used in the text. You can either translate the word into German, give a synonym in English, or explain the meaning in German or English.

		Do not use this column!
1. deny		
2. graduate		
3. astounding		
4. grant		
5. average		
6. performance		
7. recruitment		
8. age		
9. prediction		
10. findings		
11. prerequisite		
12. reputation		
13. eventually		
14. hire		
15. content		

Do not use this column!

16. adhere		
17. sole		
18. similarities		
19. announce		
20. presence		

21. opportunity		
22. opposite		
23. enroll		
24. encourage		
25. research		

26. entirely		
27. current		
28. revamp		
29. essential		
30. enhance		

NAME: _____

CLASS: _____

3. MULTIPLE CHOICE

Select *the best* alternative synonym to each one of the underlined words. Transfer your answer to the column indicated.

	A	B	C	D	E	ANSWER HERE ⇩	*Do not use this column*
1. The document was abbreviated.	burnt	lengthened	recycled	shortened	translated		
2. I will abstain from further comments.	will add	will consider	wont make	will remove	will repeat		
3. The authorities cancelled the demonstration.	government	officers	protesters	shop-owners	writers		
4. This is a dish for carnivores.	animals	athletes	ladies	meat-eaters	vegetarians		
5. The headmaster condemned their behaviour.	criticized	judged	liked	praised	reported		
6. This disease is not contagious.	appreciated	dangerous	deadly	infectious	known		
7. Don't you ever contradict me again!	come near	hurt	imitate	lie about	oppose		
8. Witnesses say the firing was deliberate.	accidental	accurate	dangerous	deadly	intentional		
9. It is my destiny one day to be king.	dream	fate	idea	plan	wish		
10. She kept a diary for memory's sake.	handkerchief	flower	note-book	photograph	secret		
11. Federalism is intended to diminish the power of the central state.	concentrate	control	increase	praise	reduce		
12. This is an elaborate research project.	detailed	easy	long-term	time-consuming	tiring		
13. He had formerly been in Norway.	fortunately	in the military	in the past	lately	normally		
14. He inserted a critical comment.	insisted on	missed	put in	removed	wanted		
15. Likewise, we will change the program.	gradually	hopefully	luckily	probably	similarly		

ANSWER HERE

Do not use this column!

16. It was an example of a fast-moving narrative.	car	career	lie	song	story	
17. It is, nevertheless, a necessary decision.	however	not at all	obviously	of course	only	
18. Our main objective was the recovery of the child.	article	item	purpose	thought	wish	
19. Elsa was now less anxious about her offspring than she had once been.	career	children	furniture	relatives	results	
20. I presume you are here on business.	have been told	hope	know	suppose	understand	

21. The results provoked criticism.	avoided	caused	resembled	saved	stopped	
22. The government rejects the idea.	aims at	does not accept	presents	supports	wants	
23. This can scarcely be coincidence.	certainly	hardly	never	obviously	unfortunately	
24. After two sessions, the council has failed to reach agreement.	disputes	emergencies	games	meetings	trials	
25. Her spouse was a dentist.	cousin	friend	husband	profession	son	

26. Germany has a surplus of teachers.	excellent	just enough	too few	too many	unqualified	
27. Her untimely death shocked the entire nation.	dramatic	painful	too early	unexpected	violent	
28. Mr Werner unveiled his new strategy.	denied	described	forgot	hid	uncovered	
29. The vehicle did not have the capacity to take the whole group.	area	bus	company	meeting room	sports club	
30. The accident took place in the vicinity of the school.	cafeteria	entrance	front	large hall	neighbourhood	

NAME: _____

CLASS: _____

4. CLOZE TEST I

Complete the following sentences. Please observe that only one word is missing from each sentence.

Do not use this column!

