

Assessing Language or Content?

A comparative study of the assessment practices in three Swedish
upper secondary CLIL schools

Helena Reierstam



**DEPARTMENT OF EDUCATION
AND SPECIAL EDUCATION**

© HELENA REIERSTAM, 2015

Licentiate thesis in Subject Matter Education at the Department of Education and Special Education, Faculty of Education, University of Gothenburg.

The licentiate thesis is available for full text download at Gothenburg University Publications Electronic Archive (GUPEA):

<http://hdl.handle.net/2077/40701>

This licentiate thesis has been carried out within the framework of the Graduate school in Foreign Language Education “De främmande språkens didaktik” (FRAM). The Graduate School, leading to a licentiate degree, is a collaboration between the universities of Gothenburg, Lund, Stockholm and Linnaeus University, and is funded by the Swedish Research Council (project number 729-2011-5277).

Abstract

Title: Assessing Language or Content? A comparative study of the assessment practices in three Swedish upper secondary CLIL schools.

Author: Helena Reierstam

Language: English with a Swedish summary

GUPEA: <http://hdl.handle.net>

Keywords: assessment, CLIL, written tests, EFL, biology, history, upper secondary school, teacher cognition

The present study investigates teachers' assessment practices in a Swedish Content and Language Integrated Learning (CLIL) context at three upper secondary schools. The aim is to explore if, and, if so, how and on what grounds the assessment practices differ in the two subject content courses biology and history due to the use of English as the language of instruction. A second aim concerns if, and, if so, how, the course content and the assessment tools in the English language (EFL) courses are affected due to the use of English in other courses. The focus is on teachers' perceptions and practices. A total of 12 teachers participated in the study: 6 subject content teachers, 4 CLIL and 2 non-CLIL, and 6 EFL teachers.

The data consists of teacher interviews, a questionnaire and assessment samples. The teacher responses and assessment samples were analyzed in relation to national course goals and written assessment features. A third objective of the study is to examine if there are common cross-disciplinary features as regards language, content and form in the tests. Students' ability to show content knowledge in a foreign language has been identified as a problematic area in CLIL assessment. So, test items were analyzed in relation to cognitive and linguistic demands, triggered by question formulations.

The results indicate that CLIL does not have an effect on teachers' assessment practices. Differences found rather seem to relate to individual preferences or teachers' perceptions of the discipline. The impact of CLIL on the EFL courses is insignificant. Some cross-disciplinary common features were identified in assessment of written production. In conclusion, the analysis suggests the development of CLIL-specific cross-disciplinary assessment guidelines, taking both language and content into account in relation to written disciplinary genres.

Contents

Acknowledgements

1.	INTRODUCTION	11
1.1	Background.....	11
1.2	Aims and research questions.....	13
1.3	Empirical research framework.....	17
1.4	Outline of thesis.....	17
2.	CLIL AND INTERDISCIPLINARY TEACHING	19
2.1	Introduction.....	19
2.2	Different variants: an overview	19
2.3	CLIL, discourse and practice	21
2.4	Assessment in CLIL	24
2.5	CLIL in Sweden	25
2.6	Summary.....	28
3.	ASSESSMENT'.....	31
3.1	Introduction.....	31
3.2	Historical and contextual impact.....	31
3.3	The Swedish context	32
3.4	Assessment literacy.....	34
3.5	Assessment in the subject disciplines	36
3.5.1	Assessing language	37
3.5.2	Assessing biology.....	38
3.5.3	Assessing history.....	40
3.6	Written assessment	41
3.6.1	Question tests.....	41
3.6.2	Production tests	42
3.7	Validity in assessment	44

3.7.1 A validation chain model	45
3.8 Summary.....	49
4. LANGUAGE AND CONTENT IN THE DISCIPLINES.	51
4.1 Introduction.....	51
4.2 Linguistic and cognitive skills	51
4.2.1 Descriptor words in the CEFR.....	52
4.2.2 Course goals	54
4.3 Learning language	56
4.3.1 Form versus function.....	57
4.3.2 Types of language	58
4.3.3 Translanguaging and interlanguages	60
4.4 Learning content.....	61
4.4.1 Knowledge structure and epistemology	61
4.4.2 Thinking skills and Bloom’s revised taxonomy.....	62
4.5 Learning language and content.....	63
4.5.1 Genre pedagogy across the curriculum.....	63
4.5.2 Cummins’ quadrant.....	64
4.5.3 Language in all the disciplines	66
4.6 Summary.....	68
5. MATERIAL AND METHODS.....	71
5.1 Introduction.....	71
5.2 Selection and data collection.....	71
5.3 The schools and the participants.....	72
5.4 Tools and analyses	75
5.4.1 The interviews	75
5.4.2 Document analysis	78
5.4.3 The questionnaire	82
5.5 Ethical concerns.....	83
5.6 Validity of the study	84

5.7 Summary.....	84
6. RESULTS.....	87
6.1 Introduction.....	87
6.2 The interviews	87
6.2.1 Biology: CLIL and non-CLIL	88
6.2.3 History CLIL and non-CLIL.....	94
6.2.4 English.....	100
6.3 The assessment samples	107
6.3.1 Design of the tests in the disciplines.....	107
6.3.2 Function and question words in the test items	110
6.3.3 Assessment in biology	112
6.3.4 Assessment in history	118
6.3.5 Assessment in English.....	124
6.4 Questionnaire	129
6.4.1 Types of assessment used	130
6.4.2 Most important factors in assessment	132
6.4.3 General views on assessment	134
6.5 Summary	135
7. ANALYSIS AND DISCUSSION	137
7.1 Introduction.....	137
7.2 Comparing assessment in CLIL vs non-CLIL.....	137
7.3 Effects of CLIL in the English language courses.....	143
7.4 Assessing language and content in the disciplines	146
7.5 Validation of assessment in a CLIL context	152
7.6 Summary.....	155
8. CONCLUDING REMARKS	157
8.1 Contributions and implications	158
8.2 Suggestions for future research	161
SWEDISH SUMMARY.....	165

REFERENCES	175
LIST OF APPENDICES	187

Acknowledgements

The road of life twists and turns and no two directions are ever the same. Yet our lessons come from the journey, not the destination. (Don Williams Jr)

I must profess that it has been quite a journey. It began in August 2012 when I started my doctoral studies, and reached its destination in May 2015 when I presented and defended my licentiate thesis. However, the process, leading up to the application for the graduate school, started long before. Curiosity and unsatisfactory answers as to the why, how and what of the assessment practices in my own teaching contexts, lead to a growing desire to investigate and learn more. My teaching experience may in some respects resemble what is referred to as CLIL in this study, but my background is mostly that of a foreign language teacher trying to find ways to enhance and make language teaching more authentic and relevant in the language classroom.

Being a doctoral student has been one of the most rewarding and challenging journeys I have undertaken. So many lessons learned. Now that the journey has come to a halt, the overarching feeling is gratitude. I say halt, not end, since I do not think of this as the end. The thesis may be completed, but the journey is not, it has merely begun. The more I have explored, the more there is to learn.

I want to express my gratitude to the FRAM steering committee for opening the door for me to pursue this dream, me together with my nine fellow doctoral colleagues in FRAM graduate school.

First and foremost I want to thank my wonderful supervisors Liss Kerstin Sylvén and Gudrun Erickson. Your support and expertise have been invaluable. Just as content and language can be said to represent two sides of a coin, you have stood out as an extraordinary team and contributed in distinct yet unanimous ways.

To my nine graduate school colleagues; it has been an honor coming alongside and sharing this experience with you. You have indeed contributed to the journey - both fellow students and all the FRAM supervisors. The same for the CLISS group. The invitation to be part of your research was a catalyst for this study.

I would also like to thank my employer Sigtuna kommun and the principals at Arlandagymnasiet for giving me this opportunity.

No informants, no study. Thank you to all the participating teachers at the three schools involved in this study. It was a pleasure meeting with you and taking part of your insights. Sölve Ohlander, thank you for helping me make improvements in my text after a very thorough reading of the final draft. A special thanks to Clas Olander, Gothenburg University and Martin Stolare, Karlstad University for sharing your expertise from the fields of the natural sciences and history during the final stages of the study. Thank you to representatives from the University of Gothenburg and Umeå and Malmö University for input regarding test design in biology and history.

I dedicate this thesis to my parents, Birgit and Ingemar, especially in memory of my beloved dad. A turn of life and you left for your final destination shortly before the thesis was completed. Words cannot express how much your unconditional love, practical support, genuine interest and prayers have meant along the way. I am forever thankful.

I also want to dedicate the thesis to my precious son and daughter, Adrian and Alina. You were only two and four when I embarked on this journey. I know that you have put up with a sometimes absent minded or absent mom as I have travelled to go to courses in Gothenburg or to different conferences. Your love for life and natural need for your mom's attention have carried me and provided me with new energy when mine has been drained. A special thanks to aunt Kerstin for being there when the family needed you to make the life puzzle work.

The thesis is also dedicated to my husband Göran. You have sighed at my frustration and smiled at my excitement along the way and I know that you rejoice with me for completing this phase of our journey. Thank you for your love and support. I love you!

And now, let the journey continue, knowing that no directions are ever exactly the same.

Vallentuna, May 2015

Helena Reierstam

1. INTRODUCTION

1.1 Background

Assessment is challenging for several reasons, some of which will be discussed in the following. One challenge is to be clear about *what* is to be assessed and to justify *how* the assessment is done, to quote Bachman and Palmer (2012:2):

We believe that despite the differences among people who use [...] assessments, what they all have in common is the need to be accountable for the uses for which their assessments are intended. In other words, they need to be able to demonstrate to stakeholders that the intended uses of their assessment are justified.

Teachers have different backgrounds and experiences, but they need to be able to describe what they do, since their assessment practices have consequences for individuals, institutions, and ultimately for society. To be able to justify the uses of certain assessment procedures in a context where the practices and consequences of a teaching strategy are unclear can be even more problematic. This is the case in many Content and Language Integrated Learning, CLIL, environments in Sweden, due to the lack of a common framework or guidelines for good practice (Socrates-Comenius 2.1, 2009; Sylvén, 2013).

CLIL is a teaching approach typically found in subject content courses where a foreign language is used as the medium of instruction, mostly English. At tertiary level in Sweden, it has become more common (Costa, 2009; Maiworm & Wächter, 2008), at least in certain academic domains such as the natural sciences (Airey, 2013, personal communication). In the present study, however, the focus is on upper secondary education, i.e. on students aged between 16 and 19.

The exact extent and scope of subject instruction through English in Swedish upper secondary schools has proved hard to determine. A survey conducted in 1999 (Nixon, 2000) reported that 23 % of all Swedish upper secondary schools had some content instruction in another language than

ASSESSING LANGUAGE OR CONTENT?

Swedish¹. Yoxsimer Paulsrud (2014) made an attempt to find out the present status. She concludes that the number is not actually increasing, unless it is defined as partial or occasional CLIL.

In a Swedish context, English finds itself in a unique position compared to the rest of the foreign languages taught in schools. Swedish authorities have even identified a need to define the roles and identities of English compared to Swedish. In 2009, a new language act (SOU 2008: 26) was passed to ensure the status of the Swedish language in Sweden, since Swedish was considered to be threatened in high status domains, such as higher academic education (Lindberg, 2009). This is rarely discussed and is not an issue among most people; on the contrary, young Swedes seem to favour English (Oscarson & Apelgren, 2010).

In international surveys and testing, Swedish students attain very high proficiency levels in English. According to the European Survey on Language Competences, Swedish students perform almost as well as young people from Malta where English is an official language (European Commission/SurveyLang, 2012b).

English is present on a daily basis in the lives of especially many young people, who are exposed to a great deal of extramural English outside of school through ICT and other media (Sundqvist & Sylvén, 2012; Sylvén, 2006). Hyltenstam (2004:53-54) lists four reasons for the prominent role of English among Swedish young people: extramural exposure, frequent travels abroad, English being considered the most useful language to learn according to a survey (European Commission, 2006), and the typology factor: Swedish and English are both Germanic languages making English reasonably easy to learn for Swedes.

The purpose for implementing CLIL may vary, one aim being to prepare students for a global world and an international context (Eurydice, 2006; Dalton-Puffer, 2007). This aim also mirrors a view that language learning in the language classroom is unsatisfactory or at least insufficient (Dalton-Puffer, 2007). Consequently, one reason is to make language learning more authentic and relevant (European Commission, 2012a). The prominence of English and the varying status of foreign languages in Sweden might raise the question why other foreign languages are not used as the medium of instruction. The

¹ 76% of the contacted schools responded. The results were self-reported and great variation in the respondents' definition of CLIL was acknowledged.

INTRODUCTION

answer is probably due to the implementation of CLIL in the subject content courses rather than in language courses, requiring both teachers and students to be proficient enough to use an L3 as the medium of instruction (see section 2.2).

At the core of assessment in CLIL are issues related to the relationship between language and subject content (see section 2.4). The same issue has received attention lately in the national Swedish instructional discourse due to immigration, causing many students with a foreign background to learn subject content in a non-L1 language. The Swedish National Agency for Education (henceforth referred to as NAE) states in a recent survey (2012b) that multilingual students need instruction with a clear dual focus on both language and knowledge development.

As regards assessment, NAE has recognized a lack of research on a national level concerning the design and use of assessment tools. It is noted that teachers employ a great variety of tools, including tests, portfolios and rubrics, but there are very few studies on how these are actually used (NAE, 2011b). Moreover, it is stated that in assessing written test outcomes, two parallel procedures seem to be prevailing: a quantitative scoring using points or grades, or the use of test items representing different complexity levels. Most likely, there are differences between disciplines and different educational levels (NAE, 2011a).

The questions raised above led to the present study: the aims of language instruction to make teaching authentic and relevant, the reality of young people; their educational needs and extramural exposure to the English language, the uncertainty in assessment procedures concerning *what* and *how* to assess, all of which create a complex teaching context for teachers. The study of CLIL adds the question whether it is possible to bring content and language closer together in the development of interdisciplinary assessment tools. If so, the CLIL practices investigated in the present study may contribute to the process of making assessment more authentic, as well as outlining a possible future framework for good practice in assessment, including a conscious dual focus on both language *and* content.

1.2 Aims and research questions

The purpose of this study is to investigate assessment in relation to bilingual and interdisciplinary teaching as carried out in three upper secondary CLIL

ASSESSING LANGUAGE OR CONTENT?

schools in Sweden. The focus is to explore if, and if so, how assessment procedures differ in two subject content courses within the same discipline. One of the courses is taught in Swedish L1, the other using English L2 as a medium of instruction, i.e. CLIL. The courses in focus are biology and history. The terms *subject* and *discipline* are used interchangeably, and sometimes in combination. One reason for using the latter is the prominent use of terms as inter- and cross-*disciplinary* in relation to CLIL, whereas the first appears in terms as *subject* content courses in CLIL. A glossary in Appendix 1 provides a summary of important terminology and abbreviations used in the study.

Another question concerns the English language courses, often called EFL-courses in the present study, if the course content and thus the assessment tools are affected where English is used in subject content courses. A third question concerns interdisciplinary similarities or differences when it comes to language, content and form in assessment. The specific research questions are outlined below:

- CLIL vs non-CLIL: do the assessment practices differ in the two subject content courses history and biology due to the language of instruction? If they do, *how* do they differ, and *on what grounds*?
- Are the assessment tools and the course content affected in the English language courses where English is used in subject content courses? If so, *how* are they affected?
- What does the assessment design look like in the different disciplines when it comes to *language, content* and *form*? Are there common features?

Each of the research questions is meant to provide an understanding of *how* content and language integrated teaching affects teachers' choices in designing assessment tools in their context. The aim is to contribute, albeit on a small scale, to the fairly unexplored field of research regarding assessment in CLIL. The focus of the third research question on content and language in relation to assessment in different subject disciplines, may possibly also contribute to the role of languages in all content courses, regardless of whether the language of instruction is the students' L1 or L2, or possibly L3 (NAE, 2012b).

INTRODUCTION

The study is part of a larger research project, Content and Language Integration in Swedish Schools (CLISS), funded by the Swedish Research Council, 2011-2014, where the main focus is to compare the development of CLIL and non-CLIL students' academic language in Swedish and English written production (for further details see Sylvén & Ohlander, 2014).

To provide an illustration of the outline of the study, the figure below offers a picture of the different components and layers.

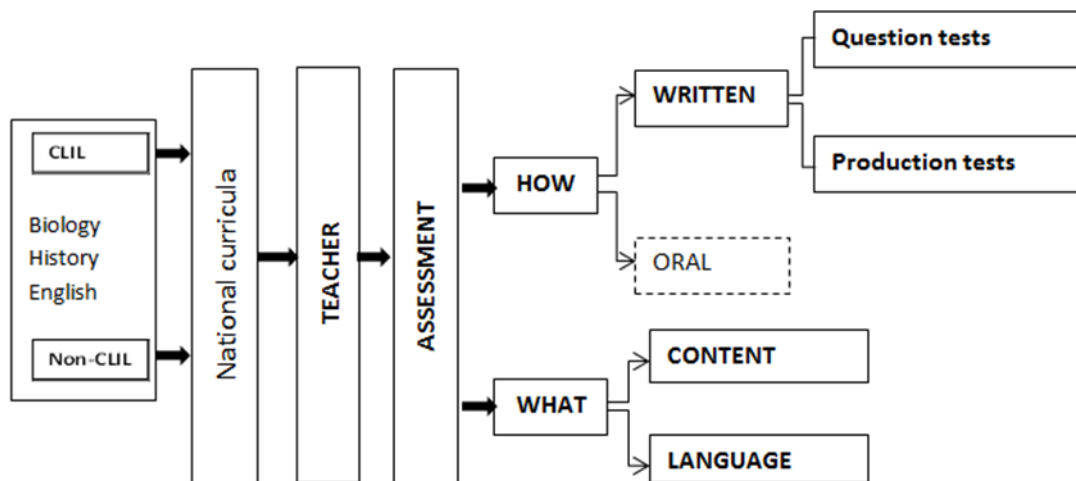


Figure 1. Outline of study

The figure illustrates how assessment is a result of national and individual contexts, the macro and micro levels. Assessment is directly affected by teacher cognition (individual context), both in test development and assessment use. The teachers' interpretation of good practice depends on experience and theories of learning, but also relates to curricula, the syllabus and course goals (national context). The term *syllabus* is used in this study to signify the national descriptions of course content in the individual disciplines. The term *course goals* is used for what sometimes is labeled as course objectives. The NAE uses the term *knowledge requirements*, a term which will appear as well, aiming at intended disciplinary learning outcomes.

Looking to the left in the figure, the context is determined by the subject course and whether it is a matter of CLIL or not, which is a local decision of the school on the micro level. The question is, as expressed in the first research question, *if* and *how* the language of instruction, as in CLIL or non-CLIL, has an effect on teachers' assessment practices. The teaching methods

ASSESSING LANGUAGE OR CONTENT?

must pass through national curricula and teacher lenses, before appearing as assessment practices. Whether the CLIL context has an impact on the English language courses is the objective of the second research question. The relationship between English as a subject and the content courses is not obvious in the figure (the box to the left), a relationship which somehow reflects the context in the present study.

The third research question focuses on the design of the assessment, dealing with both mode, the *how*, as well as the construct of the test items, the *what*, including both content and language, as shown in the figure as well. The present study focuses on the written format, found in the assessment samples. *Oral* appears in the figure, although dashed, acknowledging the equal status of oral and written production.

The distinction made between *question tests* and *production tests* relates to two different types of assessment, which could be labeled as tests or exams. In the current study the word *test* is used rather than *exam*, the latter often signifying high-stakes testing, which is not the common test type in this study. Question tests refer to multiple question tests, usually paper and pencil tests, requiring different types of answers. All other writing assignments used for assessment purposes, such as essays or lab-reports, are here referred to as production tests (cf. 3.6.2). The term *production tests* was chosen, although the writing assignments could be described as a kind of performance tests. The term seemed appropriate in relation to the term “written production”, often used in FL-courses and the CEFR. The terms written assignments and writing assignments are often used synonymously in the literature. In this study, the term writing assignments is used to denote a specific format, containing a prompt or task description designed for the written mode. The term written assignments, when used, signifies a broader category, referring to the written mode, as opposed to the oral mode.

The theoretical background in this thesis is given in three separate chapters, the first dealing with CLIL, the second with assessment and the third with language and content in the three disciplines. The design was chosen in order to provide a brief summary of each individual field even though there are overlapping features, inherent to the integrative character of CLIL. Below follows a section on the empirical and theoretical perspective of the study before a more detailed overview of the outline of the thesis.

INTRODUCTION

1.3 Empirical research framework

The empirical research perspective of the study is qualitative, consisting of descriptive data. The material is collected through a methodological triangulation using semi-structured interviews, gathering of assessment samples, used for a documentary analysis, and a questionnaire.

A cognitivist psychology perspective, found in teacher cognition and assessment literacy (see section 3.4) is combined with a socio-constructivist perspective, common in CLIL contexts (Dalton-Puffer, 2007). The focus is on how the participants' actions can be understood as part of individual as well as social practices.

The theoretical framework leans on research in foreign language acquisition; (FLA) and second language acquisition; (SLA), particularly in relation to bilingual teaching and foreign language assessment. The conceptual framework relies on the expanded view of validity (Messick, 1989, 1996; Bachman, 2005). It stretches validity beyond issues of construct coverage to considerations of issues of relevance, utility, and value implications (McNamara, 2006; Shepard, 1993). Hereby construct based interpretation and inferences are emphasized, as well as possible consequences of test use (Erickson, 2010).

The analysis of the interview material is based on thematic analysis (Rapley, 2011:274f), whereas for the document analyses, different features are used for encoding structure and content. Atkinson and Coffey (2011:80) acknowledge that documents are distinguished by certain types and genres and are marked by the use of very specific language and form, as the assessment samples in this study.

1.4 Outline of thesis

The current thesis builds on three pillars and three disciplines, representing the foundational building blocks of this study, as seen in Figure 1 above. The three main areas, already outlined in section 1.2, consist of the CLIL approach, teacher cognition and assessment. The three subjects are biology, history and English in an upper secondary educational setting. Consequently chapter 2, following this introductory chapter, offers a brief overview of CLIL and related teaching approaches.

Chapter 3 covers a range of issues related to assessment. A brief review of the Swedish context is presented and a description of assessment in relation to

ASSESSING LANGUAGE OR CONTENT?

the three disciplines and the written format is provided. At the core of the chapter is the expanded view of validity. A description of a validation chain model used in the current study concludes the chapter.

In Chapter 4, language and content are examined in relation to cognitive and linguistic demands on the student, the three disciplines involved outlining the core concepts of the Swedish national curricula and the Common European Frame of Reference, CEFR.

Chapter 5 discusses the methods used and gives an account of the material, the data collection procedure and the participants.

In Chapter 6 the results from the semi-structured teacher interviews, the documentary analysis and the questionnaire are described and presented.

A triangulation and discussion of the findings in relation to the research questions is made in Chapter 7, a chapter which also includes a tentative validation model for assessment in CLIL.

Chapter 8 discusses pedagogical implications and presents possible features for the development of CLIL guidelines for assessment before proposing areas of future research concerning CLIL and assessment.

The appendix section contains a glossary of important terminology, a sample of the questionnaire as sent to the teachers and an overview of some teacher responses to the questionnaire.

2. CLIL AND INTERDISCIPLINARY TEACHING

2.1 Introduction

Content and Language Integrated Learning, CLIL, is often referred to as an umbrella term for bilingual teaching approaches with the aim to combine language and content learning to some degree (Mehisto, Marsh & Frigols, 2008). Content appears first, and investigations have shown that the method is practiced in content subject courses for the most part, not as much in second language courses (Dalton-Puffer, 2007). Language in this context implies a second or foreign language, in other words a non-L1, used as the medium of instruction in non-language classes. Other variants, such as Content Based Instruction, CBI, or Content Based Language Teaching, CBLT, are curricular models implemented in second or foreign language classrooms. Regardless of model, the goal is to let the content or the language enhance the learning of the other.

In this chapter, a brief overview is offered of some of the most common interdisciplinary teaching methods involving content and language. The purpose is to orient ourselves among the general characteristics of and challenges associated with the methods, but also to clarify what CLIL represents in the present study.

2.2 Different variants: an overview

In interdisciplinary teaching, one of the main questions concerns the degree of integration (cf. section 2.5), which depends on the context, the model used and the users' reasons for choosing the model. Looking at CLIL, two roles of the language in content learning have been distinguished; one where the language is seen as a tool or medium of instruction, applicable in most CLIL cases, and the other where it is seen as an additional learning goal, which is consciously and systematically pursued by the teacher (Socrates-Comenius 2.1, 2009).

ASSESSING LANGUAGE OR CONTENT?

In a global context, EMI, English as a medium of instruction, may be a more commonly used term, especially in tertiary education. Yoxsimer Paulsrud (2014) chooses the label EMI to describe the practice of the teachers in her study in a Swedish upper secondary school context. She uses the term to infer that it is not a matter of content and language integration, but merely of *language alternation*, thus suggesting that the content is taught in exactly the same way as in the native language. English is only used as a tool and not consciously or systematically processed in the classroom.

Immersion is a commonly used term for content and language integrated methods in Canada, representing the original model which CLIL has developed from. Key factors to successful implementation have been the involvement of parents and support from education authorities (Eurydice, 2006). Immersion is content-driven, and the focus is to learn language “naturally” with an emphasis on the use of language for communication. Yet research shows that receptive skills improve more than the productive ones, and native-like qualities are not acquired in speaking and writing. The age of onset in language learning seems to have an effect on the results of the L2 studies; consequently, early provision seems to prompt more analytical language abilities, for instance among older students (Sylvén, 2004).

In an American context, labels such as CBI, content based instruction, or CBLT, content-based language teaching have been used (Lyster & Ballinger, 2011) and are compared with European CLIL (Brewster, 2004). As already noted, CBLT is found in language classrooms, but it is still content-driven. Lyster and Ballinger use a continuum to compare variants of bilingual teaching. The only variants that can be said to be language-driven according to this model, are those found in more traditional language classrooms borrowing content themes for authenticity in the use of language.

CLIL AND INTERDISCIPLINARY TEACHING

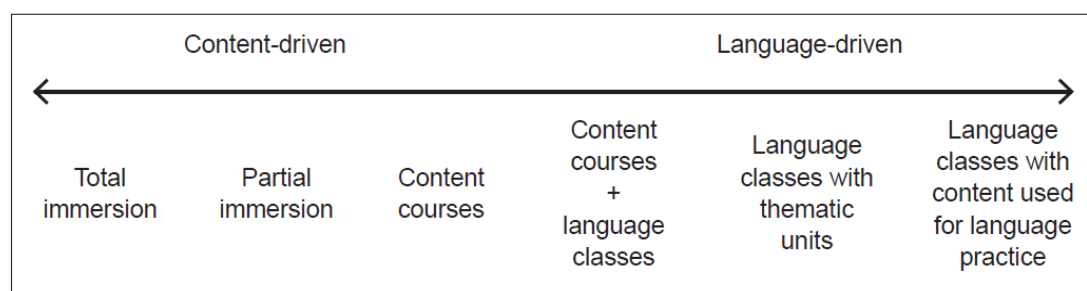


Figure 2. Range of CBLT (Content Based Language Teaching) settings

Source: Lyster & Ballinger, 2011:280

The figure shows how different teaching approaches may lean more towards content or language, but as will be seen in the present study on CLIL, it depends on the users, in this context the teachers. In one definition of CLIL it has been described as operating along “a continuum of the foreign language and the non-language content without specifying the importance of one over the other” (Coyle, 2010:2).

The next two sections provide an overview of the main features characterizing the diverse CLIL practices as well as a brief description of the Swedish CLIL context.

2.3 CLIL, discourse and practice

In 1995 the European Commission expressed their goal to increase proficiency in more than one foreign language among European citizens. Methods and measures to make students learn more languages other than English, and become more fluent in all of those, are being promoted, CLIL being one such suggested practice (Socrates-Comenius 2.1, 2009). In CLIL contexts, however, most often English is the language used (Yoxsimer Paulsrud, 2014), and so some claim that it should be labelled CEIL, as in “Content and *English* Integrated Learning” (Haataja, 2013, personal communication). The fact that CLIL is implemented in content courses rather than language courses may be part of the explanation, although that question deserves its own survey.

There exist no guidelines on how to implement CLIL (Sylvén, 2013), which has led to various efforts to distinguish some common features in order to offer a “scaffolding framework” and a coherent view (Socrates-Comenius 2.1, 2009). One of the main motives for choosing to promote the integration

ASSESSING LANGUAGE OR CONTENT?

of content and language is a belief that learners will benefit from a conscious focus on both in a learning context (Mehisto, Marsh & Frigols, 2008). On a somewhat less positive note, some previous research has shown that the CLIL approach sometimes fails to enhance the language skills of the students (Edlund, 2011; Lim Falk, 2008), at least in a Swedish context, to further be discussed in the following section. Coyle (2010:3) admits that CLIL per se does not guarantee effective teaching and learning and Coyle et al (2010:48) argue that certain fundamental principles need to be in place for CLIL to be effective; not just any kind of teaching in another language is CLIL. The teachers in the present study do not call themselves CLIL teachers, but the term CLIL is used since the results of the study will be compared with other CLIL practices. Since the CLIL approach is flexible according to Coyle et al (2010), and there is no common best-practice, the term seems relevant for the purposes of the present study.

Coyle (2010) articulates a need to state what CLIL is *not*; CLIL is not a trend, it has been around a long time. It was adopted by the European Network of Administrators, Researchers and Practitioners (EUROCLIC, 2010:5) in the mid 1990's. However, learning through a foreign or second language (L2) is ancient, dating back at least to Socrates' Academy, according to Masih (1999). CLIL has been referred to as the natural approach by Krashen and Terrell (1983) The natural approach, as well as CLIL, sees communicative abilities as a primary function of language to increase motivation to learn languages (European Commission, 2014).

Further, CLIL is not trying to replicate any other models such as the Canadian immersion model, but it is rather a range of flexible European models responding to contextual demands. Massler, Stotz & Quessier (2014) distinguish three forms of CLIL provision and assessment. The first variant (A) means CLIL in subject lessons; the second, (B), implies CLIL in foreign language classrooms; and the third variant (C) means fully integrated learning of subject and foreign language. In the schools studied by Massler et al, type A is most common in German schools at primary level, whereas in Swiss schools variant B is advocated, integrating CLIL in foreign language classes. In Swiss schools, CLIL cannot normally be integrated in science lessons. Consequently, different types of implementation are found across contexts.

In a wider European context, Dalton-Puffer (2007:3) notes that: "CLIL classrooms are seen as environments which provide opportunities for learning through acquisition rather than through explicit teaching". CLIL leans on

CLIL AND INTERDISCIPLINARY TEACHING

sociocultural and constructivist learning theory in joining together two complementary views on learning, which according to Coyle et al (2010:3) means that “parallels between general learning theories and second language acquisition (SLA) theories have to be harmonized in practice if both content learning and language learning are to be successfully achieved”.

It has been argued that CLIL differs from CBI and CBLT in that CLIL involves a “planned pedagogic integration of contextualized content, cognition, communication and culture” (Coyle et al, 2010:6), often referred to as Coyle’s four Cs (Coyle, 1999). Whether this is true or not is a matter of validation from case to case in the CLIL-context, and will not be further discussed here. Nevertheless, *content*, according to Coyle defines the topic content in a course or lesson; *communication* defines the language skills to be used during a lesson; *cognition* signifies the thinking skills needed for the class or theme; and finally, *culture* implies reference to the students’ experience and surroundings, but above all the target language culture. It is sometimes labelled citizenship. The table below, Table 1, provides an example from teaching science in English:

Table 1. Example of the use of Coyle’s 4Cs when planning a science lesson

Content	Communication	Cognition	Culture (Citizenship)
The topic: plants. Lesson and/or course content.	Language needed during the lesson: comparing, contrasting in order to analyze similarities and differences between fungi and plants. Using target language.	Thinking skills demanded of learners during the lesson, e. g. classifying, thinking about advantages vs disadvantages of growing plants in certain environments.	Find out about indigenous plants to the learners’ home country, popular plants around the world, compare fertilizers used in different countries. Understand own culture and that of others

Table adapted after Cambridge ESOL2008

The concept of integration is what differentiates CLIL from immersion and the other variants, according to Coyle (2010). Referring to De Bot (in Marsh, 2002), Coyle states that integration implies that language and content teachers need to work together to achieve a real integration of form and function in language teaching, language being promoted as a medium for learning as well as an object of learning, whilst the subject is safeguarded (Coyle, 2010:3).