1. Like you, I'm _____ a loss to explain the sudden fall in share prices. I have absolutely no idea what can have caused it. ☐
2. She just sits in front of the TV all day. She's such a _____ potato. ☐
3. I've done this several times so for me it's a piece of _____. ☐
4. If you want a flat in the centre of the city you have to pay through the _____ for it. ☐
5. She wasn't too worried about the incident. She took it in her _____. ☐
6. You didn't think I was being serious, did you Brian? It was a joke! I was pulling your _____, that's all! ☐
7. Although it has its faults, by and _____ Britain is a good country to live in. ☐
8. He couldn't keep the secret, he spilled the _____. ☐
9. It's true, I tell you! I got it straight from the _____ mouth. ☐
10. What does this mean? I can't make head or _____ of it. ☐
11. Speak up! I'm a bit _____ of hearing, you see. ☐
12. The accident was caused by a taxi driver who _____ the traffic lights. ☐
13. This is Mrs Brightwell. She is _____ charge of marketing. ☐
14. Getting a problem off your _____ is the first step to solving it. ☐
15. I bought a computer last year, but I've had nothing but trouble with it. As far as I'm concerned it was £800 down the _____. ☐

Do not use this colum!

16. Tom just kept on talking, but as Susan was in a hurry she had to cut him _____ . ☐

17. Pick whatever you want from the menu and don't worry about the prices. I'll _____ the bill! ☐

18. Thank you for attending the interview, Mr Blake. You'll be hearing from us in _____course, probably at the end of next week ☐

19. I agree, that's a bit of a _____ shot, but we're desperate and have to do something to try and save the company. ☐

20. They say that two is company but three is a _____ , but after years of observing my parents together, I have my doubts. ☐

21. Do you know what they plan for us tomorrow?
No, your _____ is as good as mine. ☐

22. Thanks for the invitation, but I'll have to take a _____ check on it. ☐

23. The viewers were up in _____ when the television station announced it was going to change the time of the evening news from 9 o'clock to 9.30. ☐

24. Every morning, rain or _____ , he goes out jogging. ☐

25. I told him he was making a mistake to accuse me, he was simply _____ up the wrong tree! ☐

26. I wasn't expecting a letter from her; it arrived quite out of the _____ . ☐

27. Being retired, he suddenly found himself with lots of time on his _____ but with little to do to occupy it. ☐

28. It isn't true that I hate pop music; on the _____ , I like it very much. ☐

29. I have to write an essay about history and since you know a lot about it but I don't, do you mind if I pick your _____ ? ☐

30. All her answers were correct, so she passed the exam with _____ colours. ☐

☐

NAME: _____

CLASS: _____

5. CLOZE TEST II

Complete the following sentences. Please observe that only one word is missing from each sentence.

Do not use this column!

1. Peter was born and brought up in Hastings and knows it like the back of his _____ . ☐

2. I was already fed up with the job, but when the boss walked into my office and told me he expected me to work overtime that was the last _____. I quit. ☐

3. He tried in _____ to catch somebody's attention. Nobody heard him. ☐

4. If you want a flat in the centre of the city you have to pay through the _____ for it. ☐

5. "And you finish work for the day at 5.30 p.m.?" "Yes, in _____, but seldom in practice." ☐

6. You didn't think I was being serious, did you Brian? It was a joke! I was pulling your _____, that's all! ☐

7. Although it has its faults, by and _____ Britain is a good country to live in. ☐

8. Was it really six years ago that we last met?
   Yes it was.
   Well, I never! Time certainly _____, doesn't it? ☐

9. It's true, I tell you! I got it from the _____ mouth. ☐

10. What does this say? I can't make head nor _____ of it. ☐

11. Speak up! I'm a bit _____ of hearing you see. ☐

12. I was only sixteen when I first fell head over _____ in love. ☐

13. She was so tired last night that she slept like a _____ until 10 o'clock this morning. ☐

14. "Is the dress too big?" "No, not at all. It fits like a _____ ." ☐

15. I bought a computer last year, but I've had nothing but trouble with it. As far as I'm concerned it was £800 down the _____. ☐