2.4 Assessment in CLIL

Assessment in CLIL is considered an underexplored area; Massler et al (2014) even calls it a blind spot. Limited empirical studies have been conducted in the field. However, several investigations have been reported over the last few years. Hönig (2009) discusses subject content teachers' perceptions and practices in oral exams in an Austrian context, whether teachers consider linguistic performance or not. Wewer (2014) investigates assessment practices in primary CLIL in Finland, with a special interest in students' progress in the target language, i. e. English. The study also looks into computer simulations as a medium of assessment in CLIL. Massler et al (2014), referring to the German and Swiss contexts, note that there are few accounts of how teachers assess progress and achievement in CLIL. They point to a lack of policy decisions and assessment guidelines and tools, suggesting a model for primary CLIL assessment in which language and subject content are combined. Gablasova (2014) presents a study performed in Slovakia on students' choice of language to communicate content knowledge in assessment in bilingual teaching, by using the language of instruction, the students' L1, or a mix of both as in translinguaging².

In a CLIL context, the effect of the language of instruction, both on comprehension and students' own linguistic production, is a matter of concern. The learner is exposed to linguistic input in a second language at a relatively complex cognitive level, and therefore has to process content knowledge and language at the same time. In a way the same is true among native speakers when first introduced to a new discipline, processing concepts as well as acquiring the new disciplinary language (Olander, 2014, personal communication). This means that there are two processes involved in the assessment, language acquisition and subject-learning, which in turn generate the question of whether language and subject content should be assessed at the same time and through the same tasks and activities. If a student performs poorly on a test in history, does that mean that he or she has not understood the question, has limited understanding of the historical concepts, or possesses insufficient language competence to express his/her comprehension clearly?

² Translinguaging refers to flexible use of multiple languages in the meaning making process in the multilingual classroom, cf. section 4.3.3.

All of the above-mentioned studies (Gablasova, 2014; Höning, 2009; Massler et al, 2014; Wewer, 2014) acknowledge the issues in CLIL assessment related to the dual focus on language and content. Kiely (2009:4) discusses the purpose of assessment in CLIL as well as the issues of language versus content; “[H]ow do we use assessment to manage an appropriate balance in CLIL practice between content and language, such that there is no fear that children [students] achieve less where the learning is in L2?”

Morgan (2006) advocates new assessment tools for CLIL since she finds that curriculum criteria and current testing procedures do not accommodate the special skills acquired by CLIL students. She underlines that external validation is particularly important in a CLIL context where the teaching and the learning “stand outside the mainstream”. The question, according to her, is whether CLIL is associated with an awareness of language skills needed in the content courses, and if students acquire different skills; if so which skills, and moreover if the proficiency level attained in English can be rewarded in the English language courses?

The students in the CLIL classroom are bilingually educated, even if the subject specific concepts are not taught in the students’ L1. This raises the question which language to choose for assessment. Gablasova (2014) lists four options: the students can be tested in the language of instruction, the L2, or in their L1; a third solution implies parallel assessment in both languages, or a mixture of both languages, implying translanguaging, recommended by, among others, García (2009).

2.5 CLIL in Sweden

CLIL is context-embedded and the application of the approach as well as the results hinge upon the cultural conditions. The positive effects of CLIL in Sweden have been questioned (Edlund, 2011; Falk, 2008), explained in part by the already prominent role of English in Sweden. Sylvén (2013) compares and contrasts CLIL in Sweden with the practices in other European countries and identifies four important factors to cater for differences in context: lack of CLIL education, lack of CLIL framework, the presence of extramural English, and the age when CLIL is implemented. She confirms that the national school policies differ markedly between countries: in some there are requirements for teacher training and in others the amount of teaching done in English is stipulated in order for the education to be labelled CLIL. Extramural exposure

ASSESSING LANGUAGE OR CONTENT?

to English differs greatly. In the Swedish context it is very high, as opposed to, for instance Spain, which of course will have implications for how to implement and evaluate CLIL.

There is no teacher certification for CLIL teachers in Sweden; rather, it is a matter of regular content teachers often with an interest in English, to teach their subjects in a non-native language. Since content teachers usually have no training in how to teach languages, the processing of, for instance, vocabulary relies on the insights of the individual teacher. As Dentler (2007:170) notes:

As there is hardly any support, neither on state nor municipal level, CLIL in Sweden manages to survive through the endeavors of some 300-400 teachers working (mostly alone) as fiery spirits against bad odds. This implies that most schools have no internal monitoring system to evaluate how the goals are fulfilled or how to facilitate further development.

Dentler (2007) adds that there are schools which take on the responsibility of evaluating the CLIL approach themselves, but she comments that there are no regulations regarding CLIL “as long as the programs conform to the school law and the national objectives are reached” (2007:167). Dentler states that the CLIL programs normally exist alongside ordinary educational programs; however, IB schools (International baccalaureate) have increased in number.

In Swedish schools, teachers give evidence of informal initiatives among colleagues to create cross-curricular and interdisciplinary themes for shorter projects. At the same time, there are upper secondary schools which use a conscious subject-integrated teaching approach to market their school on their websites. Marketing reasons are acknowledged to be one purpose for implementing CLIL, since this is considered to be attractive among stakeholders and young people in Sweden (Dentler, 2007; Yoxsimer Paulsrud, 2014). Thus there are many variants, three examples include: IB schools, following an international curriculum; international schools, following the Swedish curriculum, but do most of their teaching in English; and regular national schools with one or several classes with an international profile.

Kjellén Simes (2008) investigated the impact of English immersion by comparing IB students to students learning English in regular foreign language courses. She concludes that vocabulary competence had increased among the immersion students. After a three-year period the IB students “used significantly larger proportions of motivated tense shift as well as low-

CLIL AND INTERDISCIPLINARY TEACHING

frequency vocabulary than the NP students³” (2008:163), which, she comments, agrees with reports from Dalton-Puffer (2008).

Lim Falk (2008) compares CLIL to non-CLIL students in Swedish upper secondary school and notes that CLIL students are less confident in using Swedish than their peers in the non-CLIL classes. She also finds that often there is no interaction in the CLIL classrooms, which indicates that English is seen as an obstacle, favoring teacher dominance instead of student participation.

Kjellén Simes (2008) gives a more positive view than Lim Falk (2008), finding actual linguistic gains as a result of learning through a foreign language. However, Sylvén (2004) notes in her study that the proficiency level of the CLIL students was much higher already at the outset, compared with non-CLIL students. Moreover, she claims that the explanation is to be found in extramural exposure rather than in the use of English as a medium of instruction. As in all studies on CLIL, the language proficiency and motivation of the students at the outset have to be considered. CLIL students in previous research, as well as the IB students in Kjellén Simes’ study, measure higher on both. Motivated students who already have a good level of English seem to choose CLIL alternatives. The differing reports on the effects of CLIL initiatives in Sweden as well as the lack of teaching guidelines and teachers’ training suggest inconsistency and arbitrariness in the implementation of CLIL in Sweden.

As noted previously, the NAE performed a survey (Nixon, 2000; 2001) to investigate the spread and the scope of CLIL. The surveys found that the majority of the CLIL programs sprung from teacher initiatives. They were mostly found in municipal schools and the CLIL practices were poorly documented and the schools lacked a qualifications policy for CLIL education and teachers involved. CLIL was found in one form or another in 4% of the compulsory schools and more than 20% of the upper secondary schools. Nixon, who performed the surveys, reports an increase in the implementation of CLIL during the 1990s, as do Edlund (2011), Lim Falk (2008) and Sylvén (2004). Yoxsimer Paulsrud states that there exist no official national statistics on the number of schools offering CLIL, partly due to the lack of a definition of CLIL, since schools vary in their degree of implementation, even between lessons [and teachers] in the same school (2014:55f).

³ NP = (regular) national program

ASSESSING LANGUAGE OR CONTENT?

Haataja (2013) uses a model called the “CLIL spiral” to distinguish between different levels or degrees of integration in the implementation of CLIL in schools. The first, most basic level is characterised by single “mini-projects” in foreign language or non-linguistic subjects with an integration of target language into subject teaching or vice versa. This level is for the most part independent of systemic curricular or teacher collaboration.

Level 2 means cross-curricular arrangements of projects and trial classes. The realisation is both in language and in non-linguistic subject-classes, often in cross-curricular interconnection, by solving maths problems in English, for instance.

Level 3 represents CLIL-modules with systematic development of CLIL teaching competences in language and subject content. As a result there can be CEFR-based task-specific assessment.

The fourth and most integrated level implies a sound curriculum, according to Haataja, with planning for CLIL, including examination structures and degrees. It involves long-term programs for in-service training for teachers. It also includes organisation and accompanying longitudinal research measures.

In view of previous research on the implementation of CLIL in Sweden, it seems as if most CLIL settings would be found on level 1 or 2 according to Haataja’s model, since there is no systematic development of teaching competences, for instance. In order to see effects of the content- and subject-integrated teaching, there should be a more conscious integration and interdisciplinary cross-curricular collaboration (cf. Coyle, 2010).

2.6 Summary

In this chapter CLIL has been discussed, summarizing its most important features and considering CLIL in relation to some related variants: immersion, EMI, CBI and CBLT. Concerning the prominence of content versus language, CLIL is typically implemented in content courses and subsequently content-driven. Seen both from a Swedish and an international perspective, common issues exist regarding the lack of guidelines and documentation of the effects of CLIL, encouraging further research in the field. In Sweden, the effects of CLIL are even harder to evaluate due to the great presence and impact of English extramural exposure.

CLIL AND INTERDISCIPLINARY TEACHING

Coyle suggests a planned pedagogic implementation of CLIL, taking four Cs into account: content, communication, cognition and culture. However assessment in CLIL, the focus of the present study, is still considered a blind spot. Of particular concern in this connection is the dual focus on language and content.

This chapter has briefly presented the historical background of CLIL, in relation to immersion and the goal of CLIL: to promote the learning of more than one foreign language. The next chapter deals with assessment.

3. ASSESSMENT

3.1 Introduction

The purpose of this chapter is to provide a background and better understanding of the prerequisites for teachers' assessment practices in the subjects and the Swedish CLIL context in the present study. The main concern incorporates the *how* and *what* of assessment (Shohamy, 2008:xiv):

Matching the 'how' of testing with the 'what' of language uncovers several periods in the development of the field, with each one instantiating different notions of language knowledge along with specific measurement procedures that go with them.

Whether the construct of assessment, i.e. what to assess, is language or subject content, or both, it is affected by theories of learning and current ideologies. Hence, assessment will be briefly discussed in relation to historical and contextual factors, where the current Swedish context is given some special attention. Teachers' assessment literacy and curricular features of the subject disciplines are presented before looking into modes and features of written assessment. The chapter ends with a discussion of validity and presentation of a model for validation of assessment procedures.

3.2 Historical and contextual impact

Teacher assessment is affected by prevailing ideologies and therefore implies a need to be aware of the epistemological bases of different types of assessment (Hofer & Pintrich, 1997). Inbar-Lourie (2008) argues that assessment practices are compatible with social expectations, attitudes and values.

According to Tittle (1994), who represents an educational psychologist view, the validation arguments for assessment will be stronger when they "include evidence on the constructions of teachers and students and the meanings and use an assessment has for them in their educational situations" (1994:149). Contextual frame factors can be found both on the micro and macro level: the teachers and the assessment culture at the school in question

with its resources, the motivation and background of the students, as well as the national curriculum and historical context.

Bachman (1990:291) referring to the use of language tests in particular, considers how these are determined by political needs that change over time and vary from one society to another: “We must consider the value systems that inform test use – values of test developers, test-takers, test users, the educational system, and society at large.”

The current discourse on assessment often uses terms such as traditional versus alternative assessment, even though definitions are not clear and a dichotomy thus hardly fruitful. Alternative assessment in this case involves classroom interaction and dynamic assessment (Lantolf & Pohener, 2008; Rea-Dickins, 2004), which can be deduced from sociocultural theory, but also portfolios, which contain a collection of student work.

Shephard (2000:4) states that “it is important to remind ourselves where traditional views of testing came from and to appreciate how tightly entwined these views of testing are with past models of curriculum and instruction”. She argues that theories from the past continue to affect current practices and that, in spite of recent attention to the reform of the content and form of assessment, common practice has not moved significantly beyond the end-of-chapter test.

Some argue that there has been a paradigm shift where assessment culture has replaced testing culture (Lundahl, 2007; Taras, 2005). Assessment as a social communicative tool can help a learner move forward by developing metacognitive skills and an awareness of what constitutes topical knowledge in relation to the discipline in question, and what is the next proficiency level when it comes to cognitive and linguistic skills (Broadfood & Black, 2004; Gipps, 1999; Harlen, 2007; Inbar-Lourie, 2008; Shepard, 2000). Formative assessment means making learning visible to the learner (Black & William, 1998; Hattie, 2009; Sadler, 1989), which requires that teachers are able to describe their often tacit understanding of course goals and interpretations of learner performance.

3.3 The Swedish context

According to an OECD review of evaluation and assessment in education performed in February 2011, Sweden has one of the most decentralized education systems in the world. This dates back to a major administrative

ASSESSMENT

reform which took place in the early 1990s. The decentralized system implies that school leaders and teachers have wide-reaching autonomy in deciding on teaching content, materials, methods and study options. The NAE has developed common national curricula and syllabi, but within each classroom and school context, teachers develop the specific goals for each course based on the national documents, sometimes with the assistance of students (OECD, 2011:35).

The impact of consumer decisions has also increased due to a school choice reform. This has led to a surge in the number of individual schools, and the development of special profiles in municipal schools, e.g. international CLIL profiles, to attract students in an open market. Sweden has also become a culturally and linguistically diverse country with 20% of the population having an immigrant background, according to an OECD report from 2010. This implies that Swedish schools are faced with great challenges in adapting to a diverse student body.

Teachers in Sweden are test designers and agents in implementing what predominantly consists of their own teacher-developed tests and assessment tasks. However, in some courses there are also national tests with the purpose to facilitate fair, standardized and reliable awarding of grades (NAE, 2005). In upper secondary school, three courses are subject to national tests: the English language course, Swedish and mathematics. The OECD report problematizes the lack of guidelines as to how much weight should be given to the national test result within the overall grade assigned to students. A survey conducted by the NAE in 2009 shows great differences between teachers in this regard (OECD, 2011:50).

In the 1990s, Sweden went from a relative and norm-referenced grading system to a goal- and criterion-referenced grading system. In 2011, a new grading scale with six grades was introduced. So-called knowledge requirements (grading criteria or performance standards) exist for grade levels A, C and E, A being the highest grade. For B and D it depends on whether the students “have attained the majority of the knowledge requirements for the higher pre-established grade (i.e., A or C)” (NAE, 2013).

Teacher training in assessment is typically centred around formative assessment. Rubrics are often recommended as a tool to make learning visible (Jönsson & Swingby, 2007). In the educational discourse, different forms of assessment appear: portfolio, peer assessment and self-assessment. However, surveys reveal that these are among the least used and least preferred sources

for assessment, whereas grammar tests, essays, teachers' own tests and oral communication are the predominant instruments among FL teachers (Oscarson & Apelgren, 2010). Furthermore the OECD report notes that the use of computer-based assessments is very limited in Sweden while at the same time international test developers are now devoting significant attention to developing effective computer-based assessments.

To conclude the section on the Swedish teachers' situation, teachers themselves express a need for more training as regards assessment and grading (Oscarson & Apelgren, 2010). At the same time, the OECD report notes that little guidance is provided on how to appraise teacher performance. Measures are suggested, some of which will be discussed in the next section on assessment literacy.

3.4 Assessment literacy

Research suggests that teachers spend from one quarter to one third of their professional time on assessment-related activities, without necessarily having learned the principles of sound assessment, according to Stiggins (2007). Assessment literacy is a term that advocates evidence-informed practice and for assessors i.e. teachers, to reflect on the effect of their teaching and assessment strategies. Assessment literacy relates to validity in testing and assessment (Popham, 2006:84):

[I]f a teacher mistakenly believes that validity resides in the test itself, the teacher will be inclined to defer to whatever results the "valid test" produces. Assessment-literate educators, however, understand that education tests merely provide evidence that enables people to make judgmentally based inferences about students.

According to Popham (2009:7), teachers who are genuinely assessment literate know both how to create more suitable assessments and are familiar with "a wide array of potential assessment options". However, Malone (2008:225) states that "there is no consensus on what is required or even needed for language instructors to reliably and validly develop, select, administer and interpret tests". A gap between language testing practice and the training of language instructors is acknowledged. The CEFR is mentioned as one useful tool to bridge the gap.

Shepard (2000) claims that teachers need help in learning to use assessment in new ways in order to develop students' "robust" understanding.

ASSESSMENT

All too often, the same test types are used, implying that mastery does not transfer to new situations since students have learnt to master classroom routines and not the underlying concepts.

Assessment literate teachers consequently know how to choose and use the best method of assessment to fit the context, the students, the level and the subject. Validity, reliability, authenticity, washback, purpose, student impact and constructive alignment are identified as influential concepts for assessment literate teachers (Brown, 2004; White, 2009).

Washback does not only relate to products, as in assessment outcome, but also says something about participants and processes (Bailey, 1999; Hughes, 1994). Brown and Hudson (2002) mention that a multiple choice grammar test used to test communicative performance will have a very strong negative washback effect on a communicative curriculum. Washback is related to validity, and Messick (1996) states that there needs to be an evidential link between learning outcomes and test properties. In CLIL, as in the present study, such an evidential link may not be obvious as regards language. The intentional learning goals focus on content, which is a matter of validity in the CLIL approach and will be discussed later.

The teacher's learning intentions, as seen in the objectives, will in the best of worlds be aligned with course goals, course content, the type of assignments, material and methods used, as well as what appears in the assessment. If that is the case, the learning outcomes will agree with the learning intentions. Biggs uses the concept of constructive alignment (Biggs, 2001; 2003), arguing that effective learning is a result of a well thought-through process where teaching and learning activities are aligned with the Intended Learning Outcome (ILO), curriculum objectives and assessment tasks. Brown and Hudson (2002:48) claim:

If the relationship between testing and curriculum is solid and clear, if the objectives do indeed reflect the needs of the students, if the materials are designed to teach the objectives, and if the teachers abide by the curriculum, then, the curriculum should hold together well. And in such a situation, the tests clearly bind all the other components together.

In an ordinary Swedish upper secondary school, teachers have to abide by the Swedish national curriculum and the national objectives. The question, in relation to the passage just quoted, is whether these reflect the needs of the students, and particularly those in a CLIL environment.

ASSESSING LANGUAGE OR CONTENT?

Teacher cognition is characterized by a multiplicity of labels, according to Borg (2003), which aim at describing the psychological context of teaching and the way in which instructional practice and cognition mutually inform each other. Borg (2003:91) discusses the ‘symbiotic relationship’ between teacher cognition and classroom practice and notes that:

[L]anguage teachers’ classroom practices are shaped by a wide range of interacting and often conflicting factors. Teachers’ cognitions, though, emerge consistently as a powerful influence on their practices [...] these do not ultimately always reflect teachers’ stated beliefs, personal theories, and pedagogical principles.

Another approach to discussing teachers’ professionalism can be found in the notion pedagogical knowledge (PK), and pedagogical content knowledge (PCK), terms which are often found in research outside language teaching. The terms were introduced by Shulman (1987) and were used to define the what (PCK) and how (PK) of teaching. Shulman’s aim was to combine rather than to dichotomize the two fields of subject knowledge and pedagogy.

Sometimes the curriculum undergoes assessment and course content is compared in relation to the teacher’s *intended*, *enacted* and *assessed curriculum* (Porter, 2004). Alignment between the three is analyzed in order to answer questions whether teachers teach what is tested, whether the content of what is tested matches the content of the intended curriculum or whether the content of the textbook is the same as that of the test. Porter (2004:7) remarks that:

Teachers may teach what they believe is most important, what they think the students are ready to learn, or what is most enjoyable and easy to teach. There are many factors that can and do influence teacher decisions about what to teach.

The next section offers a brief look into assessment in the disciplines, features that also impact teachers’ assessment practices.

3.5 Assessment in the subject disciplines

In this section, an overview is offered in relation to the three subjects of the study, starting with language assessment, thus building on the foundation already laid in this chapter. After that, assessment tradition and practice in biology and history are examined.

3.5.1 Assessing language

Language competence is usually described in terms of *receptive* and *productive* skills, which can be demonstrated in different ways. Malone (2008) describes three periods of language testing by referring to Spolsky's (1977) division into the pre-scientific, psychometric and socio-linguistic approaches. The first represents open-ended tasks such as translation, composition or oral performance, where tests typically consisted of only one or two test items. The reliability of the tests was questioned due to a lack of common standards. During the second period, tests included many but shorter items, focusing on discrete aspects of language, grammar or vocabulary. Multiple-choice, true or false and short-answer questions became popular. The third approach meant a focus on assessing meaningful communicative competence, thus leading to the development of the CEFR.

Brown and Hudson (2002:15f) point to the fact that language acquisition is different from content areas in how these are typically taught and assessed in western educational traditions:

The fact that language is situated and interactional further makes its assessment different from the assessment of content knowledge. [...] The sociolinguistic context of language increases problems in what areas of language are open for testing and has a strong impact on the form that the testing takes.

Shohamy (2008:xiv) argues that theories and practices in language testing have been closely related to definitions of language proficiency. Consequently, the discrete-point testing era presented isolated test items; the integrative era meant discursal language, and the communicative era typically involved interaction and authentic texts. In the performance era, real-life tasks were used; and finally, alternative assessment recognizes the fact that language knowledge is a complex phenomenon, requiring “multiple and varied procedures to complement one another”.

By referring to the change in theories of learning, Brown and Hudson (2002) state that discrete item tests, as seen in the multiple-choice format for instance, were possible as long as language learning was concerned with specific grammar and language skills. When more complex uses of language were aimed for, e.g. pragmatic and sociolinguistic competence, performance testing became more valid, e.g. test items which cause the examinee to

perform in the language and show communicative ability for instance (2002:57).

Bachman and Palmer emphasize that there is no model language test (2010:6): “In any situation, there will be a number of alternatives, each with advantages and disadvantages”. They also point out that if we want to develop language assessments where the use is justified, there need to be justification for multiple qualities (2010:63). “[A] language assessment should consist of language use tasks. In designing language assessments whose use we can justify, it is important to include tasks whose characteristics correspond to those of TLU [target language use] tasks”.

The CEFR (2001:45) states that “each act of language use is set in the context of a particular situation within one of the *domains* [...] in which social life is organized”. The four domains include the personal, the public, the occupational, and the educational domain. Parameters assessing the quality of the language used and its linguistic form involve features such as *fluency*, *accuracy* and *range* of vocabulary.

To conclude, assessment in English today is based on communicative language competence and focuses on the use of language. The European Language Portfolio, henceforth referred to as the ELP, uses “can do-statements” as descriptors for linguistic proficiency, thereby emphasizing the action-oriented approach described in the CEFR, also acknowledging the learner as a central informant (Little, 2009). In spite of the description of language proficiency as language use both in the CEFR and the ELP, a great deal of work remains to be done to increase the engagement of learner agency in assessment, according to Little and Erickson (2015). They point out that “proficiency develops from sustained interaction between the learner’s gradually developing competences and the communicative tasks whose performance requires him or her to use the target language” (2015:124).

3.5.2 Assessing biology

In TIMSS⁴ assessment framework (2011), biology is described as one content domain within the field of science. Parameters for assessment in science, including disciplines such as physics, chemistry and biology, have been

⁴ TIMSS (Trends in International Mathematics and Science Study), reports every four years on students’ achievements in mathematics and science in fourth and eighth grade in countries around the world

ASSESSMENT

identified at different cognitive levels. Consequently, two types of domains are specified: the *content* and the *cognitive domains*. The latter includes skills-based components: to *explain, describe, compare and contrast* and to *relate*, all of which are to demonstrate the student's level of understanding.

TIMSS states that objectives in science are written in terms of *behaviors* to be elicited by items that exemplify the understandings and abilities expected of students. The objectives also represent a range of cognitive processes involved in learning science concepts (2011:50). Cognitive processes and the understanding and use of science concepts, alternatively described as subject-specific language, are integrated in the learning process. The so-called expected behavior is thus what can be seen and assessed.

In terms of progression, TIMSS (2011:84) states:

Reasoning is involved in the more complex tasks related to science. A major purpose of science education is to prepare students to engage in scientific reasoning to solve problems, develop explanations, draw conclusions, make decisions, and extend their knowledge to new situations.

Assessment in biology rests on cognitive processes, as can be seen in Bloom's revised taxonomy (cf. 4.4.2). There is a clear progression in how scientific knowledge and thus reasoning evolve from more basic knowledge of concepts towards the development of more complex cognitive skills.

Corrigan et al (2013) describe various science framework matrixes which have been in use in the American context. Overall, what is to be assessed is described as *knowledge* and *abilities*. During the 1970s, concepts such as knowledge, comprehension, application and synthesis were used, features also found in Bloom's taxonomy. In the 1980s, content areas, thinking skills and the nature of science were used. A decade later, the knowing and doing was described as conceptual understanding, scientific investigation and practical reasoning. The most recent framework from 2009 identifies so-called *performance expectations* in science content areas as well as in science practices. In Swedish national syllabuses, the same definitions appear.

Biology teachers in Sweden at tertiary level state that there is no model biology test format regarding item types, and no standards or guidelines seem to exist whether to use points or grades when scoring. The design of question tests seems to stem from a general educational tradition rather than a disciplinary tradition (personal communication with active teachers). TIMSS uses two question formats, multiple choice and constructed response. In

TIMSS, it is noted that the choice of item format depends on the subject being assessed and the format that best enables students to demonstrate their proficiency.

3.5.3 Assessing history

The knowledge structure in history is not hierarchical as in science. The question is what constitutes historical knowledge, and, consequently what generates more advanced historical thinking. Pace (2011:107) discusses the difference between science and history and regrets the absence of the same “agreement about what should be taught and what constitutes reasonable evidence that it has been learned”. He calls history a “fuzzy discipline”, again compared to science. In an American framework for assessment, historical *events* as well as the use of *specific disciplinary thinking skills* are mentioned (Serve Center, 2006). In a Swedish thesis, Rosenlund (2011) describes the importance of developing competence in thinking historically in order to handle the historical information we are exposed to in society. He notes that a cognitively advanced way of thinking is accompanied by a competence in how historical knowledge is built, which in turn is a tool when understanding a historical process. Rosenlund argues that in history it is important to practice those skills, to think historically.

Stolare (2011) defines two distinct traditions that have shaped the view of the discipline: the Anglo-Saxon and the German-Danish tradition. The goal of the first is to make students “think historically” and the discipline is based on historical concepts. The concepts are referred to as first or second order concepts. First order concepts are concrete and denote historical events, whereas second order concepts are abstract, pointing at meta-knowledge aiming at identifying underlying patterns and cause-effect problems. In Sweden, the German-Danish tradition has dominated, according to Stolare, representing a more holistic view, focusing on historical consciousness and identity. Stolare notes that recently the two traditions have come closer to finding common features. Content knowledge has influenced narrative skills which Stolare believes have been dominant in the classrooms, while objectives actually stipulate meta-skills and second order concepts.

Rosenlund (2011) discusses the difference between substantive historical knowledge and procedural knowledge by applying terms used by Lévesque

ASSESSMENT

(2008). He notes that there is a difference in historical knowledge due to historical thinking skills which become visible in how students ask questions.

Educators in history at tertiary level in Sweden have not been able to identify typical disciplinary assessment forms. Referring to a report by Forsberg and Lindberg (2010), Rosenlund notes that assessment research in the humanities is very scarce. Studies in history and social sciences show examples of standardized classroom question tests with some essay questions (Odenstad, 2010; Rosenlund, 2011)

In the next section, the written assessment mode is examined in a cross-disciplinary and generic way. The purpose of the presentation is to identify and describe common features in the design of test items and writing tasks.

3.6 Written assessment

Assessment outcomes can be presented either in an oral or a written mode. In the current study the focus is on the latter. As seen in the description in Figure 1 (section 1.4), two general formats are used for the purposes of this investigation, i.e. what are referred to as question tests and production tests. Below follows a presentation of the two formats. Written assessment is a broad topic in its own right, and the presentation below can only offer a selection of features, relevant for the present study.

3.6.1 Question tests

In the same way as test types relate to a teaching approach or theory of learning, different categories of test items or question types relate to different types of assessments. Item formats are often dichotomized into constructed response (CR) or selected response (SR), but, as Hogan (2013:2) points out, “what gets classified into each category is not always the same from one source to another”. In this study, the classification of Popham (2011), McMillan (2011) and Hogan (2007) is used, categorizing short-answer questions and completion items as CR items, rather than SR, as has been the case in some textbooks on educational assessment.

The table below presents the most commonly found definitions and how they are used in the present study.

ASSESSING LANGUAGE OR CONTENT?

Table 2. Common written test types and question/item types.

Test type	Item type/Example
Selected response tests	Binary choice/True or false Matching Multiple choice
Constructed response tests	Fill-in-the blank/Completion questions Open-ended questions Short answer questions Essay questions
Production tests	Essays Papers Lab-reports Project reports

What are called production tests in this study are labeled essay tests and performance tests by Hogan (2013). He points out that these, together with portfolios, are actually examples of constructed response items. This is true considering the requirements of the students, but in this study the CR items are solely used for test items in the question tests. The test items in the production tests are referred to as writing prompts or tasks.

In assessment literacy programs, teachers are suggested to write a short statement when test items are written, describing the *skill*, the *thinking process*, or the *strategy* required in order to answer the question. This statement, called an *item descriptor*, represents a point of learning. When item descriptors from a unit of questions are ordered by difficulty, the sequence of learning becomes clear (Brown & Hudson, 2002). In a manual for language test development, teachers are asked to consider all the competences needed to accomplish a successful response: “The task should elicit sufficient appropriate language for a judgment to be made about the test taker’s ability in the chosen competence(s)” (ALTE, 2011:14).

3.6.2 Production tests

As previously mentioned, production tests refer to fairly long essays initiated by writing prompts. Since the written texts in the current study represent a broader repertoire of texts covering both narrative, exploratory and argumentative essays as well as laboratory and project reports, an overarching label was needed (cf. section 1.2). The term essay will appear in the discussion as well, this being a term used in the literature.

ASSESSMENT

Written skills include many overlapping competences. In order to produce a readable, communicative and qualitative text, the author needs to possess *language knowledge* but also *strategic competence* (Bachman & Palmer, 2010). The complexity of required writing skills becomes even more apparent when looking at the multiple aspects in Wang and Wen's (2002) model. To exemplify, they divide language knowledge into organizational and pragmatic knowledge. The first category is further divided into grammatical and textual knowledge, where the first is defined by knowledge of vocabulary, syntax and phonology. The second, textual knowledge, consists of cohesion and rhetorical and conversational organization, including how to write different types of texts. Text types, or genre, represent a teaching pedagogy which is presented in more detail in Chapter 4.

The CEFR likewise accounts for several different competences needed to represent written language proficiency, e.g. linguistic competence defined by lexical, grammatical, semantic, and orthographic competence (cf. CEFR, p.109), as well as socio-linguistic and pragmatic competence. Research shows that students are more motivated to write and make accelerated progress when they are given clear instructions about what a quality performance looks like and they know how they will be assessed (Hyland, 2007). Once again, the teacher has to be clear about the motives for choosing one or the other written format since the construct may vary between writing skills or subject knowledge, as in question tests, or both (Tardy, 2006).

Alderson and Banerjee (2002) note that essay writing used to be questioned due to the threat of subjective grading and the lack of control that a prompt or task would elicit in the target language. They refer to the current view that writing ability is more than accuracy in vocabulary and grammar, it also includes aspects of discourse structure. Concerns about appropriate scoring of the extended writing has raised questions regarding the design and application of scoring procedures. Alderson and Banerjee conclude that the more structured the writing task and the scoring criteria provided, the more reliable the assessment.

Figure 3 below summarizes the description of different written assessment types. It shows a continuum representing the progression and complexity levels of different CR test items.

ASSESSING LANGUAGE OR CONTENT?

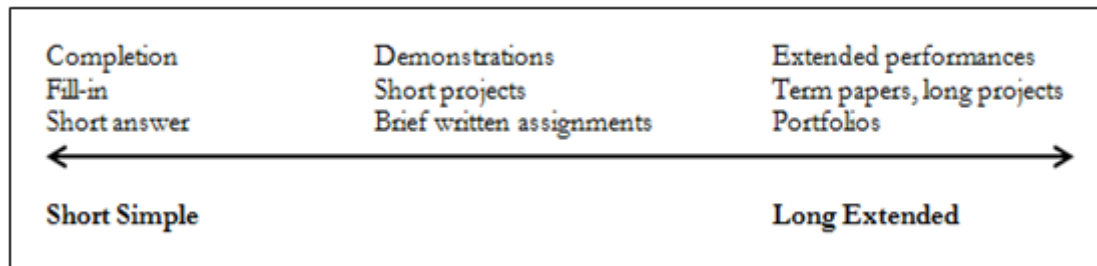


Figure 3. Continuum of constructed response test-items.

Figure adapted from Hogan, 2013

In the next section, validity in assessment and related concepts, such as reliability and generalizability, are discussed. Threats to validity are presented in relation to a validation model.

3.7 Validity in assessment

In this study, evidence of validity is sought in the teachers' pedagogical orientation as expressed in interviews and in their assessment practices as seen in written assessment samples. A validation tool used in the analysis and discussion of the assessment practices in the current study is presented below.

In an article on teaching and assessment, Erickson (2010) refers to students' implicit definitions of and views on validity, which include relevance, authenticity and construct coverage. In the CEFRL, validity is described as a measurement of how well the assessed qualities, the construct, correspond with what the tasks set out to assess. For instance, in a language test language knowledge and strategic competence represent two separate constructs (Bachman & Palmer, 2010). Bachman and Palmer comment that "the way we define the construct will have clear implications for the [...] method to be used" (2010:212); or should have implications, one might add. They further acknowledge that the way the construct is defined will guide the process that follows (2010:215).

Construct validity was first used by Crohnbach and Meehl (1955) in relation to psychological tests and was later developed by Messick (1989), to define the overarching validity concept. Messick identifies two major threats with respect to construct validity, namely *construct under-representation* and *construct-irrelevant variance*, the first one signifying too little of what is supposed

ASSESSMENT

to be measured, whereas the second implies that interfering factors affect test scores or the outcome of an assessment. In a CLIL context, interfering factors might relate to the student's use or misuse of linguistic elements which inappropriately impacts the weighing of test results.

Crooks et al (1996), comment that validity relies heavily on human judgment, as does assessment, and so can be hard to carry out and defend. Validity, however, according to Messick, is about finding evidence to support "the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (1989:13). Even though validity in assessment is desirable, it is important to note that it springs from interpretations of inferences made from tests scores or observations, where an element of subjectivity is always present. However, the better educated, i.e. assessment literate, the assessors are, the more valid the interpretations will be. Reliability and generalizability refer to more quantitative aspects that constitute common measures of quality in assessment, as will be demonstrated in the model introduced below.

3.7.1 A validation chain model

Crooks, Kane and Cohen (1996) suggest a validation chain model in eight steps which take different threats to the validity of assessment into account and can be used on existing assessment tasks. The model builds on the validity argument approach of Kane (1992) and Shepard (1993), whose models in turn build on earlier work by Cronbach (1988). Most of the threats have been identified in previous research, but never placed in such a structured model, according to Crooks et al. They argue that validation can only take place if the purpose of the assessment is well understood and the strength of each link in the chain depends on the appropriateness of the tasks to these purposes.

The importance of each link, as well as which threats apply, depends on context. Crooks et al (1996:267) suggest that the user of the model needs to identify "further threats which are associated with their particular assessment context". In the present study, for instance, the particular role of language in the assessments used in subject content courses needs to be taken into account. Validity is claimed to be limited by the weakest link, which needs to be identified in the assessment material in the current study. The model was found to be suitable for the validation of the assessment tasks in the current study, due to the transparency of the threats which have been identified in

ASSESSING LANGUAGE OR CONTENT?

relation to each of the links. The same threats apply to the validity of assessment in the CLIL context.

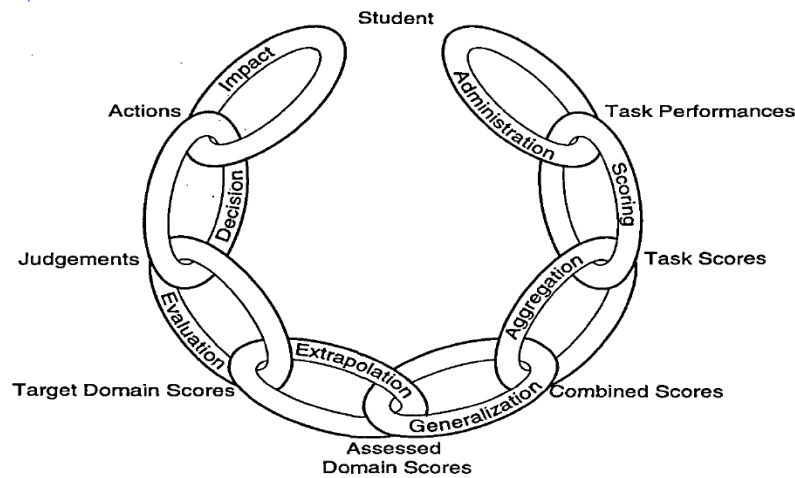


Figure 4. The validation chain model, suggested by Crooks, Kane and Cohen (1996).

The chain model starts with administration followed by scoring, aggregation, generalization, extrapolation, evaluation, decision and impact; see Figure 4 above.

As mentioned previously, many different threats can be identified, with each step depending on context. Below follow some examples used by Crooks et al (1996) to clarify how the model can be applied. The model will be used in the analysis of the assessment material in the present study and subsequently a clearer picture will appear in the analysis chapter, Chapter 7.

The *administration* link is the first, and has to do with the implementation of the assessment and the task performance. Possible threats to the validity of this link might be if the student receives too little time on task, fails to understand the instructions, or if the student is unmotivated or suffers from assessment anxiety.

Threats associated with the second link, *scoring*, consist of undue emphasis on some criteria, or scoring which fails to capture important qualities, issues which relate to what Messick (1989) labels construct under-representation and construct-irrelevant variance. Crooks et al (1996:272) use an example which is valid in a CLIL context, where undue emphasis on students' written expression, grammar and spelling, "might be doing an injustice to students whose knowledge of history and skill in historical analysis are strong, but who are poorly equipped to write well in English".

ASSESSMENT

The validity of the third link, *aggregation*, can be threatened by an inappropriate balance between different tasks, which may occur if an assessment involves two different test items, an essay and a multiple-choice test, and these are equally weighted even though the difficulty of the abilities cannot be compared, an example used by Crooks et al.

The fourth link is *generalization*, often identified as generalizability in validity discourse. It is closely linked to the reliability of a task, i.e. if too few tasks are used to represent the assessed domain. Also, failure to control for different variables, like those mentioned under the first link; time on task and task format, for instance, constitutes a threat to the generalization of an outcome.

Extrapolation is the fifth link and represents a wider sample of assessment tasks than discussed in relation to generalization. It has to do with the relationship between the assessed domain and the entire target domain: “The degree of risk to the validity of the extrapolation process varies inversely with the degree to which the assessed domain covers the target domain” (Crooks et al, 1996:275). It also relates to Messick’s wider concept construct under-representation, which can be seen in scoring and generalization as well. The assessment represents a sample of the target domain, and the question is whether the assessment succeeds to give a valid sample performance of the full range of tasks in the target domain.

The figure below (Figure 5) illustrates how the assessed domain represents a condensed sample of the entire target domain for learning (to the right). The goal is for the selected test items, or task types (1.) and number of items (2.) to represent a valid sample of the target domain (3.) so that an *extrapolation* can be made from the inferences made of the assessment to the universe score of the entire domain. The target domain refers to the course goals and may represent a thematic content unit and/or a skill. The scores on each of the individual test items of an assessment (first arrow to the left) are *aggregated* to produce a combined score for the sample of tasks used in the actual assessment or exam, which can then be *generalized* from the specific assessment to represent the assessed domain, since no single test could cover every aspect of the objectives.

ASSESSING LANGUAGE OR CONTENT?

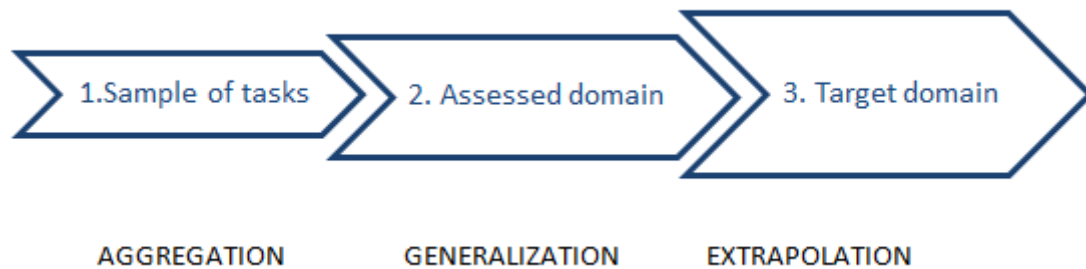


Figure 5. Illustration of the relationship between aggregation, generalization and extrapolation and target domain.

The sixth step represents *evaluation*, which means forming judgments about the student's performance. The understanding and interpretations made by the person evaluating the assessment, i. e. the teacher, can be a threat to its validity. One threat consists of teacher bias, e.g. additional knowledge about a student affects the interpretation, or teachers make comments about the wrong construct; that the student “writes well”, when the construct is knowledge of scientific concepts.

The seventh link involves *decisions* based on the judgments made, which could be referred to as the *washback effect* of an assessment (see 3.4). Crooks et al (1996) identify a threat to validity which relates to “inappropriate standards”. They describe how definitions should be available for what grade can be awarded for a given assessment score. This means that teachers need to be able to provide this information when asked, which requires insight and transparency. In a CLIL context, this can be problematic when no standards are articulated regarding the role of language in the assessments. Even if language is not explicitly accounted for in the judgment, its effects on the capacity to communicate content knowledge need to be addressed.

The last link is *impact*, and will not be immediately discussed in this study, even though it is an important issue in relation to the purpose of assessments, or as Crooks et al (1996:279) put it: “The effort involved in the assessment process can only be justified if the assessment leads to worthwhile benefits for students or other stakeholders”.

The user's awareness of the theoretical foundation of a discipline and its goals is crucial for the validity of the assessment. In the CLIL context, there needs to be an awareness why certain assessment procedures are appropriate or not, and if the purposes are identical to those of a non-CLIL context. What, then, are the intended outcomes regarding language and the possible

interferences of language on content knowledge? How the role of language, as a tool or a goal in itself, is conceived in assessments have implications for the validity of the test items. Possible threats to the validity of CLIL assessments need to be identified and accounted for, a discussion which is brought up in Chapter 7, as a result of the present study.

3.8 Summary

It has already been noted that validity is an overarching and foundational feature in establishing good practice in assessment. Validity has been said not to reside in the test itself, but in the inferences, decisions and actions based on test outcomes (Moss et al, 2006), directly pointing to the teacher who makes those inferences, and often the tests as well. Stiggins (1995:240) claims:

Assessment-literate educators [...] come to any assessment knowing what they are assessing, why they are doing so, how best to assess the achievement of interest, how to generate sound samples of performance, what can go wrong, and how to prevent these problems before they occur.

Teachers' responsibility in assessment cannot be overestimated. They are carriers of personal values and disciplinary history, executors of teaching pedagogy and learning culture in their educational context and agents of assessment. They have to be reflective and open to test impact to be able to provide students with necessary scaffolding and accommodations without jeopardizing the validity and reliability of test scores. In the following chapter, however, the focus will not be so much on the teachers, but on language and content in the disciplines and, consequently, in assessment.

4. LANGUAGE AND CONTENT IN THE DISCIPLINES

4.1 Introduction

Learning a language means relating to subject content in the same way as learning subject content necessitates language in order to communicate. Yet, language courses and subject content courses are often treated as separate entities in educational contexts, as pointed out by Mohan (1986): “In subject matter learning we overlook the role of language as a medium of learning and in language learning we overlook the fact that content is being communicated.”

This chapter starts by looking at linguistic and cognitive skills in combination, the way they are materialized in the Common European Framework of Reference, CEFR, and the course goals for the subject content courses in biology and history. Next follows a section on language, discussing form and function as well as different types of language involved in the disciplines. The notion of translanguaging and interlanguage are briefly presented, before moving on to content and a look at Bloom’s revised taxonomy and lower and higher order thinking skills. The chapter concludes by combining the learning of language and content, looking at genre pedagogy and Cummins’ quadrant, combining cognitive and linguistic demands.

4.2 Linguistic and cognitive skills

When investigating children’s development, a key issue concerns in what ways their cognitive development is influenced by their access to language (cf. Siegal & Surian, 2012). The same concern, regarding the relationship between linguistic skills and learning content, is a focal question in research done among bilingual students (e.g. Cummins, 2000). As children grow older, the role of language within the disciplines curiously seem to become less prominent, as noted in CLIL contexts. However, curricula and syllabi for

language courses as well as subject content courses reveal the co-existence of descriptors including communicative skills as well as academic knowledge. Below follow descriptions both from the CEFR for the teaching of foreign languages and from the Swedish national course objectives in the subject disciplines involved: biology, history and English as a foreign language.

4.2.1 Descriptor words in the CEFR

Descriptors from the CEFR have been used when formulating the Swedish national objectives for the English language courses in upper secondary school (NAE, 2012a). The CEFR (2001:11) focuses on *communicative language competence* and acknowledges language use and learning in combination with academic knowledge in a professional field:

All human communication depends on a shared knowledge of the world. As far as language use and learning are concerned, the knowledge which comes into play is not directly related exclusively to language and culture. Academic knowledge in a scientific or technical educational field, and academic or empirical knowledge in a professional field clearly have an important part to play in the reception and understanding of texts in a foreign language relating to those fields.

Consequently, the CEFR addresses the fact that language and academic knowledge are integrated in reception and understanding as well as in production and interaction. The CEFR (2001:160) recognizes that different competences are activated due to the different components and features of a task:

The learner's different competences are closely related to individual characteristics of a cognitive, affective and linguistic nature which need to be taken into account in establishing the potential difficulty of a given task for a particular learner.

The difficulty of a task relates to the cognitive, linguistic and affective characteristics of the task as well as to the learner's competences. The descriptors in the CEFR use "can do-statements" to describe the proficiency level of the learner and how language can be used in different tasks and situations. A six-graded scale is used, ranging from A1 to C2, describing progress from basic, to independent to proficient user of the language. In the present study, the students in the CLIL courses have attained level B1 during their first year of upper secondary school, and B2 during their second. Those

LANGUAGE AND CONTENT IN THE DISCIPLINES

levels are supposed to represent “entry level”, or, in other words, students with the lowest grade. The more proficient students can be expected to attain level C1, and possibly even C2 before graduating. The CLIL context typically attracts students with relatively high proficiency level (Sylvén & Ohlander, 2014), which will be discussed later.

As an example of descriptors appropriate for the CLIL context of this study, a couple of descriptors for written production, representing levels B1, B2 and C1 (CEFR, 2001:61f) are presented. The CEFR makes a distinction between different types of written production. The first example for each level below is found under *Overall written production* in the CEFR, and the second under *Reports and essays*.

B1: Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence.

Can write very brief reports to a standard conventionalised format, which pass on routine factual information and state reasons for actions.

B2: Can write clear, detailed texts on a variety of subjects related to his/her field of interest, synthesising and evaluating information and arguments from a number of sources.

Can write an essay or report which develops an argument, giving reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options. Can synthesise information and arguments from a number of sources.

C1: Can write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion.

Can write clear, well-structured expositions of complex subjects, underlining the relevant salient issues. Can expand and support points of view at some length with subsidiary points, reasons and relevant examples.

Without venturing a detailed analysis of the descriptors quoted above, worth noting is first the progression in cognitive demand in relation to the topic: going from familiar subjects (B1), to a variety of subjects (B2), in both cases

ASSESSING LANGUAGE OR CONTENT?

related to the field of interest, to complex subjects (C1). Second, the progression in the requirements of the student can be seen in features such as *connected texts, linking shorter elements* (B1), *synthesizing, evaluating, developing an argument* (B2), *expanding and supporting lengthier points of view* (C1). Cognitive and linguistic requirements merge, as well as subject content and language.

In the following sections a brief outline is given of the combination of requirements found in the different course objectives.

4.2.2 Course goals

A new curriculum for upper secondary school was introduced in Sweden in 2011. Diversity, but at the same time a holistic approach, is being encouraged. The different disciplines share some common features which can be seen in the examples below. The curriculum acknowledges that it is hard to decide what to teach today, since we do not know what knowledge will be needed in the future, which requires an active discussion about concepts of knowledge (NAE, 2013:6):

Knowledge is a complex, multi-faceted concept. Knowledge can be expressed in a variety of forms – as facts, understanding, skills, familiarity and accumulated experience – all of which presuppose and interact with each other. Teaching should not emphasize one aspect of knowledge at the cost of another.

Each of the disciplines states the *aim of the subject*, the *core content* and the *knowledge requirements* for the different grades, terms used by NAE. When comparing the aims of the three disciplines, common features appear, but also obvious differences revealing the different knowledge structure and aims of the subjects. *Understanding* appears in all of the three disciplines, but with a different meaning and to a different extent:

In biology:

It [the teaching] should contribute to students developing their **understanding** of the importance of biology in society.

Knowledge of concepts, models, theories, working methods of biology, and also an **understanding** of their development.

History teaching should help students develop:

LANGUAGE AND CONTENT IN THE DISCIPLINES

...an **understanding** of how history is used [...]an **understanding** of living conditions of different ages [...] an **understanding** of the present and the ability to orient themselves to the future [...] an **understanding** [...] to assess how different people and groups in space and time have used history [...] develop an **understanding** of historical processes of change in society.

In English as a foreign language:

Students should be given the opportunity, through the use of language in functional and meaningful contexts, to develop all-round communicative skills. These skills cover [...] reception, which means **understanding** spoken language and texts

Looking at the examples above, it becomes quite clear that understanding is especially prominent in history. History is a discipline which aims at making *interpretations* and developing a historical awareness, whereas in biology interpreting is not a key feature. In biology, understanding relates to the importance of the subject itself but also represents a more evolutionary perspective on cause and effect. In the English language course, understanding refers to receptive skills.

Language and communication is present in the aims of all the disciplines:

[...]to **communicate** using scientific language. (biology)

The ability to use knowledge of biology to **communicate**, and also to examine and use information. (biology)

Through teaching students should be given the opportunity to present the results of their work using **various forms of expression**, both orally and in writing (history)

The ability to use different historical theories and concepts to **formulate**, investigate, **explain** and draw conclusions (history)

In the aims of the English course goals it says:

Teaching of English should aim at helping students to develop **knowledge of language and the surrounding world** so that they have the ability, desire and confidence to **use English** in different situations and for different purposes.

In this description of aims, it is apparent that in order to develop communicative skills, students need to develop both knowledge of the language and of the surrounding world, to be able to use the language in

functional and meaningful contexts. The English national syllabus further acknowledges that:

Teaching should encourage students' curiosity in language and culture, and give them the *opportunity to develop plurilingualism where skills in different languages interact and support each other*. Teaching should also help students develop **language awareness** and knowledge of how a language is learned through and outside teaching contexts.

The English syllabus consequently encourages translanguaging, which is in line with what has been suggested for CLIL, i.e. to deepen the awareness of target language as well as mother tongue (e.g. García, 2012).

Students are supposed to “develop correctness in their use of language in speech and writing, and also the ability to express themselves with variation and complexity”. The first point of the core content in communication in the language course mentions “subject areas related to students’ education”, which may imply subject content courses, particularly those within the educational profile, often in the natural or the social sciences.

As can be seen from the examples above, language and communication are present in the aims of all three disciplines. At the same time, the objectives for the English language course stress the use of language in meaningful contexts, exemplified by subject content from the students’ content courses. The CEFR encourages the same integration of language competence and academic content. In the next section, the learning of linguistic forms, functions and registers is discussed in isolation, before looking into the knowledge structure and cognitive demands of the disciplines.

4.3 Learning language

Learning a language is often interpreted as the learning or acquisition of a *foreign* language. In primary education, however, the main goal is to become literate in the first language, L1. Nevertheless, in education the learning aim relates to the current proficiency level and age of the students, regardless of whether the goal is to communicate using basic language or to produce academic texts. Immigrants find themselves in an intermediate position: they are supposed to take part of relevant instruction at their cognitive level without always being proficient in the target language. The same issue is discussed in CLIL contexts. In foreign language teaching and learning, focus on form or focus on function represents a vast topic in its own right. The aim

in CLIL is to have a dual focus and regard language both as a tool *and* an object of study, and so focus on form and explicit teaching should be included in a CLIL context (Dalton-Puffer, 2011; Llinares et al, 2012; Pérez-Vidal, 2007; Wewer, 2014).

The chief focus here will not be on language learning theories, but rather on vocabulary, meaning and use in relation to different disciplines and thinking skills, in other words, in relation to content, communication and cognition, to borrow Coyle's categorization. The theoretical base can be found in communicative language teaching (CLT) and pragmatics (e.g. Canale & Swain, 1980; Hymes, 1971; Krashen, 1989, 2008), i.e. what people do with language, learning to use appropriate language in context. The classic quote from Hymes (1971:278) lends itself well to synthesizing the approach, stating that there are "rules of use without which the rules of grammar would be useless".

4.3.1 Form versus function

Cummins (2000) explains the development of academic expertise in terms of three dimensions: focus on meaning, language and use. The first one, *focus on meaning*, has to do with receptive skills and making input comprehensible. The second focuses on *language*, and includes an awareness of language forms and uses as well as a critical analysis of these notions. The last dimension, *use*, involves using language to generate new knowledge, create literature and art as well as acting on social realities, which could be compared with the affordances of a language, i.e. perceiving, interpreting, making sense and possibly acting in response to the environment (e.g. Gibson, 1979; Van Lier, 2004).

At upper secondary or tertiary level in Swedish education, it sometimes seems as if teachers believe that students' English proficiency is so high that there is no need to focus on making input comprehensible, or to discuss semantic nuances; students are supposed to be ready for language use (Airey, 2012). Previous research suggests that there is a need in the CLIL context to focus more on meaning, uses and the development of critical literacy, as well as target language forms, but not so much grammatical progression as in traditional FL teaching (Coyle, 2010).

ASSESSING LANGUAGE OR CONTENT?

In an academic setting, the languages within a language have to be identified, often referred to as different registers or genres, as will be seen below. Schleppegrell and O'Hallaron (2011:3) note:

Academic language refers to the disciplinary registers that students encounter in the secondary years, and using academic language calls for advanced proficiency in complex language across subject areas, posing challenges for teacher preparation.

Edlund (2011) leans on his own work as well as on previous research when concluding that the increased degree of exposure to language, according to Krashen's input hypothesis, is insufficient for CLIL instruction to successfully contribute to students' linguistic development in academic registers. He claims that there is a need to develop more genre awareness in students' English. He also suggests that this presupposes a systematic focus on genre and registers in the teaching, with teachers scaffolding and modelling language by targeting linguistic form in the interaction with students (2011:99).

4.3.2 Types of language

Yoxsimer Paulsrud (2014) notes in her study on EMI (English-Medium Instruction) in Swedish upper secondary schools that teachers as well as students seem to be unaware of the difference between academic language and everyday language. Academic language often refers to disciplinary registers and tends to be more cognitively challenging, whereas everyday language tends to involve more contextual clues. Academic language is not uniform. Subject courses involve subject specific disciplinary language, as well as general academic language related to cognitive skills. Consequently, it is important to note that multiple registers are involved in all disciplines. It is not only a matter of an L2, in this case English, versus an L1, in this case Swedish.

Coyle, Hood and Marsh (2010) distinguish three different types of language to be acknowledged in the CLIL environment: language *of*, *for* and *through* learning. The language of learning could be compared with disciplinary language involving subject specific concepts. The language for learning represents a more universal type of language, requiring specific linguistic skills. The last type, language through learning, is a combination of language use and cognitive processes which includes both BICS and CALP, acronyms coined by Cummins (1984). CALP, i.e. cognitive academic language proficiency, refers to the ability to think in and use a language as a tool for learning,

LANGUAGE AND CONTENT IN THE DISCIPLINES

whereas BICS, i.e. basic interpersonal communicative skills, is a contextual and cognitively undemanding language, used in informal settings (Cummins, 1984).

Another way of labelling the registers involved in the classroom was introduced by Snow, Met and Genesee (1989), stemming from CBI-teaching, where a distinction is made between a *content-obligatory*, (CO) language and a *content-compatible* (CC) language. The first, CO, could be compared with the language of learning referred to above, or the academic register. In the table below, some of the features of the two registers, CO and CC, are presented:

Table 3. Content-obligatory versus content-compatible language

Content-obligatory (CO) language	Content-compatible (CC) language
Necessary to learn the key content concepts	Expands the language beyond academic forms and functions
Primary, usually generated first	Provides extra language, or “filler”
Content- or discipline specific, more academic in nature	Include more communicative forms and functions
What-oriented, the “what” being the content	How-oriented, more than what
Required to learn for success with the assessment	Complement and supplement the CO-language

Some features of language objectives for the CBI-classroom. Adapted after Fortune and Tedick (From the CoBaLTT Project website: <http://www.carla.umn.edu/cobaltt/modules/>)

In tertiary education, the labels English for Special/Specific Purposes (ESP) or English for Academic Purposes (EAP) are used to describe disciplinary language which relates to certain educational or vocational areas and requires training to use domain-appropriate language. The emergence of ESP/EAP can be traced to developments in linguistics, with a new focus on the ways in which language is used in real communication (Hymes, 1971). This also meant appreciating the ways in which written and oral language vary, and how different situations require different variants of English (Gatehouse, 2001).

4.3.3 Translanguaging and interlanguages

When acquiring a language, the learner cannot possibly be fluent in all the registers from the start, which in a CLIL context means that the student moves between different levels of accuracy and fluency in order to become functional regarding both CO and CC, while very likely mixing registers. Yoxsimer Paulsrud (2014:33) refers to the sense-making process as language alternation. “This concept of using language to learn language [and content] can be extended to using all one’s language resources for learning and even alternating languages in the process of learning, moving from language to translanguaging.” Translanguaging, a term used by García (2012), has become more used in language teaching. Yoxsimer Paulsrud (2014:33) prefers not to use the term “code-switching”, claiming that “translanguaging offers a move away both on the focus of language as a code to a focus on the speakers in context and how they use language”, a view shared in this study.

Olander and Ingerman (2011) explore the role of *interlanguage* as a hybrid language in the science classroom, where everyday expressions are seen as a resource while students work on making sense of the scientific language.

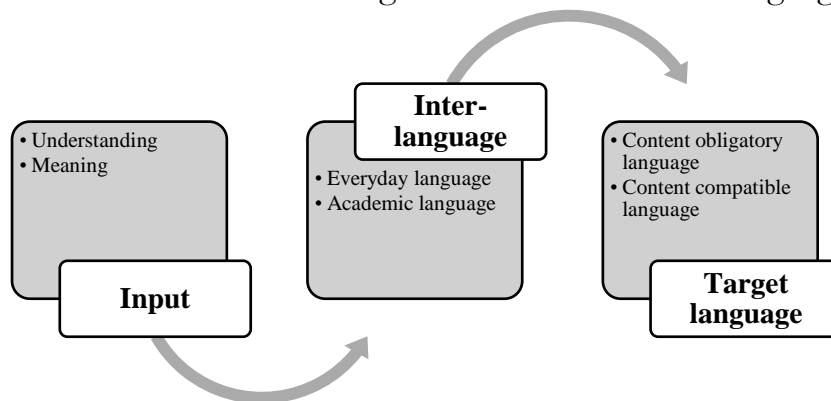


Figure 6. Model of language exposure and language use in CLIL. Sense making and acquisition process.

The above figure demonstrates how students need to first receive comprehensible input to make sense and understand the content of a course. Before acquiring the appropriate and accurate target disciplinary language, the students need to discern and use the different registers, including both academic and everyday language as well as CO and CC language. The intermediate phase may signify a period of hybrid language, an inter-language, before acquiring the accurate academic register, and learning the associated course content, which will be assessed.

If a constructivist view is adopted on assessment, where assessment can be seen as another opportunity for learning, the role of hybrid languages could be acknowledged by teachers as part of the learning process. This would include both the subject-specific academic language and the more general academic language. In a CLIL context, and in a study as the present, the possible combination of registers seems fruitful, even though the goal is to acquire the target language. In assessment, the progression toward target forms can be recognized, especially if production tests and writing assignments are used.

4.4 Learning content

Course content is an interesting concept, since it raises the question what “content” signifies in different disciplines. A historian may claim that it represents different eras or the skill to make interpretations, whereas a biology teacher might suggest scientific concepts or forming hypotheses. A language teacher may say grammar or literature. Coyle, Hood and Marsh (2010) state:

It is useful to think of content in terms of the knowledge, skills and understanding we wish our learners to access, rather than simply knowledge acquisition.

It is important to note, as in the above quotation, that content is not limited to knowledge or facts; it may also represent skills and cognitive processes, all of which presuppose language as a tool. Different disciplines have different traditions which affect the way knowledge is perceived, and consequently what should be taught and assessed.

4.4.1 Knowledge structure and epistemology

Views of learning in different disciplines hinge upon epistemology, teaching tradition and knowledge structure. Some of this has already been mentioned in conjunction with assessment (cf. section 3.5). The natural sciences are predominantly hierarchical in their structure, according to Airey (2012), who draws on the concepts of Bernstein (1999). The hierarchical structure in the sciences is contrasted with the humanities which are described as having a predominantly horizontal knowledge structure, according to Airey (2011:68): “here it is the new perspectives offered by these new descriptive languages [the academic registers] that provide the development”. Airey suggests that there is a potential conflict when a teacher of science is supposed to teach the

language of a course, especially in a CLIL setting, since two different knowledge structures intersect. Morgan (1999:30) notes that history is “a subject suitable for bilingual teaching”, since the terminology is less technical than in a science subject, which could mean less of an obstacle when communicating, which may explain the smoother fit between history and bilingual teaching in CLIL, according to her.

4.4.2 Thinking skills and Bloom’s revised taxonomy

In the natural sciences, but also in the humanities, Bloom’s taxonomy is often used as a model of reference across different disciplines. In the new revised taxonomy (Anderson & Krathwool, 2001) the descriptor nouns are changed into action verbs, going from remembering at the bottom, to understanding, applying, analysing, evaluating, to creating at the top. See Figure 7 below.

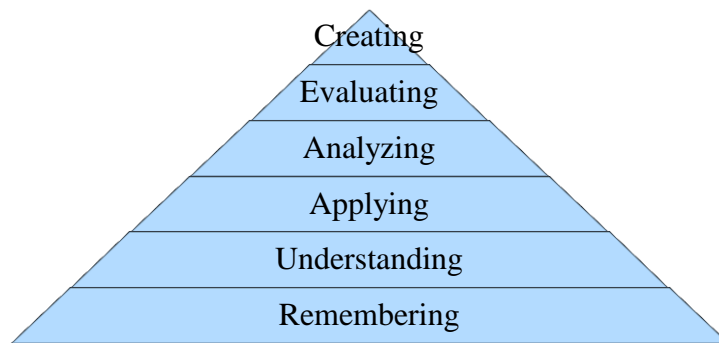


Figure 7. Bloom’s revised taxonomy

All the levels are present to some degree in the objectives of the disciplines involved in this study. Thus, they are also present in the way teachers construct assessment items, as will be seen in the description of assessment samples in Chapter 6. The levels are closely related to different thinking skills and, consequently, with the question words used.

Lower order thinking skills (LOTS) represent the two basic levels in Bloom’s taxonomy, comprehension and knowledge, which involve remembering and understanding. Typical questions asked concerning these thinking skills are *what*, *when*, *where* and *which* questions. The other levels in Bloom’s revised taxonomy involve higher order thinking skills (HOTS) and include *how* and *why* questions, which require the use of more complex language: “In CLIL contexts, and especially in science subjects, learners often have to answer higher order thinking questions at an early stage of learning curricular content” (Cambridge ESOL, 2008:5).

4.5 Learning language and content

In CLIL, where the word “integration” is used, the central idea is “fusing goals” between language and subject, according to Coyle, Hood and Marsh (2010). Dalton-Puffer (2007:5) notes, however, that despite the word “integrated” in CLIL, there is a “good deal of tension and sometimes conflict between the two areas”, which she claims seems to stem from the competition of the primacy of one over the other. Gajo (2007:564), states that integration is a “complex interactional and discursive process relevant to both the language(s) and the subject”. Whether it is a matter of building bridges between two areas or identifying existing common denominators, integration is a cross-disciplinary process, cutting through all disciplines. Below, a couple of pedagogical perspectives or tools for an interdisciplinary strategy are presented.