Do not use this column

16. She had a sweet _____ so she couldn't resist buying chocolates and cream cakes. ☐

17. None of the students liked Mr Baker. In fact, everyone was glad to see the _____of him when he left to teach in Italy. ☐

18. Thank you for attending the interview, Mr Blake. You'll be hearing from us in _____ course, probably at the end of next week. ☐

19. The cruel way some owners treat their pets makes my _____ boil. ☐

20. You'd better get some sleep, since you have to get up _____the crack of dawn tomorrow. ☐

21. We've worked enough, so let's call it a _____ and go home. ☐

22. I had to leave the clothes on the washing line. They were frozen _____ so I couldn't fold them into the basket. ☐

23. For crying out _____ Duke, could you be a little more elusive for a change? ☐

24. Securely placed in the middle-class world of a successful businessman, he is suddenly brought face to _____ with the realization that there are more things to life than making money. ☐

25. Actually, I have very little food in the house, so it's just as _____ that you're not hungry. ☐

26. I wasn't expecting a letter from her; it arrived quite out of the _____ . ☐

27. Now children, I want you to be all eyes and _____for the first part of the lesson and then we'll have questions afterwards. ☐

28. It isn't true that I hate pop music; on the _____ , I like it very much. ☐

29. "I'm a real intellectual-type guy, Tracy", James joked. "Oh, give me a _____", Tracy replied. ☐

30. I can't watch horror movies. They give me the _____. ☐

☐

Appendix 4

Abstracts

1. Abstract in English
2. Abstract in German

Curriculum Vitae

Abstract – English

Content and Language Integrated Learning (CLIL), an educational context in which a foreign language is used completely or partially as the medium of instruction in the teaching of subjects, such as History and Biology, has enjoyed increasing popularity in Austria in the last 10-15 years. The main aim of the method is to enhance students' linguistic skills due to a higher amount of target language exposure. Based on a Swedish study conducted in the years 1999-2004, the present thesis investigates whether CLIL learners have a larger and more complex English vocabulary than students taught in the traditional way. For this purpose, 33 students (21 CLIL, 12 traditional) of a Viennese grammar school have been involved in an empirical study, comprising a battery of five different lexical tests. In addition, questionnaires concerning personal background have been distributed among all students and the teachers of the CLIL group. The results show that the CLIL students clearly outperform their traditional peers, yet the degree of superiority depends on the respective test type used. Furthermore, there is a stronger tendency among CLIL learners to use English for various activities – above all, spoken and written correspondence – outside school. Besides, CLIL students are more likely to come from well-educated family backgrounds, and rate their own linguistic competence considerably higher than traditional students do. Overall, the CLIL group's lexical supremacy cannot only be traced to the method alone. Rather, CLIL is closely linked to a variety of other factors, which, in sum, have lead to outstanding scores.