4.5.1 Genre pedagogy across the curriculum

Genre pedagogies are concerned with how language is structured in particular contexts of use. Language is seen as a tool to achieve social purposes in various types of text for different purposes in different environments outside of school (Halliday, 1994; Hyland, 2007). According to Hyland (2007), genre pedagogies enable teachers to ground their courses in the texts that students will have to write in their target contexts and genre approaches see different ways of writing as “purposeful, socially situated responses to particular contexts and communities” (Hyland, 2003:17). Genre pedagogy identifies certain shared linguistic features in specific texts which can be taught, whereas writing process approaches focus on cognitive development in the writer. Without advocating one approach over the other, from an integrative CLIL perspective genre pedagogy incorporates features which lend themselves to an interdisciplinary perspective. As Hyland (2007:149) points out:

The old certainties of cognitive homogeneity which supported process writing models for so long are no longer sustainable, and there is an urgent need for more theoretically robust, linguistically informed, and research-grounded text descriptions to bridge the gap between home and school writing and prepare learners for their futures.

Setting goals for classroom teaching is about finding relevant tasks in order to prepare students for the future, academically as well as professionally. CLIL is, as stated above, about fusing goals between content courses and language

ASSESSING LANGUAGE OR CONTENT?

courses. The question for instructors is how to do this in pedagogy and practice. If genre pedagogies look outside to distinguish target contexts to facilitate learning, Cummins looks inside the learner to describe how context embeddedness supports cognitive processes and thus linguistic skills.

4.5.2 Cummins' quadrant

Cummins' quadrant, also known as Cummins' matrix (Cummins, 2000), offers a way of integrating language and knowledge acquisition, by combining a continuum ranging from cognitively undemanding skills at the bottom, to cognitively demanding skills at the top on the vertical axis, with an intersecting continuum representing the degree of context embeddedness on the horizontal axis: more context embedded to the left, meaning more contextual support, and less context embedded to the right, making it more difficult due to the lack of supporting cues. The two intersecting continua that illustrate Cummins' two dimensions of degree of context and degree of cognitive demand can be arranged in such a way that they form four quadrants characterizing language and learning activities. Figure 8 illustrates one of the original variants of the quadrant (the order of A-D, and the position of the cognitive demands may vary across variants):

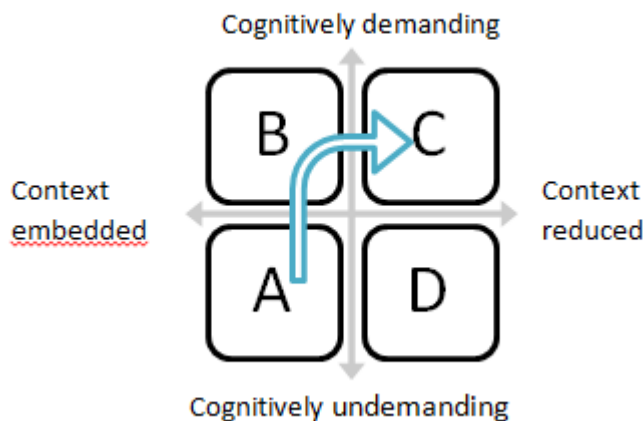


Figure 8. Cummins' quadrant: an example of one of the original versions.

Quadrant A: Cognitively simple tasks with support from context, used in everyday communication. Help found in a picture, a prompt, discussion, teacher or peer support.

Quadrant B: Cognitively demanding tasks, but with support from context. This is where most of the tasks should be to help bilingual students.

LANGUAGE AND CONTENT IN THE DISCIPLINES

Quadrant C: The goal: cognitively demanding, but ultimately not as much need for contextual support.

Quadrant D: Fill in the blank: tasks with no or very limited cognitive demand and little context. This quadrant should be avoided.

The model is intended for teachers when planning lessons and lesson content in order to consider the appropriateness of different tasks to help, especially, bilingual students to get the right linguistic input at the right cognitive level. The progression is supposed to move from A, to B and on to C. No cognitively undemanding or context-reduced tasks (quadrant D) should be used, since context together with more cognitively demanding tasks offers more learning opportunities. Cummins actually argues that the meaningful context is reduced when tasks are broken down into isolated parts (Cummins & Swain, 1986).

In the matrix below adapted by Coyle (1999), linguistic demands have been added to the model.

<u>High cognitive demand</u>		
	Generalises Compares and contrasts Summarises Plans Classifies Transforms Recalls and reviews Seeks solutions	Argues a case Identifies criteria Develops and sustains ideas Justifies opinions Evaluates critically Interprets evidence Makes deductions Forms hypothesis Predicts results Applies principles Analyses and suggests solutions
<u>Context embedded, Linguistically undemanding</u>	Reading for specific information Identifies, names, matches, retells Transfers from one medium to another Applies known procedures Describes observations Sequences Narrates with start, middle, end	<u>Context reduced, Linguistically demanding</u> Parrots, repeats Copies Reproduces information
<u>Low cognitive demand</u>		

Figure 9. Alternative of Cummins quadrant

Matrix adapted after Cummins (1986) and Hall (1995) in Coyle (1999)

Coyle claims that the challenge for teachers is to create cognitively demanding tasks, yet using less demanding language, as in quadrant B above (upper left corner). In the present study, the features of the matrix will be used in the analysis of the test items (cf. section 6.3).

4.5.3 Language in all the disciplines

The concept of academic language functions are of interest in CLIL settings (Dalton-Puffer, 2007), focusing on how cognitive thinking skills can be identified in language manifestations. The academic language functions are similar to cognitive descriptor words, as seen in the quadrant above, describing what to expect of a performance in an academic task. At a very basic level, we find words such as *define* and *describe*, whereas *analyze* and *argue* are used at more advanced levels (cf Dalton-Puffer, 2007, NAE 2012a, course plans). For this study these descriptor words are defined as *academic function words*. These are compatible with the levels in Blooms' revised taxonomy (Anderson & Krathwool, 2001), as can be seen in Table 3 below:

Table 3. Comparison of academic language functions, cognitive skills from Bloom's taxonomy and Cummins' language proficiency levels.

Bloom's revised taxonomy cognitive descriptors	Cummins' cognitive and linguistic levels
Remembering: Recalling, recognizing, listing, describing, retrieving, naming	A. Identifies, names, matches, retells, applies, describes, sequences
Understanding: Explaining, interpreting, summarizing, classifying	B.
Applying: Implementing, carrying out, using, executing	Generalizes, compares, contrasts, summarizes, plans, classifies, transforms, recalls, reviews, seeks solutions
Analyzing: Comparing, organizing, deconstructing, interrogating, finding	C.
Evaluating: Checking, hypothesizing, critiquing, experimenting, judging	Argues a case, identifies criteria, develops and sustains ideas, justifies opinions, evaluates critically, interprets evidence, makes deductions, forms hypothesis, predicts results, applies principles, analyses and suggests solutions
Creating: Designing, constructing, planning, producing, inventing, generating new ideas	

The first column in Table 3 lists the descriptors in Bloom's revised taxonomy with the bottom level at the top, the reverse order from what is usually the case in the pyramid. The second column names the descriptors found in Figure 8, using the A, B, C, levels as seen in the original matrix in Figure 7. By comparing the academic language function words and the cognitive

LANGUAGE AND CONTENT IN THE DISCIPLINES

descriptors (column one) it becomes evident that the progression in the achievement levels can be applied across disciplines.

Looking back at the matrix in Figure 8, a distinction is made between academic function words representing high versus low cognitive demand:

Low cognitive demand:

- identify, name, retell, copy, reproduce, narrate, describe

High cognitive demand:

- generalize, compare, summarize, classify, analyze, argue, interpret

Academic functions which require a higher level of cognitive skills include activities involving making inferences and integrating new knowledge with old, whereas tasks with a low cognitive demand deal with memorizing. The cognitive descriptors are used in test items across disciplines, and require cognitive skills along with linguistic skills. They are inseparable, which become apparent in CLIL contexts. The cognitively demanding descriptors usually appear in assessment tasks requiring more production of language. Comparing or analyzing requires the use of more linguistic competence than identifying or reproducing, for instance.

Krashen and Brown (2007) develop Cummins' concept of CALP (Cognitive Academic Language Proficiency) by proposing two components: academic language *and* academic content. *Academic language* is characterized by complex syntax, academic vocabulary and complex discourse style whereas the *academic content* refers to the relevant subject content. Krashen and Brown (2007) also propose a third component, *competence in the use of strategies*, which they argue can have an effect on both the acquisition of language and subject-matter learning. Strategic competence includes making academic texts more comprehensible through different reading strategies, as well as activating problem-solving through academic writing. The components suggested above are useful in CLIL assessment in the subject disciplines. For assessment purposes, students need to possess strategic competences and relevant academic language to be able to produce qualitative texts, presenting relevant subject content.

4.6 Summary

Content and language, cognitive and linguistic skills, represent cross- and interdisciplinary concerns, covering both the features of a task and the teaching material as well as the required skills of the learner, as illustrated in Figure 10 below.

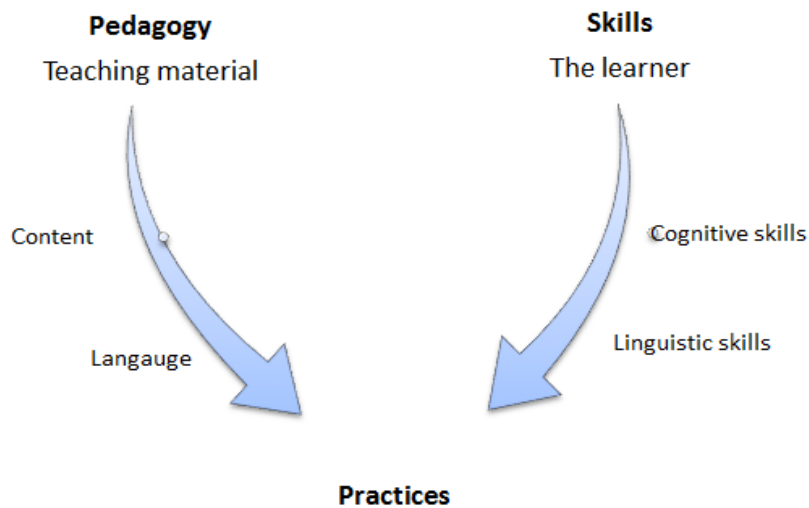


Figure 10. The twofold processing of language and content in assessment practices

The figure illustrates the two sides of the coin; the pedagogy as seen in the material used for learning, often texts, and the required literacy skills of the learner in order to succeed when processing the material. If the material is multimodal and consists of pictures and diagrams for instance, other processing and issues arise. This, however, will not be covered in this study. Content and language merge in the assessment practices, represented by the content and language of the actual assignments as well as by the twofold demands on the learner, which become particularly important in CLIL assessment. In the CEFR, it is stated that it is necessary to take both the learner's competences/skills and the conditions and constraints of a particular task into account when considering task performance in pedagogical contexts. Consequently, users of the CEFR (2001:44) are advised to consider in which domains the learner will need to be equipped and to operate, as is the goal in genre pedagogy:

If I cannot predict the situations in which the learners will use the language, how can I best prepare them to use the language for communication without over-training them for situations that may never arise?

LANGUAGE AND CONTENT IN THE DISCIPLINES

To conclude, before discussing material and methods in the next chapter, the goal for good practice is to align the method or task used with the intended outcome, which is highly dependent on whether the learner possesses the required skills or not. It is also dependent on whether intended outcomes are expressed and targeted for both content and language.

5. MATERIAL AND METHODS

5.1 Introduction

This chapter gives an account of the materials and methods used in the present study. The first section presents the data collection procedure followed by a presentation of the participating teachers and schools. Then follows a description of the tools used for the data collection and for the analysis of the material. Some comments on ethical concerns and the validity of the current study follow, including some remarks on the limitations of the methods used. The chapter ends with a brief summary.

5.2 Selection and data collection

As previously mentioned, the present study is incorporated in the large-scale CLISS project (for a fuller description see Sylvén & Ohlander, 2014). Therefore, the schools involved in the CLISS project are also the schools where this study was conducted. The schools were contacted for this specific investigation on assessment, and the material collection started early in 2013, with a first visit to school B in late February. The purpose of the study was explained, and one of the assigned contact teachers organized and prepared a selection of four teachers, two English language teachers and two biology teachers, for the first individual semi-formal interviews. One of the four teachers retired in June 2013, and a colleague at the same school with the same combination of biology and CLIL was added to the study in the fall 2013.

After the first visit to school B, the need for one more school was acknowledged for a larger sample. The same procedure was repeated at school C, where one of the organizing teachers sent contact information to four colleagues, two English language teachers and two history teachers, who after e-mail contact were interviewed individually.

In the fall, when the collection of assessment samples started, the two English language teachers at school C no longer wanted to take part in the study. Since no colleagues at their school were willing either, a third school

ASSESSING LANGUAGE OR CONTENT?

was added, school A, even though it is slightly different from the other two, due to the international profile of the entire school. Three teachers were willing to participate after direct e-mail contact, two English language teachers and one subject content teacher in history. No biology teachers were willing or available at that point. Otherwise, it would have been profitable for the design of the study to have representatives from all of the three disciplines from the international school for comparative purposes, especially in the subject content courses.

The process of collecting material was not uncomplicated, due to teachers' reluctance to share their assessment tools. Previous studies acknowledge the same problems (Hönig, 2009). Among the teachers who stayed throughout the study, the content teachers shared their assessment samples. Among the English language teachers, however, it took much longer, in several instances, to deliver any material; in some cases, nothing was presented. This will be discussed in more detail in Chapter 7.

5.3 The schools and the participants

The study includes twelve teachers at three upper secondary schools in different parts of Sweden, here labelled school A, B and C. Three of the teachers come from school A, five from school B, and four from school C. Consequently, the teachers will be labelled A1, A2 and so forth. For an illustration of the distribution of the teachers in the schools, see Table 4 below.

Table 4. Schools and teachers included in the study

School A	School B	School C
International school, Swedish curriculum	Two programs, with one class each/year mostly in English	One program with two classes/year partly in English
CLIL history teacher	CLIL biology teacher	CLIL history teacher
CLIL English teacher	CLIL biology teacher	Non-CLIL history teacher
CLIL English teacher	Non-CLIL biology teacher	CLIL English teacher
	CLIL/non-CLIL English teacher	Non-CLIL English teacher
	CLIL/non-CLIL English teacher	

As for the English language teachers in the study, since some of them teach both CLIL and non-CLIL classes they are labelled both ways (school B). In

MATERIAL AND METHODS

school A, being an international school following Swedish curricula, there are no non-CLIL classes.

School A is situated in a large city. It has a heterogeneous student body as regards the students' L1, and offers two national programs; the social sciences and the natural sciences. School B is located in a mid-sized city, as is school C. They both have a rather homogeneous Swedish L1 student body. School B offers two programs, the natural sciences and the social sciences, with one class per year mostly in English. School C offers one program, in the social sciences, with two classes per year partly in English.

Table 5 below presents a summary of the participants and background information used for the discussion in Chapter 7.

ASSESSING LANGUAGE OR CONTENT?

Table 5. Teachers' background

Teacher	Subjects	Year of diploma	Years of experience: teaching/CLIL	Degree in English	Other
A1 CLIL	English, Spanish	2007	6/6	Yes, native speaker	Examiner for ILS, Cambridge exams
A2 CLIL	English, Swedish L2, Communications	2005	7/7	Yes, native speaker	
A3 CLIL	History, religion	2007	6/6	No	Teaches geography
B1 CLIL/non-CLIL	English, P.E.	2002	14/14	Yes	
B2 CLIL/non-CLIL	English, Swedish	1982	20/17	Yes	
B3 CLIL	Biology,	-	-/-	No	Retired in June 2013
B4	Biology, chemistry	1984	28/-	No	
B5 CLIL	Biology, social science	1996	16/7	Yes	Has taught English
C1 CLIL	English, Natural science, Computers & ICT	-	-/15	Yes	
C2	English, Russian	-	15/-	Yes	
C3 CLIL	History, religion	-	20/4	Yes	Master degree abroad
C4	History, social science	2001	11/-	No	

Since the teachers are in focus in the present study, their background is of interest for the outcome of the study. The teachers' gender has been concealed for anonymity reasons, but their certification (subjects they teach), year of diploma, years of experience, both overall teaching experience and years of teaching in a CLIL context, are presented. Since certification to teach English may be a requirement to teach CLIL in other countries, but not in Sweden, this information is added. The last column, is included for any additional background information which may be of interest.

5.4 Tools and analyses

A mixed methods approach is used where different tools of analysis serve different purposes (Ercikan & Roth, 2006). The initial, tentative plan was to collect assessment samples and to interview teachers on two different occasions, before and after the rendering of assessment samples. The second interview was supposed to serve as a kind of retrospective interview, asking a few well-defined follow-up questions, whereas the first was more open and semi-formal. After the first four interviews, the material seemed too small to generate enough information for comparability, especially with regard to English. When two more schools had been added, there was enough material to exclude the second interview, also due to time constraints and the lack of rendered assessment samples for a follow-up. Instead, a questionnaire was used to substitute for the second interview.

After the analysis of the data generated from the three data collection methods, a holistic validation process of the assessment procedures in the CLIL context was performed using Crooks, Kane and Cohen's (1996) validation model in eight steps, presented in section 3.7. Below, the three data collection methods are described, starting with the interviews, followed by the documentary analysis of the collected assessment samples and, finally, the questionnaire.

5.4.1 The interviews

In an interview situation, there is always the risk that the interviewer may influence the responses of the interviewee. The semi-structured interview format was used to reduce the effect of too narrowly restrained questions, as in a structured interview. The themes covered in the interviews are presented below.

Interview guide:

- Teacher background
- Views on and experience of CLIL
- Views on teacher's own discipline/subject
- Assessment practices used
- Course and textbook material
- Course outline and plan

ASSESSING LANGUAGE OR CONTENT?

- Disciplinary and/or interdisciplinary collaboration

The present study is limited to the teachers' perceptions concerning the assessment practices in use, both in CLIL and in non-CLIL instruction. Therefore, the teachers' background regarding number of years in the profession, as well as experience of teaching CLIL, when applicable, was of interest. Since the lack of formal education and preparation for CLIL teachers is an issue which has been raised in several CLIL contexts (see section 2.), the teachers' exposure to and/or certification in English was of interest. Any possible cross-disciplinary teaching certification among the English language teachers was equally worth noting. The teachers' attitude to CLIL was also considered relevant since this may have an effect on their practice. For alignment purposes, not only assessment practices were discussed, but also course material and course layout regarding content and teaching methods.

Integration is supposed to be a key feature of the CLIL approach. Hence, this was discussed even though it is not immediately expressed in the research questions. Interdisciplinary collaboration has been stipulated in some CLIL frameworks, and even considered a prerequisite for successful CLIL implementation. Subject integration exists in other instructional contexts, not specifically CLIL, where it usually implies interdisciplinary collaboration.

A particular focus in the study is placed on the mutual relationship between content and language *within* as well as *between* subject content courses and English language courses in the same school context, especially in relation to assessment purposes. Therefore, the teachers were interviewed regarding the role of language in their courses and how they deal with language in the classroom. Since the present study is limited to interview data and no actual classroom observations, the teachers' statements stand alone and can only be compared with how they design their assessment tools in the cases where samples have been rendered.

The interviews at the first two schools were conducted in March to June 2013. The third school was added and three more teachers were interviewed in December 2013 to February 2014. The interviews lasted between 16 and 58 minutes, depending on how much time the teachers were able to spend. The paired interviews, one interviewee at a time with the interviewer, took place in the teachers' offices, or in the school library on one occasion, at the teachers' schools. A couple of the interviews (B2 and C2) were interrupted since colleagues needed to be interviewed in between for schedule reasons. Twelve

MATERIAL AND METHODS

teachers were interviewed at the three schools: six subject content teachers, three CLIL and three non-CLIL, and six English language teachers. For an overview of the teachers, their subjects and the time of the interviews, see Table 6 below. School A was added last, and the labelling of the schools (A, B and C) is the same as in the CLISS project. The table is organized after school and teacher, starting with school A and teacher A1.

Table 6. Overview of interviews with teachers

Teacher	School	Time	Date
A1 English CLIL	A	54.07	December 2013
A2 English CLIL	A	34.19	January 2014
A3 History CLIL	A	49.00	February 2014
B1 English CLIL/non-CLIL	B	34.28	February 2013
B2 English CLIL/non-CLIL	B	19.56 + 20.36	February 2013
B3 Biology CLIL	B	34.41	February 2013
B4 Biology non-CLIL	B	16.44	February 2013
B5 Biology CLIL	B	41.30	October 2013
C1 English CLIL	C	23.44	June 2013
C2 English non-CLIL	C	34.34 + 19.38	June 2013
C3 History CLIL	C	58.00	June 2013
C4 History non-CLIL	C	41.44	June 2013

Two of the English language teachers are native speakers of English, A1 and A2. Those interviews were conducted in English, all the others in Swedish, which means that quotes from ten of the twelve interviews have been translated from Swedish. Teachers B1 and B2, teach English to both international CLIL classes and non-CLIL classes. Teacher B3 retired after the interview and was replaced by a colleague at the same school, B5, who came back from maternity leave and was interviewed in October 2013.

The interviews were recorded, occasionally followed up by e-mail communication for clarification purposes. The recordings were transcribed, analyzed and compared, using the themes from the interview guide and the research questions as guidelines. Phrases and comments representing the themes were highlighted and coded, one theme at a time and one school at a time. Initially, all responses were thus categorized thematically, regardless of discipline.

In the next step the interviews were arranged according to discipline, contrasting CLIL with non-CLIL for each theme to find similarities and/or differences. Subsequently the interviews with the biology teachers were read simultaneously to find patterns within the subject, followed by the interviews

with the history teachers, and finally those with the six English language teachers, who represent the largest sample.

A selection of answers representing the research questions and interesting features regarding CLIL and disciplinary collaboration was made. The latter since it was considered relevant for the bigger assessment picture in relation to teacher perception and implementation of CLIL. The interview data is presented one discipline at a time: first biology, then history and English. At first the answers were categorized across disciplines, to focus on the comparison between CLIL and non-CLIL, however, this proved problematic. The answers relate to teacher perceptions and disciplinary features to a large extent, and this needed to be made clear by categorizing answers discipline by discipline. In Chapter 6, the headings in section 6.2 reveal the chosen themes; in chapter 7, a triangulation is made of all the three data collections when discussing the three research questions.

The aim of the interviews was to seek to understand and explain the experiences and beliefs of the teachers involved in the present study, beliefs which may affect the assessment practices used in a CLIL context (cf. Figure 1, Outline of study, section 1.2). By contrasting the views and the alleged assessment practices of CLIL teachers with those of non-CLIL colleagues, the goal was to find out if different or particular theories have an effect on the assessment procedure of the CLIL teachers compared to the non-CLIL teachers. Differing practices due to the status of the second language, i.e. English, used as a medium of instruction were of particular interest. Teachers' beliefs in relation to their discipline and assessment in general were also noted.

5.4.2 Document analysis

During the interviews, the teachers were requested to present some written assessment samples from one or several of their courses, or to send them in after being given some time to gather the material. They were also asked if they could show a plan of the different themes and assessments used during a course. Some claimed they did not have such a plan, while others presented very explicit lists.

Some of the teachers expressed concerns during the interviews regarding the anticipated workload in gathering the material. They wanted to know if this would generate any compensation. As a consequence, the teachers were given the possibility to wait till after summer and submit the material they

MATERIAL AND METHODS

used during a course as it was used or produced during the following school year. Nevertheless, several teachers hesitated, and two of the English language teachers decided not to participate any further. Other teachers claimed to be positive during the interviews, but ended up not sharing any material in spite of several reminders via e-mail, and in some cases a new visit to their school. As time passed, a decision was made to use the available material as it was and not to bother the teachers any further. Consequently, the submitted assessment samples are as follows:

Table 7. Overview of submitted assessment samples by discipline and school.

Discipline	Assessment samples
Biology CLIL, School B	4 tests 1 writing assignment
Biology non-CLIL, School B	4 tests 4 writing assignments (reports)
History CLIL, School A	4 + 2 tests
School C	2 tests
History non-CLIL, School C	4 tests
English CLIL, School A	No tests, nor portfolio description 1 standardized rubric
English CLIL/non-CLIL, School B	8 + 6 + 1 tests 1 + 1 assignment
English CLIL/non-CLIL, School C	None

As can be seen in the above table, assessment samples in English were only submitted from one of the schools, school B, where most samples were provided by one of the teachers. Samples were presented from three different courses: eight from English 5, six from English 6, and one test from English 7. English 5 represents the first year of English at upper secondary school, English 6 the second year, and English 7 the third year. In the same way, two samples of writing assignments were presented, one from English 5 and 6, respectively. For the other stated writing assignments, the teacher referred to descriptions in the course book. In one case, one of the English language teachers handed in student texts, representing samples of what students produce, and copies of rubrics, since he did not have any tests or instructions for writing assignments to submit. He handed in a brief list of different genre texts to include in a portfolio, but no task descriptions or knowledge requirements.

Assessment samples in history were submitted by one non-CLIL and two CLIL teachers at two different schools, whereas biology CLIL derives from

ASSESSING LANGUAGE OR CONTENT?

one teacher, at the same school as the non-CLIL colleague. At school A, the CLIL history teacher handed in four tests or writing assignments from history 1a, and two from history 2a. The two courses represent two consecutive courses, taught either during the first and second year, or second and third year, of upper secondary school. In history the teacher refers to all the samples as tests, but four out of the six (two of four in history 1a and two of two in history 2a) rather represent essays or writing assignments.

Altogether 42 assessment samples were collected. For reasons of delimitation, not all of the samples are described or analyzed in detail in this study. A choice was made to focus on one course in the subject content courses: the first given in the upper secondary school, called history 1b and biology 1, were chosen.

For the document analysis, a combination of models was used to describe the features of the tests and the interplay of subject content versus language. This constituted the most demanding part of the analysis; to choose relevant features and how to combine those across disciplines, since no existing model was found. The terminology relating to the different models concerning language and content was discussed in Chapter 4 (cf. Bloom's revised taxonomy and Cummins' matrix). The assessment features in written assessment and tests were presented in Chapter 3. Table 8 below provides a summary of the layers and the features used for the documentary analysis of the assessment samples. The description ranges from surface features (first and second rows) to a more in-depth description of content, as seen both in the content of the test items and in the required knowledge and skills of the test taker, i.e. the student.

The features were selected after consulting literature on assessment design (Brown & Hudson, 2002; Levin & Marton 1973; Wedman 1988; Wikström, 2013) and previous research in the field of CLIL (e.g. Hönig, 2009; Wewer, 2014), as well as assessment in the subject content disciplines (Lindmark, 2013; Odenstad, 2010; Rosenlund, 2011). National course goals, the CEFR and Bloom's revised taxonomy were used to compare the features in the assessment items. The language was described in terms of academic function words (cf. section 4.5.3), adapted after a list made by Dalton-Puffer (2007) in combination with Bloom's taxonomy and Bachman and Palmer's (2012) target language use, TLU domain. An adapted version of Cummins' matrix (Coyle, 1999) served as a tool to combine a comparison of the cognitive difficulty of

MATERIAL AND METHODS

content, lower order versus higher order thinking skills and the progress of language in the same tasks. For more details see section 4.5.

Table 8. Areas and features described in relation to the assessment samples

	Features
Course layout	Number of tests Type of assessments
Design/layout “How”	Number of questions Question type: <ul style="list-style-type: none"> - Selected response - Constructed response - Production test Modalities Scoring/Grading Time on task
Content “What”	The test items: <ul style="list-style-type: none"> - Subject theme: In relation to course goals - Language: Academic function words Question words Context embeddedness Required knowledge/skills: <ul style="list-style-type: none"> - Linguistic skills, target language - Discipline specific/Content obligatory language - Communicative function/Content compatible language - Cognitive skills, high vs low demand

The description of the assessment samples in Chapter 6, starts with a comparison of the *how*, the design of the tests in the content courses, followed by a comparison of function words and question words used in the test items in biology and history. For English, no such comparison will or can be made, due to the lack of assessment samples and the fact that the submitted test samples are used in both contexts at the school in question.

The rest of the section deals with the content of the test items, starting with biology and then history, followed by English. The presentation starts with a brief introduction of the course, called ***Course description***, where course books and course material are mentioned. In those cases where teachers provided a course plan or a list of all planned and used assessment tasks during a course, this will be mentioned as well.

ASSESSING LANGUAGE OR CONTENT?

A presentation of the *what*, the content and formulation of the questions, concludes the description. It is accompanied by some item or question samples from the tests. This part is organized according to question and test type, starting with *selected response* (SR) followed by *constructed response* (CR) and, finally, *production tests*. The intention was to present samples in order of difficulty, but some of the test items may generate a mix of grades (A, C and E), thus preventing such a categorization. However, the questions will be described in terms of what demands the questions make on students' linguistic and cognitive skills. The function word used in the question signals what is required. This also goes for the entire phrasing of the question and the question words used, whether it is a matter of low or high cognitive demand, lower order versus higher order thinking skills; remember and recall or analyze. Context-embeddedness in the question will also decide the level of difficulty. This analysis will follow after the heading ***Content of the test items***.

5.4.3 The questionnaire

In order to acquire additional information to complement the semi-formal interviews, a questionnaire was conducted in the spring of 2014. After going through all the interview data, questions remained and due to the semi-formal format of the interviews not all interviews had generated the same information, complicating comparability. Also, as noted in the introduction to this chapter, the initial idea was to include a second retrospective interview with the participants after submitting the assessment samples; this however, had to be excluded.

First a pilot study was performed with four teachers from all the three disciplines at two municipal schools, one of them using subject integration in certain courses. After the pilot test, some minor adjustments were made regarding the formulation of a couple of questions, but the questions remained the same.

The questionnaire consists of 26 questions, divided into three thematic units covering *teacher background*, *assessment and course content*, and lastly issues related to *interdisciplinary features and assessment* (see Appendix 2 and 3 for a full sample of the questionnaire). The design of the questionnaire was adapted

MATERIAL AND METHODS

from a previous questionnaire conducted among language teachers⁵, to fit the present study and its purposes, and literature in questionnaire design was consulted (e.g. Trost, 2001).

Nine out of the original twelve participating teachers responded to the questionnaire, one had retired at that point, and the English teachers at school C were no longer part of the study.

The answers to the questionnaire were compared, trying to distinguish patterns or differences in teachers' answers depending on discipline and CLIL vs non-CLIL. General features relating to the prevailing national context were identified concerning teachers' attitude to assessment and their perception of different assessment tools. Specific features relating to the research questions were identified. A selection of questions was made to be presented in more detail in section 6.4.

5.5 Ethical concerns

In this study the ethical guidelines of The Swedish Research Council (2011) were followed. All efforts have been made to conceal the identities of the individual participants as well as of the schools involved in this study. The schools were randomly assigned letters, and consequently the teachers a letter and a number. To further prevent the identities of the teachers' from being revealed, all of the teachers at schools A and C are identified as male, whereas the teachers at school B are identified as female.

No written form of consent was collected since the participating school had already agreed to be part of the CLISS project. However, the teachers who took part in this study volunteered and could decide at any time if they did not wish to participate any further. At the interview, teachers were told about the purpose of the study. Before collecting the assessment samples, teachers were informed that entire tests would not be spread, only individual questions cited, due to confidentiality in assessment documents in current use for assessment. Teachers who handed in samples were informed that the analysis would be merely descriptive and if questions of important nature arose, they would be contacted for clarification purposes.

⁵ LUB, (Lärarenkät angående färdighetsbedömning i språk)Gothenburg University

5.6 Validity of the study

The present study is qualitative with the purpose to examine the assessment aims and practices of a sample of teachers. Winter (2000) states, in Cohen et al (2011:182), that “[v]alidity in qualitative research depends on the purpose of the participants, the actors, and the appropriateness of the data collection methods used to catch those purposes”. Assessing validity always entails fallible human judgments on the part of the researcher, since validity is a property of inferences, leaning on experience with the topic (Shadish et al, 2002).

Generalizability may be limited due to the small number of participants in the study, thus affecting external validity. The validity of the construct is similarly threatened by the limited sample material for the document analysis, in terms of construct under-representation (Messick, 1989). Moreover, construct validity may be questioned since classroom observation was not possible to conduct in the limited time the data collection took place. It is known that what participants say they do in interviews and what they actually do, do not necessarily correlate. In the words of Silverman (2011:5): “[I]f we want to understand behavior and interaction, it is not enough to ask questions. We must observe the routines and practices of social actors.” However, the assessment samples did add valuable information when such documents were rendered. The aim, it should be remembered, was to contribute, albeit to a modest extent, to the underexplored area of assessment in bilingual teaching.

5.7 Summary

In this chapter, the design of the study has been described. The data collection procedure and the methods of analysis have also been outlined. In order to gain some understanding of the participants’ assessment practices in the particular CLIL contexts involved in this study, a document analysis was performed together with interviews and the use of a questionnaire. A triangulation was deemed appropriate to better understand teacher routines. Since very little research exist to date as regards CLIL and assessment, no existing method or model for the analysis was found. Moreover, a study such as the present, being cross-disciplinary, has to consider multiple variables in relation to the features of the individual subject disciplines.

MATERIAL AND METHODS

To conclude the methodology chapter, a flowchart provides an overview of the different steps and methods used during the study process.

- Preparations, deciding design of study
 - Visit for interviews at the first school (B)
 - Decision to add one more school and one more discipline (history)
 - Visit and interviews at the second school (C)
- Interview data
 - Listening through recordings, one school at the time
 - Second listening, transcribing
 - Adding one more school
 - Visit and interviews at the third school (A)
 - Listening and transcribing interviews
- Analysis of interview data
 - Reading transcriptions using interview guide
 - Categorizing answers, one school at the time
 - Comparing answers, one discipline at the time
 - Selecting interesting/relevant information
 - Compiling significant answers using the RQs
 - Selecting representative quotes
- Questionnaire
 - Designing questions
 - Making pilot-study
 - Analyzing answers, modifying questionnaire
 - Sending a digital version of the questionnaire to the participants
 - Analyzing, comparing results question by question
 - Complementing background information in thesis
 - Selecting interesting/relevant information
- Assessment samples
 - Selecting sample tests, one course/discipline in biology and history
 - Analyzing rendered tests to find appropriate descriptors/features
 - Describing the design features of the individual tests (cf Table 8)
 - Categorizing tests according to test type: Question tests and production tests
 - Describing tests one discipline at the time (the how)
 - Analyzing the content of the tests, using item types
 - Categorizing questions by item types
 - Making a frequency count of function words and question words
 - Selecting and describing test items representing different cognitive and linguistic demands (cf Table 8)
- Holistic analysis

ASSESSING LANGUAGE OR CONTENT?

- Triangulating data, analyzing using the RQs
- Validating data using Crooks et al's Chain model
- Making suggestions for assessment guidelines

6. RESULTS

6.1 Introduction

This chapter presents the results of the study. A triangulation of the findings is made, starting with a presentation of the interview data (section 6.2) followed by a description of the assessment samples (section 6.3) before some results from the questionnaire (section 6.4) are presented. The interview data and assessment samples are described one discipline at a time, comparing and contrasting CLIL with non-CLIL. The presentation begins with biology, followed by history and finally English. A summary (section 6.5) with some concluding remarks ends the chapter. The present chapter is merely descriptive; an analysis and discussion of the results will follow in Chapter 7.

6.2 The interviews

In this section, the renderings from the interviews are presented, by indirectly referring to teachers' comments or directly quoting them. As mentioned previously, the quotations are translated from Swedish, except for those deriving from interviews with the English L1 teachers, A1 and A2, which were conducted in English.

The interview data is discussed thematically in accordance with three of the major themes of the interview guide (For a presentation of the whole interview-guide see section 5.4.1). For the English language courses, the second theme regarding language is not used, since it only applies to the CLIL content classes. A summary concludes each discipline, where the findings are compared with other information from the interviews concerning teachers' background and views in relation to the implementation of CLIL. For a disciplinary overview of teachers' general experiences, see Appendix 4.

The informants are labelled according to which school they represent (A, B or C), together with a number given randomly in the presentation in Table 5, Chapter 5. In order to distinguish the teachers' subject and relation to CLIL, when cited, an abbreviation for the subject they represent as well as for CLIL versus non-CLIL is added, as in the examples below:

ASSESSING LANGUAGE OR CONTENT?

A1/En/C = School A, teacher 1, English CLIL

B2/En/C/nC = School B, teacher 2, English CLIL and non-CLIL

C3/Hi/C = School C, teacher 3, History CLIL

B4/Bi/nC = School B, teacher 4, Biology non-CLIL

6.2.1 Biology: CLIL and non-CLIL

The interview material in biology derives from three interviews with two CLIL and one non-CLIL teacher at a municipal school with an international profile in two of the programs, where one class per year is taught mostly in English. Two of the teachers teach biology and physics, and one of the CLIL teachers teaches biology in combination with social science.

Assessment

Written assessment seems to be important for summative and formative purposes in both CLIL and non-CLIL courses. There are tests, lab-reports and writing assignments in connection with excursions. The CLIL course book contains objectives for every chapter and tests which can be used to practice on for the final exam, but one of the teachers mentions that she is not so fond of those tests:

1. **B5/Bi/C:** I don't like their way of constructing the questions because the details come first. [...] I don't know where they will be able to produce a text of their own, to be able to explain to someone who is not familiar with this theme.

The CLIL teacher thus expresses one of her intended learning outcomes, that students will be able to produce a text. She also mentions that she asks "annoying" questions in the tests, which the students have already been exposed to. The first time it may have been in a lab-report on which they receive formative feedback, and then they may be asked the same question in a test for summative purposes. Thus, they have had ample opportunities to prepare for the test. She claims that she likes written tests, but also lab-reports:

RESULTS

2. **B5/Bi/C:** I think written tests have the advantage that you [as a student] have the possibility to sit on your own and to really express yourself, to use your language, Therefore I think tests at the end of almost all themes are good. However, labs are also opportunities for assessment, their [the students'] way of applying a method, and that is in the classroom, which they have to receive feedback on as well.

A reason for the advantage of written tests is noted: students can use their language. She devotes time after every test for oral feedback, all of which is done in English. She stresses the importance of feedback for students' development, but adds that it also teaches her something about the way students think and, most important, about their goals. Her CLIL colleague remarks that she tries to have some oral presentations as well and to find time to talk to students in small groups, since personal communication is important for assessment purposes. She finds it hard because there is not enough time, and some students are not very talkative; the language might be a barrier, even though they are allowed to use Swedish.

When discussing the possible hampering effect the English language may have on students' ability to present their knowledge, one of the CLIL teachers says that since the students have received all of their instruction, practice and teaching material in English, they should be able to do it in English.

3. **B5/Bi/C:** The question is whether the instruction has meant that the teacher has transmitted his or her knowledge but not put the students in a position where they have practiced how to present their knowledge [...] I believe that students sometimes find it difficult to present their knowledge regardless of language.

Both of the CLIL teachers mention progression in students' knowledge when constructing test items and when assessing:

4. **B3/Bi/C:** Well, you try to make the test items graded by difficulty, according to the grading system that is [...] so you try to make some A-questions, some C-questions and some E-questions, maybe, and then it shows how well they are able to solve those. [...] Sometimes you do it with points as well, but that maybe shows the same results, and that is as good.

The CLIL teacher in the above quotation notes that today's students should be able to draw conclusions on their own. The new grading scale does not change anything but the fact that there are more steps. Regarding the grading, she argues that there are different philosophies about whether to put a grade

ASSESSING LANGUAGE OR CONTENT?

or a score on the test. She notes that you have to answer the more demanding questions in order to receive a higher grade. Full score on the easier test items will not be enough. So, in order to answer the more demanding questions, what does this CLIL teacher believe the students need to know?

- 5. B3/Bi/C:** I don't require that they have to know specific concepts when it comes to the kidney; renal pelvis for example. They probably know "kidney", but if they don't remember renal pelvis, they can write it in Swedish. [...] I don't require that they have to know specific words, that is not the main point, they need to know relationships; how does the kidney function – describe [...] If they don't know all the words in English, then it works just as well in Swedish. They need to know how to describe what happens.

The non-CLIL biology teacher states that assessment in biology leans on evolutionary principles and that this pervades the whole subject, what governs which genes are passed on and so forth.

- 6. B4/Bi/nC:** It is not so much about going on about taxonomies, or what is the name of that class [...]. It is more about a historical perspective, how has the view on people, nature and science evolved over time; that is the foundation.

The follow-up question concerns how this affects test design and the way assessment tools are formed:

- 7. B4/Bi/nC:** There is a lot of problem solving, and you build on cases, phenomena and to explain phenomena which have occurred and compare different systems. And then there is warm-up assessment, if you think of tests, where you have to know the meaning of concepts, there are many concepts and models.

The non-CLIL teacher explains knowledge development in biology as a process where much attention in assessment is paid to the working methods and the character of the subject, and the use of academic language, i.e. to form hypotheses, draw conclusions and practice analytical competence. Therefore, classroom assessment is important to see how students tackle problems. In written tests, it is more about explaining concepts, theories and models, including questions where students need to explain a sequence of events. Students also need to show their knowledge of key concepts, something which the non-CLIL teacher relates to as "almost like vocabulary lists".

RESULTS

Language in the course

Regarding the language used in class, the situation varies. Both of the CLIL teachers note that “it is a bit of a hassle”, to quote one of them, translating everything into English. This also means that they have to think of the students and provide them with specific disciplinary concepts, so-called content-obligatory language (see section 4.3.2 on content-obligatory versus content-compatible language), which is something both CLIL teachers express, but deal with a little bit differently.

8. **B3/Bi/C:** Sometimes when you know that this is probably a term they don't know, if it is a disciplinary term, then I say it in Swedish too, at the same time, and they get an explanation so it won't be a stumbling block all the time.

When asked about classroom language the teacher declares that all the communication on her part is in English, and the students try, but sometimes they use Swedish if they feel unable to do it in English.

9. **B3/Bi/C:** It is better they that they say something, even if it is in Swenglish⁶ or Swedish, than them not saying anything. It can be a bit tricky, but it is not the language which is the main thing, it is still biology that is the main thing so to speak.

She explains that there have not been any restrictions from the school management that you have to stick with English even though the school has an international profile, but she argues that since the language may constitute a hindrance students should get all the help they need. She acknowledges that when the content theme has been particularly difficult, as when they were working on the anatomy of the human body, some students asked for material in Swedish, which they also received.

The other CLIL teacher states that she allows the CLIL students to have more time on tasks in assessment situations, which is an effect of the use of English. She also teaches social science, using English as a medium of instruction. She notes that it is hard for the students to adopt the special vocabulary, but her English colleagues have inspired her to play with words and make word games. She acknowledges that there is a great deal of content-obligatory language and subject-specific disciplinary concepts. She argues that this may be more striking in social science, where the students seem to believe

⁶ Swenglish is a term coined to denote a mix of Swedish and English

ASSESSING LANGUAGE OR CONTENT?

that it is a subject where you “can just talk” using ordinary language, which is not the case. In biology, on the other hand, people are aware that there are many subject-specific concepts. When asked if they deal with general academic content-compatible language in class, used as filler to put concepts in context, the teacher seems a bit startled at first. She starts to think aloud, discussing how she is working with the writing of lab-reports, and remarks that fluency in writing is what you strive for, and so filler language should not constitute a hindrance, or else she would have noticed.

- 10. B5/Bi/C:** I don't know how much I work on that, more than in the classroom where there is a lot of talk. [...] It is based on these key concepts, but I put them in sentences, to create a story, and in that case it is my way of talking which is either a help or insufficient help. [...] But I have not experienced that this should be a problem.

The teacher says that she provides students with linguistic input when they are listening to her speaking English. She notes that she not only comments on the content, but also on their way of using the terminology in written reports. The other CLIL teacher has a different approach, claiming that the students may use Swedish if they get stuck.

- 11. B3/Bi/C:** I don't care about the language, that is up to the English language teachers.

Thus, the CLIL teachers express different views regarding the use and role of language as part of course content and assessment.

Course content and design

When asked about how to align the English material with the Swedish curriculum and syllabi, one of the CLIL teachers admits that this involves a great deal of work, partly due to the implementation of a new national grading scale in 2011. Further, the teacher explains how the English course book they use focuses more on details and less on the larger picture. She notes that this may derive from a different approach in the English way of presenting facts. While talking about course content, the question arises whether the use of English as a medium of instruction has an effect on course content.

- 12. B3/Bi/C:** No, it does not affect the content, even though there have been concerns that you don't have time to cover the same amount of

RESULTS

content since it is a little bit trickier of course, but I feel that they [the students] are so motivated so it is not my impression that we cover less.

Both of the CLIL teachers believe that they cover the same content as their non-CLIL colleagues, something they know due to close interaction with non-CLIL colleagues. They all claim that there should not be a difference in course content since it is just the language and the literature that differ.

Summary of interviews with biology teachers

To summarize, the CLIL teachers share a positive view regarding the effects of CLIL, especially for them as teachers (cf. Appendix 4). Concerning the effect of CLIL on students, they believe their proficiency in subject-specific English vocabulary improves. As for other possible effects, they ask for research to prove what such effects may be.

None of the CLIL teachers has received any special CLIL education, but they have been able to look for learning opportunities when going to England with the school's exchange program. Using English as a medium of instruction is time-consuming and demanding, especially the need to find and create course material. Both of the CLIL teachers believe that they are able to cover the same amount of content as their non-CLIL colleagues. There is no interdisciplinary collaboration between biology and English language courses. CLIL and non-CLIL teachers agree about the analytical character of the subject and that assessment should provide opportunities to mirror progression in students' cognitive skills in biology and the use of language to express what you know, one reason why they favor written assessment.

The two CLIL teachers differ in their views regarding students' use of English. One of the teachers believes that translanguaging (see section 4.3.3) is acceptable and so allows the use of Swedish or "Swenglish", and also the use of dictionaries. The other teacher argues that students should use English. If teaching activities are aligned with assessment practices, students receive opportunities to practice and prepare for the tests. Both the non-CLIL teacher and the CLIL teachers are aware of the need for students to learn key concepts. However, the non-CLIL teacher is the only one who also mentions the need for more general academic language, although one of the CLIL teachers says that concepts are not really important, and students can use Swedish if they do not know a word. What matters is that they can describe relevant processes.

6.2.3 History CLIL and non-CLIL

The interview material in history derives from three interviews with two CLIL and one non-CLIL teacher. One of the CLIL teachers works at the international school whereas the other two, CLIL and non-CLIL, come from the same municipal school where one program with two classes every year is taught partly in English. The CLIL teachers are certified to teach religion as well and the non-CLIL teacher to teach social science.

Assessment

When it comes to assessment and the effect of the language used, one of the CLIL teachers notes that it is not only a matter of understanding; he acknowledges that he has been concerned about the ability to show content knowledge.

13. **A3/Hi/C:** Well, it is a bit tricky, because they are supposed to show their understanding, that's the thought, and knowledge of certain concepts and so forth, and sometimes when their writing is confusing, and it is not correct, it makes you wonder if it is a matter of linguistic problems or problems in understanding.

This teacher's solution to the problem, similar to that of the other CLIL teacher, is to approach the student and ask about ambiguities. He agrees that it works a little bit like a portfolio, where you can look back at written assignments for formative assessment. It is described as a written dialogue, where he comments in the margin if there is something he does not understand in the student's text. Then he may ask the student for clarification, make new comments before handing it back to the student when he has some more information, which may be a process that takes a couple of weeks.

When asked if the use of appropriate vocabulary is part of the assessment, the other CLIL teacher comments that he finds no room for that in the objectives. Instead he brings it up in his feedback to the students.

14. **C3/Hi/C:** If I am unsure [what a student means] I take a discussion, "I understand what you mean, so can you think of this next time". But the English proficiency cannot be part of an examination as little as Swedish should be part of assessment in history.

His CLIL colleague notes that teachers prefer written tests since it is easier to grade; essays are much more demanding. He uses a mix of tests and essays.

RESULTS

However, he believes it is difficult to show evidence of different proficiency levels in a regular question test. In a test situation you only have one or two hours at your disposal while you may need to sleep on it and come back the next day in order to make an analysis. There is not enough time in a question test to show proof of analytical skills, he argues.

One of the problems with written assignments and essays is the possibility of plagiarism, but the teacher claims that the language the student has used reveals if it is the student's own words or not. The teacher speaks of other issues in assessment: how to know what a student knows if he or she does not dare to speak in oral assessment? Or in group work, who has contributed with what? These are some reasons why written assessment is easier. He states that students prefer written over oral assessment as well. When questioned why, he mentions that the language barrier might be one explanation, but also that students are used to written tests from previous schooling.

The other CLIL teacher also favors essays and the writing process for assessment purposes.

15. **C3/Hi/C:** On the whole I work a lot with the writing process and try to target that it is the tool of the historian; history is basically the literary genre.

The teacher quoted above wants to distinguish a certain written genre which he believes is important in his discipline. He mentions that his colleagues work particularly with what might be labelled academic essays, with features typical of a PhD thesis, something he also uses in certain contexts. However, he feels strongly that you have to start with narrative storytelling, which he believes has to come first.

16. **C3/Hi/C:** As a historian you have to elicit interest and creativity. To write compulsively academic essays, [...] before you have the storytelling, I feel that the storytelling has to come first [...] and then you can be bridled into the writing of a thesis, into the academic structures.

He explains that he usually has four written tests in a course and starts up with a fairly traditional written test and finishes with an independent written assignment. He tries to lead students into the everyday life of the historian in the most authentic way possible.

Both of the CLIL teachers favor written production, but seem to picture different types of texts. Whereas the teacher quoted above feels that the

ASSESSING LANGUAGE OR CONTENT?

narrative genre should precede academic, more structured texts, his CLIL colleague believes that structured texts help to clarify what is required from students. He mentions that the new national grading criteria in history require more analytical skills even at the most basic level in order to receive the lowest passing grade, an E on certain assignments.

- 17. A3/Hi/C:** We try to structure it, [...] the essay so there is, well it depends a little bit, an introduction, some sort of descriptive body, an analytical part and conclusion, and then there is a bibliography. I introduce the different parts, and here in the analytical part [...] it often deals with a certain type of knowledge, or the knowledge levels.

The other CLIL teacher on the other hand, declares that he dislikes matrices and manuals, and sees a problem in what he feels is mechanization and bureaucracy in assessment. He finds the dialogue with the individual student to be the most important tool and it works just as well without rubrics. The national course goals are enough for feedback, according to him. When asked how to make progression visible and how to explain different proficiency levels to students, he states that this becomes visible in the dialogue which surrounds the development of a text, where peer feedback as well as teacher feedback helps develop the relevant skills. How well students handle the questions which appear while working with a text, constitutes a variable for grading.

The teachers mention source criticism as a key element in history. Other important skills in relation to the subject include reaching good conclusions, based on familiarity with facts, the use of the right terminology and theoretical frameworks for history, the use of different explanatory models; how to view and explain history. One of the teachers mentions how to argue around cause and effect. World war two for instance: Did it happen because of a person, Hitler, or for financial reasons?

As regards assessment, the non-CLIL teacher mentions that he prefers tests, even though it depends on the mixture of students in the class what works best. He compares the tests with a match or game:

- 18. C4/Hi/nC:** We practice and practice and then there is a game, and then we practice and practice and then there is a game, right. So of course it [the test] plays a major role in grading. That is how the students prepare as well. Now, this is it, sort of.

RESULTS

The assessment type used also has to do with the level of the course. The first course, history 1b, is more basic, which is why he feels tests work well. In the next course, where there is more specialization, other modes of assessment might be relevant. In the design of the test items, he notes that he might be stuck in “the old way of thinking”, compared with the new grading criteria. The old way is represented by conceptual questions for an E, and analytical question for higher grades. A concern in the past has been students sometimes targeting only basic questions, hesitating to even try to answer questions generating a higher grade. In making questions which may generate all grade levels, students answer all questions and so have a chance to receive a higher grade.

As to the character of history as a discipline, the non-CLIL teacher states that it is a subject suited to do things chronologically, to turn it into storytelling. While he argues for academic and structured texts, like the CLIL teacher in the international school, he also shares the desire with his CLIL colleague to develop students’ skills in storytelling, to stir their curiosity and make them interested in history. He argues that the grading is based on how articulate students are.

The non-CLIL teacher has been asked if he would consider teaching in a CLIL context, but says he is not willing to do so unless he receives training in English.

Language in the course

Both CLIL teachers claim to use only English in the classroom, even though Swedish may be used occasionally in individual conversations with students about their achievements. One of the CLIL teachers refers to what he calls the “linguistic limitation”. The teacher states that he is very careful not to correct students’ grammar, neither in spoken nor written language, since he himself makes mistakes.

19. **C3/Hi/C:** When I switch into Swedish it is mainly for student feedback on grades and course evaluation so there are no misunderstandings, if Swedish is the first language of the student that is. [...] If the students want to ask something during class, and want to ask in Swedish, I accept that of course, and answer back in English. It has to do with not inhibiting them from asking questions.

ASSESSING LANGUAGE OR CONTENT?

What matters most is to choose the right semantic nuance. This is something he claims he focuses on in the classroom: the meaning of particular words and concepts to avoid misunderstanding. As an example, he mentions the difference between farmer and peasant when discussing agriculture. The teacher says that he introduces a new theme by looking at language, thus providing the students with wordlists.

The teacher in the international school explains that he sometimes gives students some terminology or sentences in Swedish, especially in their first year of upper secondary school. He feels that they need it and notes that students say they wish more teachers did like him, as not all students in their school are highly proficient speakers of English when they start. The teacher remarks that tricky subject-specific concepts become even harder when English is used as a medium of instruction.

- 20. A3/Hi/C:** I translate certain words. [...] I do a lot of power points. Sometimes the translation is in parenthesis, some words, some terminology and such may be tricky. I have subjects, social science for instance, where there are lots of words that are tricky even in Swedish to explain and such, so doing it in English doesn't make it any easier.

He also notes that he makes clear to students in the international school from the very start that it is possible that the use of English as a medium of instruction may inhibit their understanding, which in turn may slow down the learning process. The fact that this very likely means that students receive a lower grade than they would have if their first language had been used, he sees as only natural.

Course content and design

Regarding teaching methods, the non-CLIL teacher claims to be rather traditional; just like the students, he prefers lectures and to have a course book. Both of the CLIL teachers mention the use of pictures and other multimodal instruments, both during classroom lectures and in assignments and tests. Concerning the alignment between national course goals and the use of English textbooks one of the CLIL teachers notes that the course goals leave room to design the course, but a great deal of content should be covered.

- 21. A3/Hi/C:** The course goals are tough, which means that there is a lot of content to be covered, [...] if you want to respond to the grading

RESULTS

objectives in all of the themes, there is not enough time for them [the students] to be able to make in-depth analyses and to have analytical teacher briefings and assignments, it takes several weeks. I know of no teacher, even in other schools, who feels that this can be done.

Time constraints are mentioned by both of the CLIL teachers. One relates to how the allocation of time in the schedule is much tighter in Sweden compared with what the international course literature assumes, where they have the double amount of time to cover course content.

22. A3/Hi/C: You almost have to erase certain [things] from the course plan to have some quality in what you do. It is a little bit slower pace than what would have been the case if it had been done in Swedish.

One of the CLIL teachers does not believe that the language of instruction in the CLIL approach has an effect on course content. He thinks it has to do with the teacher's personal preferences, pedagogical ethos, general interests and student input. His CLIL colleague, however, acknowledges that he has to slow down the tempo due to the language of instruction, which means that he has to skip certain themes that are stipulated in the course goals

One of the CLIL teachers sees the advantages in having access to a much broader source material in English. At the same time, he notes that the English literature obviously implies a more international perspective and when the national course goals stipulate insights regarding the Vikings from a Nordic perspective, for instance, Swedish sources have to be used. He does not believe in translating all the material, but accepts that some material in Swedish has to be used.

Summary of interviews with history teachers

Both of the two history CLIL teachers are positive to CLIL. They mention the possibilities for students to study abroad as a result of English being used as the medium of instruction. One of the two CLIL teachers has a background from studying abroad himself and does not see teaching in English as a problem. The other CLIL teacher would not necessarily have chosen to teach in English, but was offered a job which he needed. He also feels that teaching in English has been very time-consuming and challenging, especially in the beginning. After investing all the extra time, he now feels happy to have "gained a language". None of the CLIL teachers has received any training from their schools before starting to teach in English.

ASSESSING LANGUAGE OR CONTENT?

The CLIL teachers do not feel that the language of instruction has an effect on the selected course content, but one of them believes it will have an effect on the amount of content covered. His CLIL colleague refers to teachers' and students' preferences. There is no current interdisciplinary collaboration in either of the schools, but in the international school an interdisciplinary project between an English language teacher and a content teacher has been stipulated during the school year.

Regarding the language used in class, both teachers say they only use English, and provide the students with key concepts when introducing a new theme. One of the CLIL teachers discusses semantic nuances with the students in the choice of one word over another; at the same time, he thinks students should be allowed to use some Swedish if they prefer.

In assessment all teachers prefer written tests, although the non-CLIL teacher favors so-called traditional tests and the CLIL teachers advocate essays or students' own production. One of the CLIL teachers notes that writing a text is more in line with the course goals, since analytical skills require that you have time to go home and "sleep on it", which is not the case in a traditional test. He and the non-CLIL teacher at the other school both mention academic essays as a model to teach students, whereas the other CLIL teacher speaks about the writing process in terms of the narrative genre which he believes should come first. His non-CLIL colleague mentions the narrative character of history, but does not feel there is time within the course to work with the writing process the way you want as a teacher.

Even though there are similarities between the teachers' views and perception of their discipline and assessment, there are obvious individual preferences and experiences which have an effect on their actual practice.

6.2.4 English

Six of the participating teachers are English language teachers, two from each of the three schools. They all teach English as a foreign language in combination with one other subject, in most cases another language (Swedish, Spanish or Russian) or a content subject (sports, natural science). Two are native speakers of English, and the rest are native speakers of Swedish.

RESULTS

Assessment

The teachers, except for one at the international school, state that they use a mix of tests to assess different skills. When asked what they include in their assessment of a course, they particularly mention different skills and vocabulary.

23. **B1/En/C/nC:** Well, it's all the skills; reading comprehension, listening comprehension, speaking and writing. I may have a vocabulary test, but I don't do much words. They can show that in their other assignments, their vocabulary and so forth.

The teacher describes that she is not interested in assessing vocabulary homework, since she is only interested in the use of words. The non-CLIL teacher mentions the importance of vocabulary as well and says he has homework every week, but without quizzes or tests. Another theme in the English language course is realia.

24. **B1/En/C/nC:** [W]e learn about Great Britain during that period [...] and we finish with a knowledge test on Great Britain, maybe a little vocabulary test too.

However, her colleague, who uses the same test, states that she does not put much emphasis on the test results; the purpose is to check that the students have learned what they have been told. She adds that she actually thinks that teachers should not be involved in grading their students at all, and that Sweden should adopt a system where someone from outside grades the students through a final exam.

All teachers use oral assessment in different forms. They state that the students speak very fluently in these activities. In writing assignments they require the students to write things in class for reasons of reliability. It is so easy these days to copy things online and write book reviews without even having read the book.

25. **A2/En/C:** So sometimes I write exams with questions that you only know how to answer to if you've read the book, and that's just me being evil. So tests, I like tests, oral presentations because that also gives them a chance to show their knowledge in a way that I know they've done it, they're doing it here. [...] An assignment that you let them do at home, it worries me sometimes that once again the internet is the devil in everything.

ASSESSING LANGUAGE OR CONTENT?

The teacher says that there are so many things that you can grade in an assessment, language and “information”. This means that he favors tests which cover a mix of things, tests which include a multiple choice section, vocabulary tasks including defining different words, and a section with longer essay questions. The assessment in English 7, the last year of upper secondary school, is more geared toward essays in general, whereas the first course, English 5, includes a mix of tests and assignments. The first course generally contains more diagnostic tests, typically grammar and vocabulary checks, as well as old national tests covering the four skills. The non-CLIL teacher expresses the same views as his CLIL colleagues in this regard.

Written tests seem common and very often teachers mention scores when grading or correcting tests. Sometimes it is only a matter of pass or fail and to decide students’ initial proficiency level when they start upper secondary school. However, one of the teachers at the international school claims he does not use tests, except for the national tests and a few grammar tests in the past, targeting certain aspects that come up in students’ own production. The teacher explains that he prefers to work on the writing process, but also on dictations, which he feels works really well, even though he comments that this may seem old fashioned. When asked if he is grading the dictations, he admits that it is not done in that sense, even though he states that he has told the students that everything is graded. He prefers to work with portfolios and have students work on their own texts. In grading, the teacher claims he focuses on the language, not on content or argumentation even though he admits that these are important too.

- 26. A1/En/C:** I tend to focus on the language, how well they are able to express themselves, the vocabulary, that they have as you know an indicator of their general fluency, and being able to express themselves in a coherent way; which is the intent behind the thoughts.

When asked how he works with the writing process and how he is able to clarify to students what is required for a certain grade, he explains that he makes students think about the choices they make in their writing, but not necessarily using a rubric or anything.

His colleague at the international school argues that writing essays and sending them back and forth is too time-consuming, since he has 130 students and so it consequently would “be the death” of him. One of the CLIL teachers at the international school states that writing assignments could be

RESULTS

suitable to work on together with colleagues, since they might have other goals or criteria that they would be looking for. However, he would not be comfortable grading a subject he is not familiar with. Even though he might be able to interpret a question, he might not know how the students should apply their knowledge in that discipline, but doing it with a colleague, looking at it from different perspectives, would be a possibility.

When one of the teachers is asked if the writing assignments are used for interdisciplinary collaboration around assessment, she notes that this would be difficult since assessment in those disciplines concerns two very different things.

27. **B1/En/C/nC:** We never assess language in math or geography for instance. If you present something in a written test in biology, it is not about language, they [the students] won't be assessed in language there, but they will be assessed on knowledge in the subject, right, so we don't have such collaborative assessment.

However, the non-CLIL teacher notes that he uses three parameters in the assessment of writing skills in essays: *content*, *structure* and *language*. He claims that he has used argumentative essays all the way down to ninth grade and he believes it works as long as you choose a topic which is relevant and appropriate at that level.

The national tests are high-stakes tests and important in the English language courses. Although all teachers administer the tests and acknowledge their importance, the teachers somehow have different confidence in the tests.

28. **C2/En/nC:** I compare everything I do with the national tests, and if you are lucky there is a correlation, then it is fine. If you get totally different results on what they have done in class and on the nationals, I have to test them more, which I do.

The teacher explains that this procedure helps in attaining a valid grade, especially since none of the teachers grade their own students' essays, and the oral exams are recorded in case a colleague needs to listen for a second opinion.

One of the teachers at the international school explicitly raises concerns regarding the effect of English as a medium of instruction in a CLIL context.

29. **A1/En/C:** We were talking a lot about how we can help our students improve and potentially achieve higher grades, because they're very concerned about grades and rightly so, but it's very hard when you're

ASSESSING LANGUAGE OR CONTENT?

studying something in another language and you're being, you know, graded on that understanding, that you will achieve the same goals, although the language is not necessarily there in all cases.

He also talks about different ways and assessment tools to “extract that information” about students’ proficiency and explains that the teachers at his school have mentioned oral exams or other ways to assess; otherwise, it should be writing a text, he concludes, not a question test.

Course content and design

When discussing methodology and course design, all of the English language teachers refer to the national tests in the courses; teaching to the test could be used in this context. Teachers use old tests to practice for the real tests close to the end of a course. They make sure they practice on both receptive and productive skills.

The teachers state that they like variation: individual work, group work, discussions and oral presentations. They also state somewhat different methodological preferences, though not explicitly due to a deliberate CLIL approach.

- 30. B1/En/C/nC:** We have like a basic plan for all of the courses [English 5, 6 and 7], what we think we should cover, but how we do it, yes..., how is a little bit up to the individual.

One of the teachers at the international school notes that at the beginning of a school year he prefers working on spoken goals so that students can interact in a good way in their different classes. He also likes to work structurally with essay writing, which means basic things like paragraphing, since students are not always familiar with this. Or looking into what makes scientific reading a little bit different from reading a novel. He says that he checks for comprehension and understanding of words, using the words himself as a native speaker, and hoping to get students to use those words:

- 31. A1/En/C:** That's you know the real test, if they're producing them afterwards. Modelling I think good behavior, or good language behavior I guess in a sense; what kind of words would I use, or a native speaker, surrounding a certain text type.