Abstract – German

Content and Language Integrated Learning (CLIL), ein Unterrichtskonzept, im Zuge dessen Fächer wie etwa Geschichte oder Biologie gänzlich oder teilweise in einer Fremdsprache unterrichtet werden, hat sich in Österreich in den letzten 10-15 Jahren einer immer größer werdenden Beliebtheit erfreut. Hauptziel dieser Methode ist es, die Sprachkompetenz von SchülerInnen durch die verstärkte Auseinandersetzung mit der Zielsprache zu verbessern. Basierend auf einer schwedischen Studie, erstellt in den Jahren 1999-2004, geht die vorliegende Diplomarbeit der Frage nach, ob CLIL-Lernende über ein größeres und komplexeres englisches Vokabular verfügen, als SchülerInnen, die auf traditionellem Wege unterrichtet werden. Zu diesem Zweck wurde eine empirische Untersuchung, bestehend aus einer Testbatterie von fünf verschiedenen lexikalischen Tests, mit 33 Schülerinnen und Schülern (21 CLIL- und 12 RegelschülerInnen) eines Wiener Gymnasiums durchgeführt. Zusätzlich wurden an alle SchülerInnen und die LehrerInnen der CLIL-Gruppe Fragebögen hinsichtlich persönlicher Hintergründe verteilt. Die Untersuchungsergebnisse zeigen, dass die Leistungen der CLIL-SchülerInnen deutlich jene ihrer traditionell-unterrichteten AltersgenossInnen übertreffen, jedoch variiert der Grad der Überlegenheit in Abhängigkeit vom jeweiligen Testtyp. Weiters tendieren CLIL-Lernende stärker dazu, die englische Sprache für verschiedene Aktivitäten – vor Allem mündliche und schriftliche Korrespondenz – außerhalb des schulischen Umfelds zu verwenden. Darüber hinaus stammen CLIL-SchülerInnen eher aus einem akademisch gebildeten Elternhaus und schätzen ihre eigenen sprachlichen Fähigkeiten wesentlich höher ein, als dies RegelschülerInnen tun. Insgesamt lässt sich die höhere lexikalische Kompetenz der CLIL-Gruppe allerdings nicht nur auf die Methode selbst zurückführen. Vielmehr ist CLIL eng mit einer Vielzahl anderer Faktoren verknüpft, die in ihrer Gesamtheit zu den überragenden Testresultaten beigetragen haben.

Curriculum Vitae

Angaben zur Person	
Name	Eva Maria Seregély
Staatsangehörigkeit	Österreich
Geburtsdatum und -ort	16.05.1984, Oberwart
Ausbildung	
Daten	seit Oktober 2002
Bildungseinrichtung	Universität Wien
Studienrichtung	Lehramtsstudium Unterrichtsfach Englisch Unterrichtsfach Mathematik
Zusatzqualifikation	Certificate in Teaching English for Specific Purposes (CerTESP)
Schwerpunkte und Studieninteressen	Applied Linguistics, Content and Language Integrated Learning (CLIL), English for Specific Purposes (ESP), Lexical learning, Angewandte Mathematik, Statistik
Daten	August 2004 (einmonatiger Kurs)
Bildungseinrichtung	Malvern House, London
Schwerpunkte	General English, English conversation
Daten	1994 - 2002
Bildungseinrichtung	Bundesgymnasium Oberschützen
Abschluss	Reifeprüfung mit ausgezeichnetem Erfolg
Schwerpunkt	Fremdsprachen (Latein, Spanisch, Französisch)
Besondere Tätigkeiten	Mitarbeit an einem dreiwöchigen Entwicklungshilfeprojekt (im Rahmen des Spanischunterrichts) mit Waisenkindern in Bogotá und Villavicencio, Kolumbien, September 2000
Berufserfahrung	
Daten	seit September 2006
Beruf oder Funktion	Lehrerin
Zuständigkeiten	Nachhilfeunterricht in Englisch und Mathematik für alle Schulstufen Englischunterricht für Volks- und Vorschulkinder administrative Tätigkeiten
Name des Arbeitgebers	lernaktiv das Lern- und Sprachinstitut, Wien XII und Purkersdorf
Daten	August 2005
Beruf oder Funktion	Betreuerin/Lehrerin
Zuständigkeiten	Unterrichtsvorbereitung in Deutsch, Mathematik und Englisch, Freizeitgestaltung
Name des Arbeitgebers	Österreichisches Rotes Kreuz – Bereich Gesundheit und Sozialer Dienst, Bezirksstelle Purkersdorf
Daten	2002-2005 (Sommermonate)
Beruf oder Funktion	Museums- und Fremdenführerin
Zuständigkeiten	Einzel- und Gruppenführungen, administrative Tätigkeiten
Name des Arbeitgebers	Stadt- und Tuchmachermuseum, Stadtgemeinde Pinkafeld
Sprachkenntnisse	Deutsch (Muttersprache), Englisch (C1-C2), Spanisch (B1), Französisch (A1)