When asked if the CLIL students ever bring material or topics from their content courses into the English language classroom, the teachers give the

RESULTS

impression that it is rare or non-existent, although it does happen at the international school when choosing a topic for comparative essays, for instance.

The non-CLIL teacher tries to integrate some course content which relates to students' field of study and notes that students like the fact that texts are authentic.

- 32. C2/En/nC:** “In a little bit over a year from now you will be studying at the university, and this is what you will be facing”, for instance. And I picked Tim Jackson’s *Prosperity without Growth*. [...] “This is a popular scientific text in English in the field you have chosen, you have chosen economics”.

The teacher explains how he worked around the text and helped students with key concepts, such as *recession*, and *economic growth*, before they discussed the text, listened to a speech, and were given a test. The same teacher also states that he has tried to synchronize his course with a history teacher when introducing the book *Animal Farm* by George Orwell right after the students had dealt with the Russian revolution in history class.

When asked what they include in their course layout, the teachers feel they need to cover the content described in the national syllabus, and that the time is limited which prevents them from adding other content.

- 33. B2/En/C/nC:** I focus on doing what is in the national syllabus, [...] and it is very clear in our local plan that in English 5 we are working on Great Britain and in English 6 we work on the US. You can say that is the ideal.

The national syllabus stipulates what skills to include, but teachers are free to choose what material to use. At the international school, the teachers do not use a textbook since it would contain wordlists in Swedish and would be too basic for their students. The teachers at the other schools use textbooks (see section 6.3.5 below, Table 13), but explain that this only constitutes one part of the material they include in their courses.

- 34. C2/En/nC:** [The textbook] is ambitious. [...] there are a lot of really good exercises to build vocabulary. It depends on how much you feel you can do. It is impossible to do it all, so I am guessing, less than half of the course is in the course book.

Even though the non-CLIL teacher quoted above notes that there is more than enough to do in the textbook, he and his CLIL colleagues agree that they

want to include other things in their courses. At the same time teachers feel that there is too little time to do anything else than what is required in the syllabus. Some of the teachers state that they plan some of the course content with other colleagues within the discipline. Therefore, course content is very similar regardless of whether it is CLIL or not, whereas interdisciplinary collaboration is somewhat more scarce, partly due to the lack of time, as previously noted.

Summary of interviews with EFL teachers

The background of the teachers as well as the individual school context has an effect on the attitude toward CLIL. At one of the schools the teachers claim that students in the CLIL programs are motivated and proficient at the outset, whereas their colleagues at the international school express a need to encourage students and help them with the language during the initial phase.

Not much of the content in the English language courses can be traced to the CLIL approach. Teachers rely on national course goals and preparation for the national tests when planning the course design. They feel that there is too little time to cover much else. The English language courses deal with vocabulary, grammar and other formal aspects, essay structure to a large extent, as well as the four skills: reading and listening comprehension, writing and speaking. Teachers state that they use various oral activities described as spontaneous interaction around students' everyday interests, but sometimes a formal speech, book reports and discussion around different topics are also included.

Teachers like written tests or assignments. Essays typically appear more and more as students reach the higher courses, English 6 and 7, except for one of the teachers, who prefers working with students' text production as a general method and tool for assessment already from the start. Written production and argumentative essays are used in all of the courses, but the use of rubrics is rare. Feedback to students' is done through written comments. One of the teachers uses portfolio for assessment purposes, and some teachers claim to work with peer assessment as a tool in the writing process.

CLIL advocates interdisciplinary collaboration, which some of the English language teachers claim they might consider if it alleviates stress for students. All of the teachers state that they are concerned that it would take too much time, and several of them say that they do not really see how it could happen.

RESULTS

This is partly due to organizational issues at their school and what they perceive of as colleagues' unwillingness, partly to differences between disciplines. Some of the teachers have tried interdisciplinary projects and believe it has worked out well. Regarding assessment, they generally do not see how interdisciplinary collaboration could be done. At one of the schools, they have been asked to read each other's course goals, which they feel could be the start of more interdisciplinary projects.

There is no apparent difference due to the CLIL approach in the English language course in any respect. The only non-CLIL teacher expresses more concern over authentic texts and describes more spontaneous interdisciplinary exchange with content colleagues than do many of the CLIL teachers.

6.3 The assessment samples

In this section, the collected assessment samples are described, starting with a cross-disciplinary summary of the design of the tests, the *how* in biology and history. It is followed by an overview of the function and question words of the test items, combining and contrasting CLIL with non-CLIL and the disciplines.

Next follows an outline discipline by discipline, starting with biology, history, and, finally English. The presentation provides a brief course description and the content of the test items, *the what* is described in more depth regarding the linguistic and cognitive demands that the test items make on the test taker. This section shows some test items from the assessment samples. For a more exhaustive description of the method and features, see Table 8 in Chapter 5. Chapter 7 will deal with the research questions in more detail, such as the comparison of the CLIL and the non-CLIL assessment features.

6.3.1 Design of the tests in the disciplines

Table 10 offers an overview of the general design features of the tests in the different subjects, including *number of questions*, *types of test items*, *scoring/grading* and *modalities*. The test types in the second row are in bold, to make the categorization into different types of items clearer.

ASSESSING LANGUAGE OR CONTENT?

Table 10. Design of the tests in the disciplines

	CLIL biology	Non CLIL biology	CLIL history	Non CLIL history
Number of questions	7-8 (10-21)	7-11 (14-26)	5-9 (14-26) 1-3 Prompts/Essay topics	4-6 (8-15)
Types of items	Selected response/SR: Matching Putting in order Constructed response/CR: Short answer questions Production tests: Lab-reports	Selected response/SR: Matching Constructed response/CR: Completion questions Short answer questions Essay questions Production tests: Lab-reports	Constructed response/CR: Completion questions Short answer questions Essays Production tests: Essays	Selected response/SR: Putting in order Constructed response/CR: Short answer questions Essay questions
Scoring/grading	Both: A-E + points	Both: A-E + points	Both: A-E + points	Both: A-E + points
Modalities	Mostly text Pictures or diagrams in 4 of the questions	Mostly text Pictures or diagrams in 3-4 of the questions	Mostly text Many pictures in one of the tests No pictures in the other teacher's samples	Mostly text Pictures or diagrams in some questions

Biology

The biology teachers submitted three tests each, representing the same course content and themes. As can be seen in Table 10, CLIL and non-CLIL assessment practices in biology share many common features regarding number and type of questions, as well as scoring and the multimodal nature of the tests. As regards the number of questions, several test items include sub-questions, in the table represented by the number in parenthesis. The CLIL tests contain between 10 and 21 test items, including sub-questions. Similarly, the non-CLIL tests contain between 14 and 26 test items.

The tests contain mostly CR questions and short answer questions. Some of the questions in the non-CLIL tests require longer answers and could be described as essay questions.

The scoring of the questions is done using grades. Consequently, a question may generate one specific grade, an E for instance, implying that the expected answer is factual or very basic, whereas other questions may generate

RESULTS

two or three grades depending on the quality of the answer. This will be presented more in detail under *Content of the test items*.

All of the biology tests are multimodal, containing text but also several pictures of plants, animals, figures, diagrams, thematic illustrations and disciplinary symbols. In connection with some questions, illustrations refer to experiments made in the classroom, e.g. a jar when the students are supposed to describe an experiment containing a jar.

History

In history, the CLIL assessment is more varied than the non-CLIL, since essays are used alongside question tests. Further, the two CLIL teachers differ in their practices as well. One of the CLIL teachers (C3) submitted two of his four written assessment samples which represent two rather dissimilar test designs. The first test contains a mix of different CR completion and short answer questions. The test includes a total of 22 questions, including the sub-questions. A number next to each question reveals the maximum score.

The second test by the same teacher is called “History novel”, in which students are supposed to write three articles containing about 800 words each. Two should be written in English and one in Swedish, even though all the test items, consisting of writing prompts, are in English. No instructions concerning scores or grading are provided.

The other CLIL teacher (A3) uses two essays and two question tests, containing eight and nine questions respectively, the latter comprising sub-questions resulting in a total of 14 questions to be answered. The first essay offers a range of topics, whereas the second is an assignment on World War I.

The non-CLIL teacher (C4) uses four written question tests. All tests include sub-questions, implying a total of between eleven and 15 test items. The second part in two of the tests includes an essay-question where the test-taker can choose one of three topics. The grading of the questions is done using grades or points. On the first two tests, the possible grade is marked after every question. Consequently, the grade level is predetermined. Some of the longer essay questions may generate any of the grades E, C or A, implying that the quality of the answer is decisive. No rubrics or knowledge requirements are attached to the tests.

Pictures are included in the tests from teachers C3 and C4, whereas the second CLIL teacher uses merely text in the written tests.

6.3.2 Function and question words in the test items

The next area of interest concerns the function and question words used in the test items (for a description of function words, see 4.5.3). An important question is *how* the test items are phrased and what is required of students when dealing with the test items regarding language and cognitive skills.

The presentation below starts with a comparison of academic function words used in the test items in CLIL vs non-CLIL to find out which ones are most frequent. Next, the use of question words is presented. The function words as well as the question words represent different cognitive demands where some require higher order thinking skills, while others require lower (see section 4.5.3).

The instances counted in Table 11 represent six of the submitted CLIL and non-CLIL biology tests and four of the submitted CLIL and non-CLIL history tests. Looking at the table, the function words listed at the top are function words connected with lower order thinking skills, LOTS, words found in SR test items such as *put in order*, *match*, or less cognitively demanding completion questions, such as *write in the right place* and *name*. The three function words *name*, *state* and *mention* may be considered to be more or less synonymous, representing similar level of difficulty. Since they all appear in the test items, they have been given separate entries in this frequency count. The discussion on complexity will follow, both in this chapter, and in Chapter 7.

Explain is by far the most common function word in both CLIL and non-CLIL test items in biology, but there are four instances in the non-CLIL history test as well. In the biology CLIL tests, *describe* and *explain* sometimes appear in the same test item, as will be seen in one of the item samples below. *State* is common in both CLIL and non-CLIL biology tests, but not in history, where *name* and *mention* are more common, representing the same cognitive level. *Analyze* appear once in the non-CLIL tests, but no instances are found in the CLIL tests. However, analytical skills can be triggered by other wordings.

RESULTS

Table 11. Function words used in the test items

Function words	CLIL biology	Non CLIL biology	CLIL history	Non CLIL history
Encircle	1			
Match		1		
Put in order				1
Write in the right place	1			
Name	1	2	4	
State	6	6		
Mention			5	
Mark	1	1		
Give an example	1	1	1	
Define	1	2		1
Describe	2	1	4	1
Discuss	1	1		2
Explain	12	14	2	4
Draw		2		
Show	4	2		1
Compare	2	1		1
Motivate	1	2		
Analyze		1		1
Give arguments			1	

The CLIL history tests did not show very many instances of function words, but contain more question words instead, as will be seen in Table 12 below.

Table 12. Question words used in the test items

Question words	CLIL biology	Non CLIL biology	CLIL history	Non CLIL history
What	7	3	43	2
When	2	1		
Where			1	
Who			2	
Which	3	5	1	2
How	6	3	2	1
Why	2			1
Yes/No-question	1		2	

The question word in itself often signals a level of cognitive difficulty even though the wording of the rest of the question also has to be considered. *Why* and *how* are generally considered to represent more cognitively demanding

questions. Questions containing the question word *what* may display a continuum of difficulty depending on context; consequently it merits some special attention to be given below.

Looking at the instances of *what* and *which* above, it is worth noting that in Swedish the word *which* (cf. *vilka*) is more typically used than *what* in certain questions, which may explain the difference in the number of instances found. As regards the level of difficulty of the test items, it is not enough to consider the question words used; instead, an analysis of the test items is necessary as done below.

In the history CLIL tests, the total number of function words is higher than in the non-CLIL tests, 17 instances compared with 12 in the non-CLIL test items. Looking at both tables above (11 and 12), the difference in the number of instances of both function words and question words is notable. The number of question tests used for the count is the same, but the non-CLIL teacher often uses one question entrance containing the instruction, followed by several sub-questions without further question words, e.g. “Explain the following concepts”.

In the next section, an overview of course content and a more detailed description of the test items are provided, one discipline at the time, starting with biology. The description of the test items starts with the lowest cognitive level found and progresses toward more cognitively demanding questions. Academic function words and question words in the test items have been highlighted for easy identification.

6.3.3 Assessment in biology

Course description

The assessment samples represent a one-year course of biology, taught during the students’ second year of upper secondary school, called biology 1. A course book is used, in the CLIL class a book in English. Other teaching materials, such as handouts, pedagogic film, power point presentations, visuals in the form of laboratory exercises and field trips, are used in both classes⁷. The teachers present four written question tests each during the course, tests

⁷ Since classroom observations are not part of the present study, the information relies on reports from the teachers and observations from a few visits in some of the classes.

RESULTS

used to assess students' knowledge in relation to four content areas, three of which are identical in the CLIL and non-CLIL courses. Thus, three question tests from each context are used for the description below, tests representing the same content themes.

Biology CLIL & non-CLIL

- Structure and dynamics of ecosystems/Systems ecology:
Forms of life, energy flow, recycling materials, ecosystems, disturbances in the ecosystem, variation and composition of species
- Genetics:
Cell division, DNA, gene expression, heredity and environment, genetic applications
- Evolution and diversity:
Origins and development of life, natural selection, behavior of organisms, taxonomic systems

For assessment purposes, the teachers also mention the use of writing assignments in the form of laboratory reports, field notes, other reports and oral activities in class, especially related to laboratory assignments. The design and content of these are discussed below. Next follows a description of the content of the test items.

Content of the test items

Selected response/SR

An SR-question usually requires little, if any, production of language. It does not offer much context, but relies on the test-taker's memory and understanding of the individual concept. It may, as in the example below, offer multimodal scaffolding in the pictures; each phylum is illustrated by a related symbol. The function word *combine*, signals lower order thinking skills, typically generating an E, the lowest grade. However, the test item below may generate an E or a C. The maximum score is seven, although ten characteristics are listed.

ASSESSING LANGUAGE OR CONTENT?

Example 1/Bi CLIL:

Combine pictures with the right characteristics. Put the right letters in the boxes underneath the phyla below. E/C (5/2)

- | | |
|-----------------|----------------------------|
| A) No tissue | E) Tissues |
| C) Coelomate | D) Acoelomate |
| B) Diploblastic | F) Radially symmetrical |
| G) Asymmetrical | H) Bilaterally symmetrical |
| I) Sessile | J) Triploblastic |



Characteristics phylum 1	Characteristics phylum 2	Characteristics phylum 3

Likewise, in an example of a selected response test item from a non-CLIL test, the test-taker is requested to match concepts with letters to fit them into the right place in a figure.

Constructed response/CR

Most of the questions in the present material in biology are CR questions at different cognitive levels. Examples of less cognitively challenging questions include naming and completion questions, e.g. “**Name** the plants”, accompanied by pictures. In an example in a non-CLIL test, students are supposed to write a name on a line next to each picture, thereby identifying “**which** phylum” the animals belong to. Both these examples represent test items at the E-level. They are slightly more demanding than the SR question in example 1 above, since the student has to remember the names of the actual phylum. The question-word *which* signals an easier cognitive level. Some slightly more cognitively and linguistically demanding short answer questions very often involve questions asking to briefly *explain* concepts. They appear both in CLIL and non-CLIL tests. They are not context-embedded to a high degree, but rely on the test-taker to remember facts, i.e. subject-specific concepts, often generating no more than a C.

RESULTS

Short answer questions may contain a mix of different cognitive levels within the same test item and a mix of function and question words, as seen below. In Table 11 above, *describe* and *explain* appeared to be the most commonly used function words, here used in the same question.

Example 2/Bi CLIL:

The nitrogen cycle (E, C, A) 2/2/4

- a) **Why** is it so important for a living organism to be part of the nitrogen cycle? What is N_2 used for in life?
- b) Nitrogen, N_2 , is a major part of the atmosphere. **Describe and explain** how nitrogen can transform into forms for living organisms to use.
- c) **How come** the level of N_2 stays the same?

Example 3/Bi CLIL

Dogs (E/C/A)

Among cocker spaniels the colour of the fur is inherited from two different loci at two different chromosomes. The phenotypes are:

AB black colour Ab red colour aB brown ab yellow colour

The black cocker spaniel Lufsen mated with Lady, a beautiful yellow coloured female. They got a yellow puppy. Not long after Lufsen mated with the she-dog Black Lady (according to Lady a real bitch), who has got the same genotype as himself.

- a) **State** the genotypes for Lady and Lufsen respectively.
- b) **What** gamets can be produced by Lufsen?
- c) **Could** Lady and Lufsen's puppy turn out to have another colour than yellow? **Explain** your statement.
- d) **How** big is the probability that Black Lady and Lufsen give birth to a brown puppy?

Explain your statement.

The question words, *why* and *how come*, used in example 2 above, entail higher order thinking skills since the test taker is expected to analyze and apply

ASSESSING LANGUAGE OR CONTENT?

knowledge. The test item may generate all grades, E-A. Once again a similar test item can be found in the non-CLIL test.

In example 3, the answers to the first three sub-questions represent lower order thinking skills. Sub-question number three is a yes-/no-question, but the urge to *explain* makes the expected answer slightly more demanding. The fourth sub-question is introduced by the question word *how*, which is considered to represent a question demanding higher order thinking skills. In this question, the student may attain the highest grade, an A

In the non-CLIL tests, there are some examples of essay questions where a longer answer is expected, as in the following example:

Example 4/Bi non-CLIL:

Evolution (E/C/A)

- a) **Give an account of how** scientists think life evolved during the early years of earth's history.
- b) **How** has it been possible to illustrate this in an experiment?

In the example above, the test-taker has a full page to answer this single question, an indicator that a fairly long answer is expected. The test item may also generate all grades. Out of this test's seven items, five are essay questions. The CLIL test on the same topic does not contain long essay questions.

Production tests

Except for question tests, other types of writing assignments are used for assessment purposes. The assignments are similar to essays and represent more extensive text production and genre writing. In the biology courses, CLIL and non-CLIL, these consist of laboratory reports and reports from excursions. In the non-CLIL course, students receive guidelines for what to include in a report. A report often follows a model, but the model can vary depending on content and scope. There is a description of the language to be used: "formal writing with objective and focused content". The language should be correct, free from slang expressions, the words are to be chosen with care and exact concepts and terms may need to be defined. (Guidelines for reports, school B.)

The structure and content of a report should contain the following: informative title, purpose/problem, background/theory, material/method,

RESULTS

outline/implementation, results/observations, conclusion/explanation, discussion and sources.

The guidelines for written reports also include knowledge requirements: knowledge of biological concepts, models and theories, ability to analyze and look for answers to disciplinary questions, to identify and solve problems, and ability to use knowledge in biology to communicate. The accurate use of academic language is also described as a prerequisite for different grade levels.

The following example is an extract from a laboratory report in the CLIL biology course. The first part of the assignment consists of background information, as seen in the quotation below, as well as instructions for the task. The assignment is done in groups in class. The topic is evolution:

Scientific theories about the origins and development of life. Evolutionary mechanisms, such as natural selection and sexual selection and their importance in speciation. Behaviour of organisms and the importance of behaviour for survival and reproductive success.

The students are supposed to use various tools representing different animals: chopsticks, tweezers, fork and spoon to “chase pasta” on a table. “Who gets the most?” The results should be written in three tables before students are asked to observe and draw conclusions of their own:

Example 5/Bi CLIL:

Conclusions – show your line of argument.

What conclusions can you make due to the information in the tables?

Explain the results.

Use the theory of natural selection and the following key terms: competition, extinction, adaption, fitness.

Does it matter which student got a certain device?

Motivate your answer.

Did time matter? **Motivate** your answer.



The laboratory report is highly context-embedded in classroom practice, making it somewhat difficult to grasp the entire task without observing the procedure. Four disciplinary concepts should be included in the answer. A theoretical framework is stipulated, i.e. natural selection. The writing assignment, even though context-embedded and scaffolded in the classroom exercise, requires the student to make inferences and to interpret the evidence, implying that students have to use broader linguistic registers and higher order thinking skills.

In the next section the assessment samples from history are dealt with in a similar way, first the design and then the content of the test items.

6.3.4 Assessment in history

Course description

The assessment samples represent a one-year course in history, taught during either students' first or second year of upper secondary school, called history 1b. A course book is used, in the CLIL class a book in English. Other teaching material used, according to the teachers, consists of handouts, film and power point presentations.

The teachers present four written tests each during the course. The non-CLIL teacher uses four paper-pencil question tests and one oral assignment for summative purposes. The CLIL teachers have four written tests or writing assignments each, to assess similar course content. They also indicate that they use oral assessment forms. Four question tests from the two CLIL teachers and the non-CLIL teacher, respectively, have been used for the analysis, together with writing prompts. The content areas in the tests vary to some extent. The following themes are used in the present sample material.

History non-CLIL:

- Eras and source criticism = Question test
- The double revolutions (the 18th century and enlightenment) = Question test
- The interwar years = Question test
- World war II = Question test

RESULTS

History CLIL (School C):

- Antiquity and the medieval period = Question test
- 16th, 17th and 18th centuries in Italy, Germany, France and the US.
= Production test

History CLIL (School A):

- Ancient Greece and Rome = Production test
- The industrial revolution = Question test
- World War I = Production test
- World War II = Question test

Content of the test items

Selected response/SR

Only one question represents the SR category, being found in the first test in the non-CLIL history course. This test item is considered less demanding by the test designer, i.e. the teacher, and can only generate an E, the lowest grade.

Example 6/Hi non-CLIL:

Place the following events **in chronological order**. The timeline starts year 1000 B.C. and ends 1500 A.C. (E)

- a) The Black Death
- b) Christianity becomes the official religion in the Roman Empire
- c) Alexander the Great conquers the Middle East
- d) Sweden, Denmark and Norway are united in the Kalmar Union
- e) High Middle Ages start in Europe
- f) Julius Caesar becomes a dictator in Rome
- g) Athens is the leading city-state in Greece

The original language is Swedish, and the student only needs to recall information, identify and list the events in the right order.

Constructed response/CR

The CLIL tests include only constructed response items, as do the rest of the non-CLIL tests. However, they vary greatly in their range of complexity, as will be seen below. None of the non-CLIL tests use test items requiring only one word. Both of the CLIL teachers' tests do.

The use of the question word *what* was discussed in connection with Table 12 above. The example below represents a question of *low cognitive demand*, where the student only needs to remember and recall names. It is found in the same CLIL test as the previous question. The same test contains 34 instances with the use of *what* in similar types of items, requesting a name in most of the cases. As noted previously, this test generates a C at the most; no higher grades, A or B, can be attained. The scoring has not been described, but the two names give four points.

Example 7/Hi CLIL:

What was the popular name for the young Macedonian king who was in charge of the Macedonian troops and personalized this development? He had also a very famous teacher, almost as famous as himself. **What** was his name? (4)

Answer:

The king: A_____

His famous teacher: Ar_____

A similar function word requiring a response of low cognitive and linguistic demand is the verb *name* in a CLIL-test item: “**Name** three reasons why”. The expected answer to this test item may require the use of several sentences, but the teacher does not want the student to describe the reasons in any depth. A total of three points may be awarded.

The next example shows another instance of an item using the question word *what* and comes from one of the other CLIL tests. As seen from the example, the teacher specifies that a lengthier explanatory answer is required. The first part of the question deals with the understanding of a concept. The test item provides some context, but the question gives rather little scaffolding unless you are familiar with the word *appeasement*. The answer requires a fairly long explanation, and can generate four points.

RESULTS

Example 8/Hi CLIL:

More explanatory answers needed. Points in brackets.

What does the concept appeasement mean? Also **give an example** of **when** one can say that the UK and France used that way of acting. (4)

Other test items including the question word *what* usually generate low grades or scores. Examples found generating one point or an E, the lowest grade, include questions such as “**What** was the proletariat?” (CLIL) or “**What** is the difference between stories and remains?” (non-CLIL). Both of these questions could generate a short answer or a longer essay, but the grade/score signals the first, even though the question is open-ended. No context is provided so the student has to know the concept.

In another example from a CLIL test, the question “**What** happened that made USA join the war?” may give two points and is found under the heading “Descriptive questions”. Again, the scores signal the scope of the expected answer. This question, just like the previous one, is open-ended and could represent a writing prompt where a sequence of events could be reported, thus representing an essay question.

Two of the four non-CLIL tests have a sort of dual design, where the test items in the first part require explanation of concepts, and the second part contains essay questions, requiring longer answers. In one of the non-CLIL tests, the item type requesting explanation of concepts may only generate an E; in the other tests, no scoring is mentioned. The essay question in one of the tests may generate any grade on the scale, F-A. In the other tests, no such information is given. One explanatory question contains a total of twelve concepts to be explained, representing World War II and the Cold War: “Explain the central concepts in three to four sentences”. The answers are to be written on a separate sheet of paper. Examples of concepts to be explained are The Spanish Civil War, The Molotov-Ribbentrop pact, concepts which could constitute writing prompt for essays.

Below is one example of a test item from a non-CLIL test. The essay-question requires both a supposedly lengthier definition, on a separate sheet of paper, of the concept “totalitarian” and higher order cognitive skills in order to be able to *compare* two ideologies. This, however, is done in Swedish. The test item is accompanied by two pictures and a caption, providing some context.

ASSESSING LANGUAGE OR CONTENT?

Example 9/Hi non-CLIL:

The interwar period's political and economic alternatives.

The interwar period implied the emergence of two totalitarian states in Europe, the communist Soviet Union and Nazi Germany.

Define what totalitarian means and next **compare** the two states and their ideologies with one another (communism and Nazism).

The next example is from one of the CLIL tests. It is found in a test on World War II and is context-embedded but in turn contains several sub-questions. The answer is supposed to be analytical and argumentative expressing the student's opinion. The test item requires both higher order thinking skills and more demanding linguistic skills, since a thorough answer is needed, containing both subject-specific concepts and content-compatible academic genre words.

Example 10/Hi CLIL:

Level requiring a more **analytical** answer. Try to answer as thoroughly as possible and, if needed, point out what your opinions are.

The Holocaust is the name given to the atrocities during which the Nazi regime in Germany systematically killed millions of people, not least Jews. It is easy to see the responsibility of the Nazis in this. But **what** responsibility did the Germans as a group and as individuals have concerning the Holocaust? And **what** about other countries... to what extent could we say that other countries had a responsibility and could have acted in ways that perhaps could have prevented or stopped the Holocaust?

Try to **give arguments** based on for example...

- your knowledge of people's awareness of the treatment of Jews
- your knowledge of how other countries acted towards Germany and Hitler

The next section deals specifically with writing assignments, categorized as production tests.

RESULTS

Production tests

Essay-questions are often found at the end of the question tests, thus implying written production. Teachers distinguish production tests, or essays, from mixed question tests. Sometimes writing prompts are used together with short answer questions in the same tests. The non-CLIL teacher only uses the mix in the rendered tests, and no “pure” essays or writing assignments are used for assessment purposes. Consequently, the two examples below are both found in CLIL tests, one from each teacher. The first example represents one out of three prompts, called “First scene” in a test called “history novel”.

Example 11/Hi CLIL:

Believe it or not but you are a famous 16th century talk show host in Italian TV. At last you are able to present one of the most well-known and controversial characters in your time, the 80 year old Florentine multi-skilled artist Michelangelo Buonarroti. You have a lot of questions about his artistic dreams, his many conflicts with both profane and spiritual rulers and his secret private life. This article will be about 800 words and written as a **speech manuscript** with a frequent use of quotation marks. (English)

The prompt describes the expected genre, a speech manuscript. It provides a great deal of context, but the test item is fairly open-ended. The student is supposed to include questions and quotations, suggesting that he or she has come across the relevant genre and context in the course. The genre is basically narrative in character rather than analytical and argumentative, requiring a specific type of language.

The example below represents a somewhat different test design, although the impact of the surrounding society is included here as well. The example comes from a production test in the other CLIL context. This is not a test done during one class, as the previous example. The assignment should be worked on and handed in several weeks later. Students receive a three pages long description of the assignment, including a page giving the relevant knowledge requirements, mentioning that “students can [...] give an account of processes of change, events and people from different time periods”. Topic suggestions and instructions regarding length and which font to use are given in the instructions. The suggested topics cover a range of themes mentioned under the headings Individuals, Groups in society, Wars, Entire societies, Concepts and processes, all of which are exemplified.

ASSESSING LANGUAGE OR CONTENT?

Example 12/Hi CLIL:

After studying Ancient Greece and Rome during a couple of weeks, you now get a chance to work on your own within this field of studies. You will write about some person/event/phenomena from the two civilizations mentioned above – both providing facts and trying to find what kind of impact your chose subject has had on the world since the ancient of times.

[...]

What to include?

Facts: Provide facts to describe your subject. This will of course look a bit different depending on your choice of topic. What would **an encyclopedia article** on your subject include? Use that as a guide on what to bring up. This is seen as a strictly descriptive part and needs to be well sourced.

Influence/impact: **In what ways** was your subject influenced by earlier times? And **how** has it influenced later times? [...] This part is more analytical and gives you the chance to make comparisons between cultures/times [...] This part might consist of a mix of your own thoughts and information taken from sources.

A writing assignment, as found in the above example requires certain skills from students. Except for linguistic and cognitive skills, metacognitive skills and skills in the writing process are needed: how to plan the work and how to compose the text (cf. section 3.6.2).

Below follows a description of assessment samples from the English language courses, mostly consisting of national test samples.

6.3.5 Assessment in English

Six teachers of English participated in the interviews, two from each school. Two teachers teach both CLIL and non-CLIL students, depending on which classes they are assigned every year. One more teacher, who only participated in the interview, represented a non-CLIL EFL teacher; the rest teach in a CLIL context. By the end of the study, only one of the original six teachers had handed in assessment samples. The two participating teachers at that school claimed to follow more or less the same plan, using the same assessment material. One of the CLIL teachers at the international school presented portfolio prompts used in English 7, but no instructions or grading criteria for those written assignments.

RESULTS

Below is a brief course description of the English language courses, including a summary of the reported course material and assessment types. Due to the small sample, and the alleged importance of the national tests, a brief description of the writing assignment in the Swedish national test is included. The Swedish national test is compulsory during the first and second year of English at upper secondary school thus representing a test design and content to which all students are exposed.

Course description

The teachers and the assessment samples represent all of the three English language courses taught at upper secondary school, English 5, 6 and 7. Since the interviews provided such varied pictures of the courses, Table 13 gives an overview of differences and similarities between the three schools and the potential influence of the CLIL approach:

Table 13. Course material and assessment types in the English language courses

	School A	School B	School C
Course material	Novels, film, articles	Textbook: “Blueprint” Novels, film, articles, field trip to England	Textbook: “Context” Novels, film, articles
Assessment types (written)	Portfolio Written production National tests Tests	Old national tests Book presentations Vocabulary test Grammar test Realia test Essays	Old national tests On line quizzes Vocabulary quizzes/test Written production Grammar test

A course book is used in two of the schools. The teachers at the international school (A), do not use a typical course book since the textbooks found in the English as a foreign language (EFL) classrooms usually have wordlists translating words from English to Swedish, which these teachers did not find suitable.

Other teaching material used, according to the teachers at all three schools, is represented by novels, film and articles. All of the teachers use old national tests for practice and preparation. In English 7, teachers recognize students’ higher proficiency level and use advanced writing assignments, where the main focus is on producing different academic text genres.

Design of the tests

A range of designs are used depending on content and purpose. Completion tests with fill-in-the-blank test items are used to assess vocabulary and grammar while multiple-choice tests are used to assess cultural knowledge about English speaking countries. These tests are not multi-modal; they only contain text. A flag representing the country can be found on top of the tests on Great Britain and the USA. The grading of the tests is done using scores: one point for each correct answer.

Old national tests are used to assess listening and reading skills. These tests consist of matching, multiple-choice, completion and short-answer questions. Two to four pictures are included in each test. The tests are divided into sub-parts, each with individual scoring. The production tests typically contain a longer description of the assignment, sharing information on form/genre/type of text, topic, and content. Grading criteria and instructions what the assessment will be based on are usually included. Writing assignments used in the different courses during the three years of upper secondary school include the following:

English 5:

- Book presentations, film reviews (CLIL/non-CLIL)
- Portfolio (CLIL):
- A narrative text, e.g. short story, poem, texts based on personal experiences
- An informative text, e.g. news article, topic summary, personal statement, statement of intent, presentation slides (possibly an argumentative text)
- National test: essay (CLIL/non-CLIL)'

English 6:

- Argumentative essay, discussion essay (CLIL/non-CLIL)
- Portfolio (CLIL):
- An argumentative text, e.g. personal project, argumentative essay, formal letter
- An analytical text, e.g. history assignment, film/book analysis, blog posts, short answer
- National test: essay (CLIL/non-CLIL)

RESULTS

English 7:

- Novel analysis, exploratory essay, argumentative essay, university application (CLIL/non-CLIL)
- Portfolio (CLIL):
- All of the above-mentioned (English 5 and 6)

Content of the test items

In this section where English is in focus, no frequency count regarding function and question words was made. Instead, a couple of examples of SR test items follow, for cross-disciplinary comparative reasons. No CR items were found. The most important investigation concerns the content of the writing prompts involved in the production tests.

Selected response/SR

In the question tests used at one of the schools, the items in the vocabulary and grammar tests are usually represented by completion questions, i.e. a gap and a word in parenthesis in need of translation, as in the example below:

Example 13 En CLIL/non-CLIL:

The guitar has _____(kosta⁸) over £100.

Each test consists of 85 to 100 similar test items, usually generating one point per correct answer. The sentences are mutually independent, so there is no real coherence or context shared between them. The test taker is supposed to know the word to be able to make the translation into English.

Other test items, in one of the tests, consist of sentences to translate from Swedish into English. In one of the grammar tests, the suggested words in parenthesis are sometimes in Swedish, sometimes a word in English to inflect. This implies that students have to do some code-switching while taking the test.

The next example represents a typical item from the knowledge tests on English-speaking countries. This one comes from the test on the UK.

⁸ Kosta = Cost in Swedish

ASSESSING LANGUAGE OR CONTENT?

Example 14 En CLIL/non-CLIL:

In the following tasks (1p each), circle the correct alternative:

We know from history that England has been attacked by

1. Vikings, Romans, Angles, Jutes and Saxons

X. Vikings, Romans and Russians

2. Vikings, Romans and Chinese

The above example is a multiple-choice item. The entire test consists of 17 such items, one including matching of eight concepts represented by countries and geographical terms. The above test item is not linguistically or cognitively very demanding. The content is interdisciplinary, involving geographical terms.

Since the research question in the present study regarding the English language courses concern whether or not the content is affected by the CLIL profile of the school, the national tests are not described. However, the writing part represented by a writing prompt and the production of an essay will be presented below. This is done to be able to compare the expected proficiency level of students in EFL compared to expected writing proficiency in the subject content courses.

Production tests

Writing assignments and essays are a natural part of the English language courses at all levels: only the topics differ. For English 5, where students should have reached the B1 proficiency level according to the CEFR, so called familiar topics are used. In English 6, more academic topics and formats are used. This can be seen in the national tests, as exemplified below:

Example 15, English 5:

Music means a lot to most people. We are surrounded by music; in the supermarket, on the bus, at work and at home. Can you imagine life without it? Do you prefer to listen to classical music, or is pop/rock music your cup of tea? How does your taste in music reflect your personal life-style? Do you ever go to live concerts, play an instrument or sing in a choir?

Your task is either to write a text about a specific musical memory (Topic 1) or to discuss what effects music can have (Topic 2).

RESULTS

During students' second year of English at upper secondary school, the character of the theme is somewhat different, in line with the curricular goals. Students are supposed to present arguments and the theme is "Temptations". Six titles are suggested, e.g. "Smoking ruins your health", or "Gambling might lead to trouble".

Example 16, English 6:

Try to convince your reader that your position is the right one and remember to bring up some of the counterarguments as well. **Define** your issue clearly. **Develop and support your arguments** with examples.

In the English language courses at the schools in this study, the same kind of progression in text genres can be seen. In the courses English 6 and 7, argumentative and analytical tasks are used. The topic, according to one of the assignments in English 7 at school B, can be one of the following: politics, society, religion, literature, film, art in an English-speaking country of the student's own choice. The text should measure 4-5 pages, as well as a title page and list of references.

To conclude, written assessments in the EFL courses usually involve some genre-based writing assignments. The more advanced the proficiency level, the more academic and subject-content-oriented the assignments get. English 5 deals more with narrative and with personal experiences and thoughts, whereas English 6 is more argumentative and analytical.

6.4 Questionnaire

Nine of the participating teachers responded to the questionnaire. The purpose of the questionnaire was to complement the interviews and the evidence of teachers' practice found in the assessment samples. Consequently, questions perceived as adding to the results were selected. For the full quest of the questionnaire see Appendices 2 and 3. Two individual questions are presented below, followed by a summary of some general attitudes. The two individual themes refer to different *types of assessment used* and *primary focus in assessment* in relation to the teachers' disciplines.

6.4.1 Types of assessment used

One of the key questions concerns what basis for assessment teachers use for summative and formative assessment and which assessment types they find most useful for grading. The purpose was not to dichotomize the two types of assessment, summative and formative, but rather to cover teachers' assessment practices in the best possible way. Apparently, in spite of extensive discussions about formative assessment in recent years in Swedish schools, there is still some uncertainty surrounding the topic. One of the teachers comments, when describing her summative assessment forms: "The students are more summative than I am, they want to know 'where am I now'". This, however, may refer to assessment for learning, as in formative assessment.

Two of the English language teachers do not mention what types of assessment they use, but one of them comments on assessment in general when asked about summative assessment types:

Having worked a lot with other tests of English such as IELTS and CAE/CPE as an examiner, I question the quality of the National Tests in some respects. Generally, I feel that the reading and writing sections are of good quality, but that the listening section leaves A LOT to be desired if compared with the Common European Frame of Reference for Languages which is underlying the course goals for English in GY11⁹.

Table 14 below presents the individual teachers' responses also providing an overview of subject-specific features. All of the teachers claim to favor written assessment over oral, even though two of the English language teachers and one of the CLIL history teachers says they use speech and oral presentations. For EFL teachers, it is mandatory since they are supposed to assess oral skills. The non-CLIL history teacher claims to use oral checks in the classroom for formative purposes.

In response to the next question, what basis for assessment is found most useful when grading, there are three options: written, oral or both. Three out of nine teachers claim they find both written and oral assessment equally useful, whereas the rest, i.e. the majority, respond "written". Comparing CLIL with non-CLIL teachers, the latter seem to rely more on what one of them refers to as "traditional" written tests, which is also the case for the non-CLIL biology teacher as well as the non-CLIL history teacher.

⁹ GY 11 = The curriculum for the upper secondary school which was introduced in 2011.

RESULTS

Table 14. Overview of teachers' responses regarding assessment types.

Teacher	Summative assessment	Formative assessment
A1 CLIL English	(Questions the national tests, especially the listening)	Use points/scores, not letters to prevent the student from focusing too much on a grade
A2 CLIL English	-	-
A3 CLIL history	Written tests, essays, home assignments, homework	Written feedback and talks with students about assignments made. Looking at course goals and the student's previous level, discussing what can be done to reach a higher level of understanding
B1 CLIL/non-CLIL English	Essays, comprehension, realia, grammar, speech	Student presentations
B2 CLIL/non-CLIL English	Written assignments ; exploratory and argumentative essays, literature response-papers, reading and listening comprehension, speech, oral presentations,	The same as for summative
B4 non-CLIL biology	Written tests	Lab- and excursion reports
B5 CLIL biology	Written tests and matrices based on the knowledge requirements	Labs, problem-solving individually or in a group with peers, excursions, written reports, discussions
C3 CLIL history	Tests, hand-ins, essays, oral presentations etc	Process-oriented, continuous tutor dialogue about individual assignments/exams. The dialogue shows the individual student's development in relation to the goals
C4 non-CLIL history	Traditional tests, essays, home exams	Oral checks in the classroom, written feedback on assignments, individual talks with students

A couple of teachers choose to comment on written versus oral forms of assessment. One of the CLIL history teachers says that by oral he does not refer only to class presentations, but also to oral discussions in small group settings. The other CLIL history teacher states that, in order to find proof of a deeper level of understanding, the oral discussion has to reach a certain level of cognitive complexity, which is hard to achieve in a class of thirty students while ensuring validity and reliability. The CLIL biology teacher claims that oral proficiency in class and during laboratory assignments is important, but

ASSESSING LANGUAGE OR CONTENT?

biology rests largely on written skills, even though the oral and the written go hand in hand.

Other comments concern the time-consuming aspect of oral assessment forms, and that it is easier to rely on written documents, to quote one of the EFL teachers:

More guidance should be provided for non-native teachers of English when it comes to judging student's communicative abilities. I have seen a lot of variation in what is seen as passing when it comes to the students.

To conclude, teachers rely more on written assessments due to validity and reliability concerns.

6.4.2 Most important factors in assessment

The question “What factors are most important when assessing students’ skills” refers to both knowledge and skills. Nine options are provided to see if any differences can be found relating to disciplinary features or the teacher’s cognition. The nine response options emanate from curricula and course goals in the different disciplines. Other factors might be of relevance, but these were chosen since they were considered to represent a valid sample of the linguistic and cognitive skills involved in the disciplines in the present study.

Table 15 presents the teachers’ responses, with EFL teachers together at the top, followed by teachers of biology and history, in order to facilitate comparability between disciplines. Worth noting is that there were no restrictions as to how many options the teachers could choose. One of the EFL teachers only picked one alternative, whereas the rest chose anything from two to seven.

RESULTS

Table 15. What teachers claim to be most important when assessing students' skills.

Teacher:	Written & oral performance	Subject content	Use of subject-specific language (CO)	Use of general academic language (CC)	Mastery of various forms of expression and IT	Linguistic accuracy	Linguistic complexity	Mastery of disciplinary written genre	Analytical skills
	A	B	C	D	E	F	G	H	I
A1 CLIL English	X				X				
A2 CLIL English	X								
B1 CLIL/non-CLIL English	X		X	X	X	X	X	X	
B2 CLIL/non-CLIL English	X			X			X	X	
B4 non-CLIL biology	X	X	X					X	X
B5 CLIL biology		X	X				X	X	X
A3 CLIL history		X	X						X
C3 CLIL history	X	X						X	X
C4 non-CLIL history	X	X	X						X

Alternative E, mastery of various forms of expression and modern information technology, is articulated as an aim in the course goals. This does not appear to be one of the most important goals in assessment among these teachers. Alternative F, only one of the respondents, an EFL teacher, believes linguistic accuracy to be a main concern in assessment of students' skills, which is noteworthy.

Another interesting result for this study is found in column D: none of the content teachers consider the use of general academic language to be one of the most important goals. It is also interesting to note that all teachers but the CLIL biology teacher and one of the CLIL history teachers believe oral and

written production (A) to be important when assessing students' skills. Further, analytical skills (I), are chosen only by the content teachers.

Below is a summary of the teachers' views on assessment as seen in the other questions in the questionnaire.

6.4.3 General views on assessment

All teachers, except one, consider national course goals to be very important when grading. Only one of the CLIL history teachers views them as only fairly important. In the interviews, she described them as gates in downhill skiing that you should stay inside, referring to them basically as points of reference.

When given four alternatives, as to what they include in their assessment, the CLIL history teachers state *only content*, whereas none of the teachers claim to include *only language*. In the interviews, on the other hand, several of the EFL teachers claimed to assess only language. Three of them report that they include both language and content, and one that it depends on the task. Both of the biology teachers, CLIL and non-CLIL, claim to assess both content and language, commenting that language refers to "terminology".

Five out of the nine teachers find grading in their discipline difficult or fairly difficult. They include both of the biology teachers, CLIL and non-CLIL, both of the CLIL history teachers and one of the English language teachers, the same who claimed in the interview that external examiners should do the grading. The other EFL teachers state that grading is fairly easy.

In response to the question in what areas they would want further training, two teachers state that they do not need to learn more; namely one of the EFL teachers and one of the CLIL history teachers, the latter also acknowledging that grading is rather difficult. The other CLIL history teacher claims she wants to learn more about how to interpret national course goals, test development and how to assess written production. The CLIL biology teacher states that he wants to learn more in all suggested areas (see question 20 in Appendix 3). Four of the teachers want to learn about alternative forms of assessment: the non-CLIL content teachers, the CLIL biology teacher and one of the EFL teachers. One possible interpretation of this is that the other teachers feel the tools they already use are sufficient.

RESULTS

6.5 Summary

In a comparison of assessment practices described by the teachers, there are both common features and differences, but not specifically due to the CLIL approach. The English language courses seem rather unaffected by CLIL; there seems to be no or little influence from the subject courses on the content of the English language classrooms.

The two biology teachers show a great deal of consensus concerning how to assess as well as what to assess. They collaborate to some extent when planning courses and use the same tests, more or less, translating some of the test items or questions into English.

The history teachers show greater variation in their assessment practices than the science teachers, apparently not due to different schools, but rather to differences in teacher cognition. Regarding similarities between the three, the two CLIL teachers show more common features, favoring essays over question tests.

The English language teachers are very similar in displaying greater intra-disciplinary variation, i.e. they seem to use a larger mixture of assessment tools in their courses. The national tests provide the common denominator. All of the teachers use them as a frame of reference and sometimes use old tests for their formative assessment and as preparation for the “real” tests.

To conclude, assessment practices seem to vary depending on several converging/intersecting factors:

- The character of the discipline
- The existence of national tests
- The level of the course, whether English 5, 6 or 7 (see section 5.4.2)
- The L1 of the teacher, whether (s)he is a native English speaker
- The school context, especially when an international school context
- The teacher’s personal preferences/experience

Some of the factors listed above are a natural part of the assessment procedures, such as the level of the course and taking the skills and proficiency of the learner into account. Others, as the last three, should not be parameters in assessment for validity reasons.

The next chapter provides further analysis and discussion of the findings, combining and delving into the connections between the results from the different data collections, as well as performing a validation process.

7. ANALYSIS AND DISCUSSION

7.1 Introduction

In this chapter, the results accounted for in Chapter 6 are analyzed in relation to the research questions formulated in Chapter 1: whether there are differences in assessment format and practices due to the CLIL approach and the language of instruction; whether CLIL has any effect on the course content and assessment in the English language courses; and finally if it is possible to distinguish any cross-disciplinary similarities in assessment. The analytical method is to triangulate the findings from the different types of data collection carried out in the study.

7.2 Comparing assessment in CLIL vs non-CLIL

The first research question focuses on possible effects of the language of instruction on the assessment methods in the subject content courses:

- CLIL vs non-CLIL, do the assessment practices differ in the two subject content courses history and biology due to the language of instruction? If they do, *how* do they differ, and *on what grounds*?

In the first chapter, Figure 1 served to present the different layers involved in the study. These represent components that have an impact on assessment design and what skills are to be assessed, i.e. the construct. In a CLIL context, where no CLIL curriculum exists, the validation of the intended, enacted and assessed curriculum becomes even more important. The national curriculum and course goals are mandatory for all contexts to help avoid variability in the quality of education. However, continuous validation is necessary. Many variables are involved, as seen in the present study, including subjective interpretation of policy documents, which constitutes a threat to validity. The parameters in this study involve teacher cognition, course goals and disciplinary tradition, representing both a macro and micro perspective. The

ASSESSING LANGUAGE OR CONTENT?

macro perspective is present in the course goals, and the micro perspective relies on teachers' perceptions of the goals, the discipline and CLIL, where applicable.

The intended curriculum in CLIL, since no specific CLIL curricula exist, is the same Swedish national syllabus as in the non-CLIL setting. The enacted curriculum differs, at least when it comes to the language of instruction. For the consequential validity (Messick, 1989) of the CLIL approach, it is of interest to find out if the assessed curriculum is affected by the use of an L2, as expressed in the research question quoted above.

Using Biggs' (2003) constructive alignment model, the intended learning outcomes, ILOs, are formulated first, from which assessment criteria and assessment design are derived. Accordingly, the role of course goals and national objectives is brought up by all teachers. This may reflect the teachers' desire to answer professionally, or reveal the status and importance of national policy documents among Swedish teachers. Regardless of which, all teachers acknowledge their importance in the questionnaire, thus confirming the results of the interviews. No difference is found between CLIL and non-CLIL teachers in this regard. The subject content teachers are very careful to point out that the same national standards should be used to cater for the validity of assessment in CLIL contexts. However, the purpose and consequences of bilingual teaching need to be articulated and problematized in order to define appropriate assessment procedures. The ILOs in CLIL are represented by content and disciplinary language learning as well as enhanced language use, which differs from FL learning (Nikula, 2007). In combination with content, students will encounter and use a wide range of the target language and academic functions (Dalton-Puffer, 2007; Llinares et al, 2012), broadly referred to as "the language of schooling" (Schleppegrell, 2004).

The CLIL teachers express certain concerns about the possible impact of the use of English on assessment outcomes. Different accommodations are offered, sometimes by allowing the use of Swedish, providing dictionaries or providing more time on tasks. Another strategy is to offer students the opportunity to make clarifications afterwards if ambiguities appear.

In the previous chapter, it was noted that no evident differences were identified in the assessment practices due to the CLIL approach. Instead varying assessment strategies could be explained by teachers' different preferences and perceptions of their disciplines. In order to examine "on what grounds" teachers' assessment practices are based, the context of the

ANALYSIS AND DISCUSSION

individual teachers and the prevailing societal and disciplinary views need to be identified (Bachman, 1990; Hofer & Pintrich, 1997; Inbar-Lourie, 2008; Shephard, 2000; Tittle, 1994). Shephard (2000) argues that traditional views of assessment are entwined with models from the past still affecting the prevailing practices of teachers.

Nikula (2007), referring to CLIL in Finland, notes that the diversity in the forms of CLIL poses a challenge for research, making it difficult to draw firm conclusions. In the present study, the small sample represents another restriction. However, drawing on the current observations, the personal preferences on the part of the teachers become apparent.

The assessment practices differ more between the CLIL teachers in history than between CLIL teachers in biology. This may be explained in part by “the fuzzy” history discipline (Pace, 2011), leaving teachers without clear assessment strategies. History is described as narrative in character (Rosenlund, 2011), which can be identified in teachers’ reports, claiming that “history is basically a literary genre”. One of the CLIL teachers prefers the use of narrative texts for assessment purposes, but without using rubrics. He believes “storytelling comes first” and progression can be found in the dialogue with students. Since he claims not to include language in the assessment, it would have been interesting to find out more about the effects of students’ use of English in this situation. This, however, is beyond the scope of the current study.

His CLIL colleague leans more toward structured texts “to help clarify what is required”. His non-CLIL colleague favors question tests and argues that the history discipline is based on how well students deal with questions, draw conclusions and use explanatory models and terminology. In order to identify progression, he prefers the use of questions at varying levels of complexity. The use of questions also refers to the students’ own use of them, as a sign of analytical thinking, as mentioned by all teachers of history.

Alderson and Banerjee (2002) argue that the more structured the more reliable assessment gets, which in this case might speak for the use of question tests rather than written production, especially if no task-related knowledge requirements can be presented. This is a validity concern, and relates to the opening quotation in the first chapter: to be held accountable and be able to demonstrate the intended uses of an assessment (Bachman & Palmer, 2012). On the other hand, if the goal is for students to show analytical skills in more lengthy lines of reasoning, production tests may be more appropriate, as

ASSESSING LANGUAGE OR CONTENT?

argued by one of the CLIL history teachers. Clear performance criteria, offering consistent standards (Hyland, 2007), should describe how the analytical skills will be evaluated in production tests, to support the validity of the inferences made from test scores (Messick, 1989). The targeted skills, including both the cognitive and linguistic skills of the students, have to be considered, in relation to the features of the task, as seen in Figure 10 in Chapter 4.

As regards language, one of the history CLIL teachers uses only English in class, whereas the other uses Swedish, when needed. The view of the role of language constitutes an important difference as to the basis for the intended learning outcome, as well as implementation, of CLIL. This impacts assessment, as seen in one of the examples where one of the teachers used writing prompts allowing Swedish to be used. His CLIL colleague never allows Swedish, which is the policy of the entire international school context where he is working. As noted previously, translanguaging and interlanguages are often considered positive in the classroom (Creese & Blackledge, 2010; García, 2012; Olander & Ingerman, 2011), but at the same time it may cause stress for students when the language of instruction and the language of assessment are not aligned (Gablasova, 2014).

None of the history teachers claim to include language in their assessment, only content. To quote one of the CLIL teachers:

English proficiency cannot be part of an examination as little as Swedish should be part of assessment in history.

Yet one of the CLIL teachers states that the correct use of terminology is part of the intended learning outcomes. He has told students that they need to accept the possibility of attaining a lower grade due to lack of linguistic proficiency. This reveals a certain discrepancy between the ILOs and what is targeted in the assessment, or at least what is claimed to be.

In the question tests, some scaffolding can be identified in the tests used by to one of the CLIL teachers. The cognitive complexity level of the test items appears to be reduced, giving rise to modified items, such as sentence starters as in one of the tests. As Siegel et al (2014:683) state, “With scaffolding learners can achieve more advanced skills and reach levels of advanced cognition”. However, since no retrospective interviews were performed, it is difficult to know if the design of the test items was due to the language or if it would appear in a test in Swedish by the same teacher as well.

ANALYSIS AND DISCUSSION

To conclude the discussion on the assessment in history, there seem to be some minor modifications in the CLIL assessment design due to teachers' perceptions of students' proficiency level and the perceived difficulty involved in the use of English. Certain test items seem to be at a lower cognitive level and students are given the chance to make clarifications after the test. However, most of the differences seem to relate to teachers' varying individual preferences and views on the discipline.

One of the CLIL teachers believes that oral assessment is unfair to students who do not dare to speak English; assessment would only include what they "dare to say". He prefers the structured academic written genre and believes essays help students to show their analytical skills which ordinary question tests fail to do due to lack of time to process. However, he thinks question tests are easier to grade. A similar situation exists among the CLIL teachers: they favor written production, but differ in preferred type of genre. The non-CLIL history teacher mostly uses question tests, but favors an academic written genre. He still argues that the disciplinary genre is narrative, in common with both of the CLIL teachers, advocating both narrative and academic genres. The national course goals in history (NAE, 2012a) express that students should develop "the ability to use different historical theories and concepts to formulate, investigate [and] explain", thus pointing to competences requiring some sort of constructed response or written production.

In biology, the assessment samples rendered by the CLIL and the non-CLIL teacher at the same school are almost identical. The course layout and the writing assignments are comparable as well. As reported in TIMSS (2011), constructed response is a common test item, as in the teachers' tests in the present study. According to the literature, assessment in the biology discipline rests on cognitive processes and hierarchical performance expectations (Corrigan et al, 2013; Airey, 2012). Both disciplinary knowledge and relevant abilities need to be assessed.

According to the teachers in the present study, the abilities include problem solving, working methods, drawing conclusions and forming hypotheses. The non-CLIL teacher goes more into detail regarding cognitive skills as part of the ILOs, whereas the CLIL teacher focuses more on communication and content. All three – cognitive skills, communication and content, together with cultural references – are advocated for CLIL (Coyle, 2010). The CLIL teacher uses a range of function words in the test items, but

ASSESSING LANGUAGE OR CONTENT?

does not express the same concern for disciplinary tradition. Language is perceived as an instrument to describe relationships. Both teachers, regardless of L1 or L2, recognize a need to provide students with wordlists of disciplinary concepts. The CLIL teacher specifically claims that activities and language input in class should be sufficient to prepare students for linguistic output and production. According to her, if teachers do it right, students should acquire the tools they need, which resembles the views expressed by many scholars of immersion and language bath, or Krashen's input hypothesis: exposure to language results in language acquisition (Morgan, 1999).

The non-CLIL teacher in particular claims to draw on the character of the discipline when designing the assessment tools. One of the CLIL teachers claims to prefer written forms of assessment, acknowledging the importance of students having the time to sit down and express themselves, as opposed to oral communication. This is an interesting comment, since Llinares et al (2012:244) state that "the role of writing as part of learning in CLIL contexts is, at present, largely unrecognized, with much more interest being shown in the development of oracy." Whittaker et al (2011) note that, within CLIL research, very little information is available when it comes to discipline-specific writing. As already noted, the CLIL teachers in the present study claim to prefer written assessment modes, yet cannot expect to receive much research-informed guidance. However, suggestions have been introduced, presented in section 7.4 below.

To conclude the comparison of the assessment practices in biology, it seems as if differences in assessment practices are based on different perceptions of students' needs. In CLIL more focus on language use, in non-CLIL on students' production of academic genre and cognitive skills. The CLIL teacher's focus on language is testimony of her awareness of the impact of the language of instruction, which in turn may prevent her from focusing as much on other features, such as academic genre. However, the test design is almost identical in CLIL and non-CLIL. According to the national course goals (NAE, 2012 a), the students are supposed to "use knowledge of biology to communicate and [...] examine and use information", an opportunity offered especially in the writing assignments.

A concluding remark should be made regarding the CLIL teachers in both disciplines. In the questionnaire, they acknowledged assessment to be difficult and some of them stated that they would like more training in the use of

assessment tools. Nevertheless, they were willing to share their assessment samples, and they also expressed enthusiasm for their CLIL endeavor.

7.3 Effects of CLIL in the English language courses

The second research question focuses on the influence the CLIL profile may exert on the English language course format in the CLIL schools:

- Are the assessment tools and the course content affected in the English language courses where English is used in subject content courses? If so, *how* are they affected?

Before examining the results, issues in the data collection procedure need to be addressed. During the interviews, all EFL teachers gave their consent to share samples of their written assessment tools, but only one out of six ended up doing so. Two of them left the study and three never rendered any material, or submitted student texts instead of tests, prompts or assignment descriptions. Standardized tests in the form of national tests are compulsory in English 5 and 6 to measure all four receptive and productive skills. One of the teachers, who used old national tests for summative assessment purposes, submitted those together with some other standardized tests. The EFL teachers' hesitation to share their assessment material was unexpected. However, one possible explanation may relate to the researcher's background as an EFL teacher, causing them to feel more vigilantly analyzed (cf. Dalton-Puffer, 2007). CLIL being implemented in the content courses, and not in the EFL courses, implies that the content teachers' practice is more experimental, in a sense. The EFL teachers' assessment-tools, on the other hand, are designed in "their" language of instruction. To judge from the interviews, they use tools they feel comfortable with. In the questionnaire, three out of four responding EFL teachers claimed grading to be *fairly easy*, whereas one stated *fairly difficult*. Only one of them expressed a need for more training, referring to alternative forms of assessment.

The literature acknowledges the complexity of language testing (Bachman & Palmer, 2012), and Shohamy (2008:xiv) notes that "language knowledge is a complex phenomenon, which no single procedure can be expected to capture". In the CEFR and national course goals (cf. Chapter 3), it is

ASSESSING LANGUAGE OR CONTENT?

suggested that relevant and authentic subject content should be included in the EFL courses. When describing language use in assessment, Bachman and Palmer (2010:41) claim that topical knowledge has to be considered, since this provides the information base that enable students to use language. Yet, in the CLIL schools, where there seem to be a natural cross-curricular context, this is not immediately embraced by the EFL teachers. Teachers refer to lack of time, both to plan interdisciplinary projects, but also in their courses, where they need to cover a certain course content. Considering the changing status and presence of English in the lives of young Swedes (e.g. Sylvén & Sundqvist, 2012) it may be fruitful to consider different formats in the EFL courses. The communicative shift (Hymes, 1971), language as a tool *and* object of learning, as well as students' high proficiency level in English, have all made language teaching more complex.

Oscarson and Apelgren (2010) found in their survey that language teachers use a mix of classroom observation, written assignments and tests for assessment purposes. Self-assessment and portfolio were among the least used. At the international school, one of the teachers focuses on written production for student portfolios. His colleague at the same school does not favor students' own writing but, for different reasons, prefers to "spread it all over". The character of the discipline, as well as the syllabus, requires attention to multiple complementing features (Shohamy, 2008), a possible source of stress and confusion as to how and what to assess.

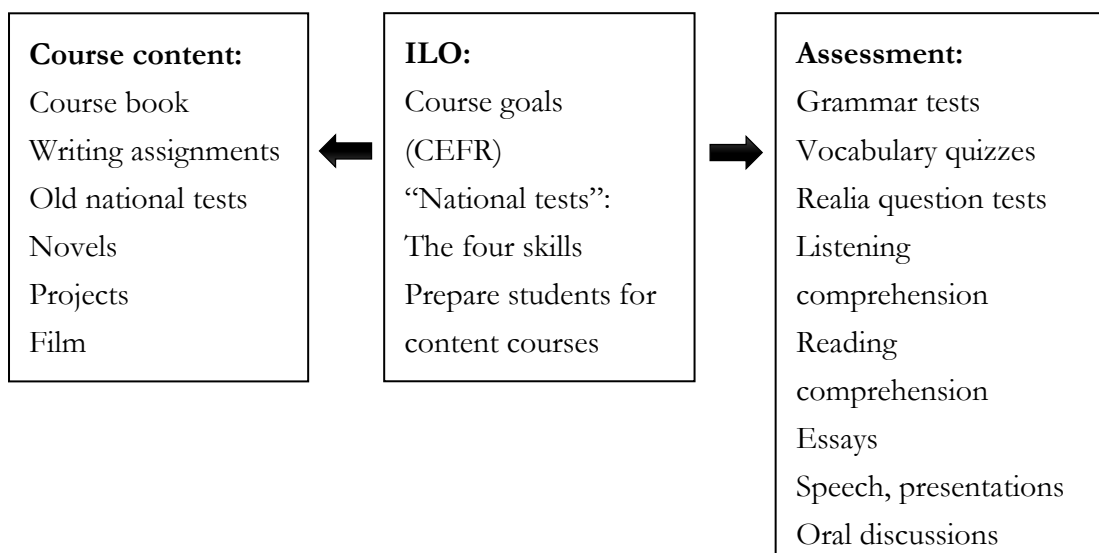
Returning to the second research question, the answer appears to be mainly "no": the content and assessment in the EFL courses are not significantly affected by the use of English in the other courses. At the international school, the EFL teachers acknowledge what they seem to perceive of as their responsibility, to prepare students, during their first few weeks in upper secondary school, for the English medium instruction in the content courses. One of the teachers has helped students go through material and difficult terminology from other classes. His EFL colleague has focused on spoken goals so that the students will feel comfortable using English in their content classes. The rest of the EFL course is characterized by assignments and content related to teachers' preferences and the syllabus, according to their own reports. None of the teachers acknowledge any integration, neither on their part, nor on the students', of disciplinary terminology from other courses in the language used in the classroom.

ANALYSIS AND DISCUSSION

In the CLIL literature, the success of CLIL as a teaching strategy is linked to the degree of teacher collaboration (Coyle, 1999; Haataja, 2013). In the present study, however, teaching methods involving collaboration across disciplines appear even less prominent in the participating schools than in contexts without bilingual teaching. Only the non-CLIL EFL teacher, who left the study after the interviews, claimed to work occasionally with a subject content teacher in history.

The result gives rise to a complementary question, namely why course content and assessment are *not* affected. By attempting to merge all the teachers of the study into one profile, as in the figure below, a very diverse and multi-faceted image of the EFL course format emerges. Intended learning outcomes, ILOs (Biggs, 2003), include many features, as seen both in the CEFR and the national course goals. Here the ILOs reported by the teachers are listed. The CEFR is clearly present in this connection, although not mentioned by any of the teachers. National tests are not learning goals, but the teachers speak about them as if they were. The diverse construct featured as ILO in the figure below can be said to reveal the occasionally “fuzzy” nature of the intended learning outcomes in the EFL course format.

Figure 11. Alignment of course content, Intended Learning Outcome and assessment in the EFL courses



The figure aims at combining the perceptions of the teachers, as expressed in the present study, regardless of school and course level (i.e. English 5, 6 or 7). This is to highlight the central role that the intended learning outcomes seem

to have concerning both course content and assessment; the ILOs define the assessment task, and teaching methods should be aligned with both ILOs and assessment. The question “Why are the English language course content and assessment tools *not* affected” may be answered, in part, by looking at the teachers’ learning goals, although merged, in the above figure. The teachers mention that they need to prepare students for the national tests, which cover the four skills: reading and listening comprehension, written and oral communication. The possibility suggested by the national syllabus, i.e. to find relevant content for purposeful language use in relation to student’s educational profile, does not appear to be a focal area. In some instances, the teachers express doubts as to their own ability to deal with subject content they are unfamiliar with. In the present study, the EFL teachers acknowledge very little interdisciplinary integration, mutual planning or sharing of thematic content. This is considered to be too time-consuming. Collaboration is not viewed as time-saving or an advantage, although a couple of teachers note that it would probably be beneficial for students if their workload could be diminished by merging assignments.

The results of the present study confirm what has previously been stated about CLIL, namely that it is usually implemented in content courses, not in language classrooms (Massler et al, 2014). If the aim of CLIL to make language learning more authentic and relevant is to be regarded as valid, EFL courses need to be included somehow. As Nikula (2005:55) argues, “there seem to be no principled reasons why certain aspects of CLIL instruction that seem to be conducive and meaningful foreign language use could not be brought into FL contexts as well”. A common framework for assessment in CLIL has here been advocated to strengthen a more coherent view.

7.4 Assessing language and content in the disciplines

The third research question focuses on possible similarities in the disciplines’ assessment procedures:

- What does the assessment design look like in the different disciplines when it comes to *language*, *content* and *form*? Are there common features?

ANALYSIS AND DISCUSSION

Before discussing the actual assessment material, an important finding concerns the teachers' views of the students. The EFL teachers, especially in the non-international schools, view the CLIL students as very proficient English users. The content teachers, on the other hand, seem to have a somewhat different perception, as they acknowledge the students' need of support and scaffolding in order to manage the extensive texts in the course books and all the new subject-specific vocabulary. The EFL teachers speak of the students as top students, easy to have in class since they are already so competent, whereas the content teachers think they need accommodation and support. The content teachers mention that they do not consider language, or refrain from correcting grammar mistakes since they themselves are not native speakers. This reveals a contradictory picture of the Swedish students' proficiency in English.

An explanation for the different views of the students, as stated above, has to do with the different registers and the difference in course focus. In EFL classes, the focus is often on BICS (basic interpersonal communicative skills) while in content courses the intended learning outcome is academic language, as in CALP (cognitive academic language proficiency) (Cummins, 2000). Further, the language in focus in content classes is disciplinary as in CO (content obligatory) language, rather than CC (content compatible) academic language. When students are expected to use subject-specific concepts in classroom discourse, these notions are supposed to integrate with their general language of schooling (Schleppegrel, 2004). However, students' communicative proficiency in English has been shaped in the language classrooms and by their extramural exposure (Olsson 2011; Sundqvist & Sylvén, 2012; Sylvén, 2006). The often neglected linguistic register is the general CC academic language needed in CLIL settings. It has to be dealt with either in the EFL course context or in the content courses.

After analyzing the data, it becomes clear that a certain discrepancy exists between the disciplines regarding what to identify as content. Course content in biology and history often relates to themes and topics. In the national syllabus, the core content in biology is defined under four headings: ecology, genetics, evolution and the nature of biology and its working methods. This is consistent with the thematic content of the course in the present study. Below each heading, smaller components are listed in four to ten bullet points describing, e.g., theories, models, structures and mechanisms. In history, core content is defined in five bullet points without any thematic headings. It can

ASSESSING LANGUAGE OR CONTENT?

be summarized by the European classification of time periods, the industrialization and democratization during the 19th and 20th centuries, historical source material, its interpretation and use, and finally “[h]ow individuals and groups have used history in connection with current conflicts and attempts to cooperate” (cf. NAE, 2012a, syllabus of history). The difference between the disciplines has also been recognized by researchers such as Martin (1993:213, in Llinares et al, 2012):

However, researchers have found a difference between science and humanities subjects in the treatment of terminology: ‘For many students, abstraction probably forms more of a problem than technicality, since science teachers do teach concepts and terms that make up scientific discourse whereas [...] history teachers do not focus explicitly on nominalization...’

Llinares et al (2012) note that the language of science is recognized as part of subject knowledge, but this is not equally clear in history.

In the English syllabus, core content is described by the use of three headings: content of communication, reception and production, and interaction. Below each heading, three to seven bullet points are listed, covering subject areas related to the students’ education, e.g. literature, cultural conditions, the spread of English in the world, texts of different kinds and for different purposes, oral and written production (NAE, 2012a). The CEFR defines thematic areas and domains which should be covered in language teaching, such as, personal identification, free time and travel, areas which would be used in relation to BICS (Cummins, 2000). Further, four types of “knowledge” are distinguished; declarative knowledge, communicative skills, existential competence, i.e. attitude and motivation, and finally ability to learn. Competences are defined as linguistic, socio-linguistic and pragmatic.

The intention here is not to make a complete comparison of Swedish national course goals and the CEFR, but to illustrate how the notion of content may signal and comprise many intersecting features, themes and even competences or skills. The same quandary can be traced in the scoring rubrics and task descriptions in the present study. In a task description for book reports in the assessment samples of this study, content is described in terms of; a very brief summary of the plot, setting, and personal opinion about the book, thus indicating the layout and structure of the presentation. Further, other aspects to be discussed are noted, including, for instance, subject message, genre and language. The varying views of what constitutes content

ANALYSIS AND DISCUSSION

are important to identify and describe before aiming for interdisciplinary projects.

Looking at language instead, only one of the six subject content teachers interviewed has a degree in English, two semesters at university level. This could be compared with the four semesters required for an English language teacher at upper secondary school level, in order to be certified to teach. The subject content teachers, CLIL and non-CLIL, show individual differences in the ways they relate to language in their courses.

The results of the questionnaire reveal that only one of the respondents, an EFL teacher, mentions linguistic accuracy as a main concern in the assessment of students' skills. It is not surprising that non-language teachers chose other alternatives.

In the questionnaire the teachers refrain from choosing a general content-compatible language as an important goal in their teaching. The lack of such a language will make it difficult for students to express their content knowledge in cognitively demanding contexts. The teachers may also lack an understanding of what this general language stands for, as found by Yoxsimer Pausrud (2014). Subject-specific concepts and terminology on the other hand, are easier to see the importance of. They are also included in the course goals.

A framework for assessment in CLIL was mentioned earlier, highlighting the need to cover both language *and* content within the disciplines. Therefore, a CLIL framework for assessment necessitates an integrative approach between disciplines, where language registers are identified in relation to common target language use.

In the table below, some of the cross-disciplinary features from this study are compared to show how writing assignments, which appear in all of the disciplines, cover different cognitive levels and thinking skills. The inclusion of LOTS (lower order thinking skills) to the left, and HOTS (higher order thinking skills) to the right shows the range of cognitive complexity. In the material in the present study, no assignments requiring the highest order skill, creating new ideas, can be identified. One explanation may be found in the content courses included in the study, biology 1 and history 1b, taught during the students' first year of upper secondary school.

ASSESSING LANGUAGE OR CONTENT?

Table 16. Cognitive function words in writing assignments across disciplines.

	Function words in production tests/writing assignments Cognitive levels				
	<LOTS		HOTS>		
	Narrate/ Remember	Explain/ Understand	Use/ Apply	Compare/ Analyze	Generate new ideas/ Create
Biology CLIL		X	X	X	
Biology non-CLIL		X	X	X	
History CLIL	X	X	X	X	
History non- CLIL		X	X	X	
English 5	X	X	X	X	
English 6		X	X	X	
English 7		X	X	X	

The instructions in the writing assignments and essay questions in some of the question tests have been compared, revealing that the level of cognitive complexity is rather similar across disciplines. All of them require students to explain, apply concepts and theories, as well as to analyze and compare. No descriptions including the generation of new ideas have been identified in the assessment samples. The lowest order thinking skill, to narrate and recall information, only appears in the first English course, English 5, as well as in narrative essays in one of the CLIL history courses. This could be compared with the CEFR (2001:61), level B1: “Can write very brief reports [...] which pass on routine factual information”, and the prevalence of BICS in the first course. During the next course, English 6, more focus is placed on CALP. In history one of the CLIL teachers favors the narrative format, thus implying a use of both narrative and analytical skills. None of the English language teachers chose analytical skills as one of the most prominent features in assessment in the questionnaire. Yet, in the essays in the high-stakes national tests students are supposed to discuss and compare, or, in other words to analyze (cf. section 6.).

Regardless of discipline, the course syllabi include content to communicate and learn. In the present study a distinction is made between content and language, where the former term is used to denote subject themes. Referring to Bloom’s revised taxonomy as well as lower and higher order thinking skills, a progression in complexity can be identified. This, in turn, is closely

ANALYSIS AND DISCUSSION

integrated with a student's language competence. For assessment purposes, one or the other, content or language may be in focus, or at least be said to be the target of assessment. By using Coyle's (1999) version of Cummins' quadrant (cf. section 4.5.2), the linguistic and cognitive complexity of the tests could be described and combined. The more linguistically demanding test items, which also require higher order thinking skills, are found in CR test items or production tests. Test items using academic function words requiring analysis and argumentation are found in both biology and history. In the EFL courses they may be traced in essays in English 6 and 7.

The question after analyzing the material of the present study, is whether there are enough common features to fuse goals across disciplines in order to find common assessment formats and features. Writing assignments seem to provide such a tool. Llinares et al (2012:244) refer to research where students' struggle with a foreign language in writing has led to deeper processing of content:

Writing about content is, on the one hand, a way for students to find out what they know and don't know about what they have studied. It is also a way to develop and expand language resources in the foreign language.

In CLIL discourse, assessment tools are requested which are capable of measuring the special skills used and acquired in CLIL settings (Llinares et al 2012; Morgan 2006). A validity concern in assessment refers to whether teachers focus on what is perceived as most important or most enjoyable and easy to assess (cf. Porter 2004). In order to assess "special CLIL skills", these need to be identified, described and included in a CLIL curriculum. At present, no such curricula exist, at least not in the Swedish context. Therefore, it is not hard to realize the complex, not to say impossible, endeavor CLIL teachers are facing. One cannot be accountable and justify the intended uses of an assessment (Bachman and Palmer 2012) in relation to CLIL and language acquisition, when no such intentional learning goals exist.

Below follows a discussion of possible threats to the validity of assessment in a CLIL-context.

7.5 Validation of assessment in a CLIL context

In Chapter 3, a chain model was presented, offering a systemic approach to validation, where different threats to the validity of an assessment procedure are identified (Crooks et al, 1996:282):

The primary purpose of the model is to guide and assist the validation of assessment procedures, interpretations and consequences. [...] different purposes imply substantially different emphases in validation, because the relative risks associated with each of the links and with the specific threats vary greatly with different assessment purposes.

This section offers a discussion of the validity of the written assessment procedures in the present study, using the eight steps in Crooks et al's chain model. Consequences for students as well as the educational context in which they appear are considered. Assessment interpretations, as seen in the assessment format chosen by the participating teachers, are evaluated. To the same extent as some of the steps in the chain overlap, not all of the eight steps are equally relevant for the present study, as indicated in the above quotation. Crooks et al describe how the model should be adapted to fit the current situation, while acknowledging the importance of identifying the weakest link to ensure validity. The validation is not exhaustive, but offers a view of the complexity involved in the practices.

The first step in the chain model is called *administration*. Threats to the validity associated with this first link refers to students' performance being misinterpreted due to lack of language, which may interfere with students' ability to demonstrate their content knowledge. This may cause test anxiety. In order to avoid what Crooks et al call inappropriate assessment conditions, research suggests that students should be encouraged to translanguage to avoid test anxiety and bias due to poor administration (Gablasova, 2014; García, 2009; 2012). In the present study, some of the content teachers claim to accept students responding in Swedish or even mixing languages, whereas others state that students should use English only, since this is the language used by the teacher. One of the teachers mentions the use of dictionaries during tests, representing a possible threat to validity, not because of the language, but because of the extra time and skills needed to use the tool.

The second step relates to the *scoring* of the tests. As in the previous link, CLIL students' lack of linguistic skills may interfere with their analytical skills or content knowledge. If undue emphasis is placed on certain aspects in the

ANALYSIS AND DISCUSSION

scoring, such as poor spelling, students may not receive relevant credit for their knowledge and skills. Even when content teachers claim not to assess language but only content knowledge, certain test items are highly dependent on student's linguistic performance. In a bilingual teaching context, awareness is needed of how to balance too detailed scoring, covering many separate areas, with too holistic scoring, where no assessment and scoring of specific aspects is done. The use of question tests versus production tests may serve as an example where this awareness is needed.

Threats to the third link, *aggregation*, relate to the design of individual tests as well as the assessment procedures within an entire course. Overlapping with the threats associated with the balancing of test scores, aggregation has to do with inter-task correlations. The balancing refers to the diversity of topics during a course, as well as between test items in the same question test.

The validity of the interpretation of test scores is also threatened if there is too wide a range of topics to assess the target domain. This is not a threat specific to CLIL contexts, but still important when the teacher considers the purpose of an assessment in relation to course goals, target language and written genre in relation to discipline and a student's L2. Is the purpose to test factual knowledge, analytical skills or linguistic repertoire? The purpose should align with the choice of assessment format. The use of essays and portfolio in some cases, question tests in content courses and completion tests in English language courses need to be validated. Some have already questioned the validity of traditional assessment tools in CLIL contexts (Kiely, 2012; Morgan, 2006).

The use of portfolio and writing assignments represent a broad construct-centered assessment where reliability can be at risk. *Generalizability*, which is the next link, improves when scoring criteria can be made more similar, by using standardized criteria and scoring rubrics. Scoring rubrics can be either task-specific or generic (Crooks et al, 1996). In this study, some of the English language teachers, as well as one of the content teachers, found it difficult to formulate or show proof of criteria used in the assessment of writing assignments. In English, the NAE provides guidelines for the national tests, but language teachers in the present study were hesitant to assess written production including unfamiliar content areas, and did not always refer to the above-mentioned guidelines in other assessment tasks. The lack of task-specific criteria, as well as generic interdisciplinary common features applicable in bilingual teaching contexts, represents a threat to the validity of

ASSESSING LANGUAGE OR CONTENT?

the generalization and accuracy of students' scores, when writing genre texts in their L2.

Messick (1989) uses the expression "construct under-representation", indicating that the validity of assessment is undermined if too constrained item types are used, preventing extrapolation and conclusions for the entire target domain of a course. The validity of *extrapolation* is strongly dependent on content coverage and cognitive complexity in the assessed domain. The lack of content-compatible academic registers in the subject content courses in the present study may prevent students' ability to attain the targeted proficiency levels in academic writing. The lack of content-obligatory registers in their L1 and content-compatible language in their L2 may restrict cognitive complexity and thus prevent extrapolation. Teachers may consciously or subconsciously try to lower the level of difficulty in the choice of test items, as may be inferred from some of the CLIL tests when compared with non-CLIL equivalents, making modifications by providing the initial letters of the answers, or allowing students to make clarifications after the test.

The validity of the *evaluation* relies on teachers' perception of course goals and students' proficiency level. The choice of words is an important factor here, indicating what the construct is, e.g. performance or ability. A CLIL teacher who states that "I don't assess language; that is up to the English language teachers", how can he/she differentiate between a mere passing grade and an excellent performance? Since the same objectives are used for CLIL as non-CLIL, can there be inter-rater consistency in the evaluation of scores regardless of language? The CLIL teacher may end up making allowances for students' poor quality of the answer, explained by poor written expression due to the use of an L2. In the present study one of the CLIL teachers mentioned that he had to approach students and ask them about the meaning if there was ambiguity in the answer before deciding on a grade. This, of course, may be considered relevant in a context where the effect of the L2 on the quality of the answer in content courses is an acknowledged concern. At the same time, however, this may constitute a threat to validity. In the end, it is a pedagogical decision determining the impact of the assessment, which according to Crooks et al (1996) directly influences the assessment's validity. The *decision* is informed by the standards the teacher uses, the explicit standards as found in curricular goals, or possibly informal ones, in the mind of the assessor. In the present study, all the participants refer to the use of the same national course goals when planning and assessing their courses,

ANALYSIS AND DISCUSSION

especially when asked about possible effects of CLIL. Yet, the CLIL teachers do realize the issues related to the language of instruction, showing, both in the interviews and in the assessment samples, that this is somehow taken into account (cf. section 7.2).

The last link in the chain model deals with the *impact* of an assessment, also expressed as consequential validity, referring to Messick (1989) once again. For students, this sums up all the previous steps and is possibly the most important link (Crooks et al 1996:280):

An essential part of the validation of an assessment process is an examination of the extent to which the assessment achieves the purposes for which it was intended, and the extent to which both intended and unintended effects of the assessment are positive or negative for the participants.

Thus stating that there will be both intended and unintended effects of the assessment, the important thing is how the effects are made manifest. Examples of positive and negative consequences are listed by Crooks et al (1996:279): enhanced motivation and greater confidence in skills and future performance on the positive side; reduced motivation, increased anxiety, focus on factual learning at the expense of higher cognitive level outcomes on the negative. The examples given depend on what the teacher focuses on in the grading and the feedback provided. The findings in the present study do not include students; therefore the validity of this link cannot be fairly estimated.

7.6 Summary

In this chapter the results in relation to the research questions have been analyzed and discussed. The findings reveal the absence of a specific CLIL method for assessment, but also the absence of subject-integration in the participating schools. The pedagogical purposes for implementing CLIL are unclear. If the language of instruction is not taken into account in assessment, what then is the intended learning outcome, other than subject content?

Crooks et al (1996) note that threats to the validity of assessment use vary greatly, depending on situation and different assessment purposes. Conversely, construct validity depends on clarity in the purpose and intentions of an assessment. The question is what the strongest, not the weakest, link is when determining the purpose of assessment in CLIL situations. The

ASSESSING LANGUAGE OR CONTENT?

language in CLIL cannot be disregarded as a bonus or side effect since it will have implications for student outcomes.

Nikula (2007) notes that CLIL and EFL teaching should be seen as complementing each other; CLIL situates students as language users rather than language learners. In an expanded view of validity (Messick, 1989), the relevance and use of assessment instruments are in focus. Brown and Hudson (2002) mention the needs of the students when aligning testing with curriculum. In the present study, the choice of assessment format is influenced by traditions inherent to the disciplines and the particular school context. In the next and final chapter, an effort is made to identify assessment features which may contribute to the needs of the students and of bilingual teaching approaches such as CLIL, integrating the use of content and language in the same tasks.

8. CONCLUDING REMARKS

The present study represents a limited contribution to the allegedly problematic and underexplored field regarding CLIL and assessment. However, there is now an increasing body of research and several studies have recently set out to investigate key areas of interest and concern regarding assessment in CLIL, especially the role of language in assessment and different modes of assessment in bilingual content instruction (e.g. Gablasova, 2014; Hönig, 2009; Wewer, 2014).

The issues related to language and content integration in assessment also find intersecting areas of interest with the teaching situation in bilingual immigrant education. Issues related to language in all content courses have been highlighted as a focal question and are becoming increasingly significant in Sweden as in other parts of the world (Hönig, 2009:3; Liberg, 2009; Lindberg, 2011; NAE, 2012b). In Sweden, close to twenty percent of the students in elementary school have another first language than Swedish, which means that subject content teachers face a linguistic diversity in the classrooms they are not always prepared for (NAE, 2012b).

Previous research on CLIL has concluded that there is a need for teacher training, a shared framework for good practice and a raised awareness of the role, functions and forms of different academic linguistic registers (Edlund, 2011; Gablasova, 2014; Hönig, 2009; Morgan, 2006; Wewer, 2014; Yoxsimer Paulsrud, 2014). The present study agrees with previous findings in all these regards. Teachers seem to recognize the need for subject-specific language, including terms and concepts, while at the same time they are unaware of the need to consciously develop an interdisciplinary academic register, sometimes referred to as content-compatible language (CC). In an assessment situation, the lack of such language may have considerable implications for students' ability to express content knowledge and to show proof of higher order thinking skills in cognitively demanding tasks. Without enough such language, they may be prevented from expressing a higher level of understanding of the course content, which in turn will have consequences for the outcome of assessment. This is specifically true in written assessment types requiring a constructed response, and even more so in essays. In genre pedagogy, it is

acknowledged that established patterns often form the basis of any variations (Hyland, 2007).

The written genre consists of typical, interdisciplinary academic features along with subject-specific concepts, both of which can be taught. The question in CLIL seems to be by whom: the language teacher or the subject content teacher, or both? At the same time, as with any model, there is a risk that teachers focus on genre approaches as the only way, instead of using other complementary instruments as well (Liberg, 2009).

In order to provide the best conceivable conditions for integration of language and content in bilingual teaching, and to be able to cater for validity in assessment in both domains, assessment guidelines need to be in place. In the next section, the pedagogical implications of the findings in the present study are discussed and a possible way forward by distinguishing cross-curricular writing features is suggested.

8.1 Contributions and implications

A study such as the present, which sets out to find common features in the assessment practices in CLIL, obviously has pedagogical implications: for the validity of CLIL as a teaching method, but also for its practitioners to be considered assessment literate, standards and guidelines for good practice need to be developed. First of all, assessment practices in bilingual teaching should differ from those of regular teaching in L1. A specific approach for assessment in CLIL, where standards how to deal with language are articulated needs to be in place. Subject content teachers, who take on a tremendous workload in developing course and assessment material in English, need training to recognize the different registers involved when performing subject-specific tasks. This includes identifying different academic registers comprising CO (content obligatory) and CC (content compatible) language along with written genre (Llinares et al, 2012; Schleppegrell, 2004) and TLU (target language use) domains (Bachman & Palmer, 2012).

Secondly, subject-content teachers need an awareness of the cognitive and linguistic demands of different academic function and question words in order to appreciate the required skills of the students, including language, when processing different test items. The appropriateness of different types and modes of assessment has to be considered and validated.

CONCLUDING REMARKS

In CLIL schools, the integration and interdisciplinary collaboration between subject content courses and English language courses should be a top priority. EFL teachers have an opportunity to access authentic content to make their courses more relevant, and, conversely, subject-content teachers can focus more on cognition and the relevant topic. However, to quote Sylvén (2004:227):

Is it then, necessary to use CLIL to enhance the teaching of English? Considering the lack of dually qualified teachers and the exposure to English available in so many areas in Sweden already, how justifiable is it to use it as the language of instruction in school?

One response to this question may be that, by developing a genre based written pedagogy and reading strategies in upper secondary school this may contribute to bridging the gap to the academic world at large. Although, preparing students for tertiary level is not the aim in all CLIL contexts. Other genres may be relevant in vocational programs where CLIL is used. However, teachers' genre awareness and the pursuit of shared cross-disciplinary academic registers should be the overarching aim in an integrative teaching approach offered through CLIL. The purpose of CLIL is to fuse goals and work across disciplines and curricula. If genre awareness and general linguistic skills can be taught regardless of discipline, making connections and addressing the fragmentation in the educational context, CLIL could help to make education more relevant and homogeneous to students.

Another possible perspective would imply using CLIL in other FL contexts. For instance, to find content teachers willing to teach in French and enough students in one school prepared to learn subject matter in any other language but English does not seem likely in a Swedish context. However, if CLIL were to be implemented in the FL courses this would be different. The use of any other language but English implies a different construct. Even though the national course goals are the same, the proficiency level at the outset at upper secondary school differs, as does course design for the same reasons. To respond to Sylvén's question, maybe CLIL would be more valuable using an L3 or L4 as the medium of instruction? This would better align with the goal of the European Commission (1996), to enhance the learning of more foreign languages.

Either way, a future assessment framework for CLIL will necessitate more awareness of the role of languages within and across disciplines. Language and

ASSESSING LANGUAGE OR CONTENT?

cognition are inseparable and, therefore, both of them should be acknowledged in assessment practices regardless of discipline. In the same way, language and content are two sides of the coin and cannot be assessed separately (Mohan, 1986). It is a matter of defining and deciding what content means in each case, and what language is needed and used to perform various written tasks, applicable to oral communication as well.

Following Cummins' CALP, the academic language for each discipline that is used in assessment should also be taught. The target language use, TLU, for an assessment needs to be aligned with assignments used in class. Hyland 2007:149 states:

By making explicit what is to be learnt, providing a coherent framework for studying both language and contexts, ensuring that course objectives are derived from students' needs, and creating the resources for students to understand and challenge valued discourses, genre approaches provide an effective writing pedagogy.

In the present Swedish context, the first place to look for objectives is curricular goals and syllabi. Linguistic and cognitive skills are stated for each discipline. While some of them are discipline-specific, many are indeed cross-disciplinary. By merging similar goals and applying genre-based writing instruction across disciplines, a specific assessment framework and pedagogy may be established, in line with students' needs. The same could apply to constructed response test items.

Gajo (2007) notes that lists of content-obligatory and content-compatible language should be established, thus promoting collaboration between language and subject teachers. Linguistic and cognitive goals have already been merged in Cummins' matrix (cf. section 4.5.2). Furthermore, within CLIL other suggestions have been put forward to combine features across disciplines, e.g. Coyle's four Cs (Cambridge ESOL, 2008), analytic assessment grids (AECLIL, 2012), grammatical features in the subjects (Llinares et al, 2012; Schleppegrell, 2004), to mention a few. Previous research claims that portfolio seems like a good assessment option in CLIL contexts (Wewer, 2014). Without specifically advocating portfolio, the findings in the present study point to common practice already in place in Swedish CLIL classrooms, where writing assignments provide a common cross-curricular denominator.

Building on the findings in the present study, a few preliminary guidelines are suggested below. These align with suggestions in genre pedagogy

CONCLUDING REMARKS

(Gibbons 2003; Hyland 2007; Schleppegrell & O'Hallaron, 2011) as well as previous research in CLIL (Gajo 2007; Llinares et al, 2012; Wewer, 2014):

- Identify and define relevant written genres for the relevant context and discipline (e.g. narrative, informative, expository, argumentative, analytical texts)
- Exemplify by using subject specific types and formats (e.g. novel, article, letter, project, laboratory reports)
- List content-obligatory language and concepts
- List content-compatible language and expressions (e.g. linking words and expressions to create coherence)
- List grammatical features needed
- Define language needed for higher order thinking skills
- Incorporate CEFR in the process
- Set up transparent scoring rubrics, involve colleagues and students in the process

As seen in the validation of assessment practices, reliability can be at risk when using portfolio and writing assignments. One way of improving generalizability and validity is to find generic as well as task-specific rubrics and benchmarks. When more research has been conducted in the field in various disciplines, a possible framework may be launched. Until then, some suggestions for future research are presented below.

8.2 Suggestions for future research

The focus in this study has been on teachers and their assessment practices in a Swedish upper secondary context. There are many aspects of assessment in CLIL which have not been addressed. Three disciplines were included in this study: English (EFL), biology and history. The results of the present study would benefit from more in-depth research within the disciplines as well as a inclusion of other disciplines. By comparing the findings in this study with teachers' assessment practices in other contexts, e. g. other disciplines and other schools at the same level, elementary school or tertiary education, a deeper understanding would be possible.

In a further study involving teachers, other methods could be used, such as retrospective interviews and stimulated recall, where teachers reflect on

ASSESSING LANGUAGE OR CONTENT?

student achievements in the tests, and also their own grading, possibly making a validation of their own tests.

Oral assessment and formative assessment are often in focus in the discourse surrounding CLIL. However, the present study was particularly concerned with written assessment. More research involving other types and modes of assessment in a Swedish educational context would add to the overall picture. Comparisons with assessment formats in other international CLIL contexts would also be rewarding in order to find a basis for good assessment practices in bilingual education or CLIL.

The present study has been limited to interviews and analyses of documents. Future research could include classroom studies to better understand context and the alignment between course content, classroom practice and assessment procedures. Furthermore, this study was limited to teachers' perspectives, but shifting the focus to students would also be worthwhile. One of the teachers in the present study suggested interviews with former CLIL students who are now university students, to find out how well they succeed in their academic writing and overall achievements. Student interviews paired with an analysis of their performance could serve as method.

A deepening of the analysis of genre and written performance would be beneficial for CLIL and bilingual teaching, but also for validation of written assessment in the various disciplines. Whittaker et al (2011) note that there has not been much work published on the written production of CLIL students, and as to discipline-specific writing, even less information is available. Comparing CLIL schools with IB programs or regular schools that work with subject integration could also add valuable information regarding assessment design and assessment issues.

For future research, it might be of interest to compare the results of the present questionnaire to a larger sample. Some of the questions are similar to those in a survey made by Oscarson and Apelgren in 2010 among language teachers; thus a partial comparison might be possible in the future.

In conclusion, teachers possess a great deal of experience and valuable insights which are not always taken into consideration. Offering research-based teacher training and support in how to design relevant and reliable assessment assignments, taking both language and content into account, could promote awareness that there are obvious gains to be made in more interdisciplinary collaboration. There are indeed shared features among

CONCLUDING REMARKS

academic genres and languages used for similar purposes in and across disciplines.

SWEDISH SUMMARY

Inledning

Content and Language Integrated Learning, CLIL, är en tämligen utbredd undervisningsform där ett annat språk än elevernas, och ofta även lärarnas, första språk används som undervisningsspråk i ämnen som, exempelvis, biologi och historia¹⁰. Denna studie utforskar och jämför 12 svenska gymnasielärares bedömarpraktik i ämnena engelska, biologi och historia i tre olika sådana CLIL-kontexter där språk- och ämne integreras. Upphovet till studien återfinns i en önskan att förstå bakgrunden till lärares användande av olika bedömningsinstrument i engelska. I kontakten med CLIL-kontexten kom studien att inkludera bedömningsformer inte enbart *i* engelska, utan även *på* engelska, där engelska är undervisningsspråk i andra ämnen.

Bedömning inom CLIL har lyfts fram som ett i hög grad utforskat område, det har t o m beskrivits som en blind fläck (Massler et al, 2014), även om några studier har bidragit på senare tid (t ex Hönig, 2009; Wewer, 2014). Ett problem som har uppmärksamats i samband med CLIL är att användandet av ett främmande undervisningsspråk ibland befaras ha en hämmande effekt på elevers uttrycksförmåga när de ska redovisa sina ämneskunskaper (AECLIL, 2012; Kiely, 2010; Morgan, 2006). Därutöver saknas en gemensam pedagogik och riktlinjer för hur ämnesintegreringen ska gå till, inte minst i samband med bedömning (Sylvén, 2013). Bedömning och betygsättning är ofta förenat med viss vanda från lärares sida, inte minst utifrån kravet på validitet och reliabilitet. I ett CLIL-sammanhang, där det saknas en gemensam och medveten undervisningsstrategi som omfattar både ämnesinnehåll och språk (Socrates-Comenius, 2009), ställs validiteten i bedömningspraktiken inför ytterligare utmaningar, vilket diskuteras i föreliggande studie. Frågor som står i fokus berör *vad* som bedöms i förhållande till innehåll och språk i de olika ämnena, samt *hur* detta bedöms.

¹⁰ Språk- och ämnesintegrerad undervisning går i vissa svenska sammanhang under beteckningen SPRINT (cf. Nixon, 2000).

Bakgrund

CLIL och ämnesintegrerad undervisning

Under senare år har det blivit allt vanligare att gymnasieskolor i Sverige erbjuder en internationell profil där undervisningsspråket är engelska. Ett av syftena som uttrycks är att förbereda eleverna för en global värld (Lim Falk, 2008). Vid fortsatta studier på högskola i Sverige är kurslitteratur och undervisningsspråket många gånger på engelska (Costa, 2009; Maiworm & Wächter, 2008). I en undersökning från 1999 (Nixon, 2000), bedöms 23 % av svenska gymnasieskolor ha någon form av CLIL. Yoxsimer Paulsrud (2014) gjorde en ny undersökning, och bedömer att antalet inte har ökat sedan dess, såvida det inte definieras som delvis eller tillfällig CLIL. Ett mål som uttryckts för CLIL internationellt är att ge språkundervisningen ett uppsving med autentisk och förhoppningsvis motiverande innehåll (Dalton-Puffer, 2007). Europakommissionen (1996) har framhållit CLIL som en metod värd att främja för att uppnå målet med tillägnet av ett tredje främmande språk i Europeiska skolor. I Sverige har dock användandet av CLIL framförallt förekommit med engelska som undervisningsspråk. För att uppnå det önskade flerspråkiga målet skulle CLIL snarare behöva användas i samband med andra främmande språk.

CLIL betraktas som en samlande term för det som ofta beskrivs som en mycket skiftande och heterogen undervisningspraktik (Dalton-Puffer, 2008; Mehisto, Marsh & Frigols, 2008). Internationellt används även termer som English Medium Instruction, EMI, eller Content Based Language Teaching, CBLT, Content Based Instruction, CBI, samt immersion för att nämna några av de vanligast förekommande.

Validitet i bedömning

Lärares bedömning har konsekvenser för individer såväl som för samhälle. Med detta följer att lärare förväntas kunna motivera och beskriva bakgrunden till dragna slutsatser rörande elevers kunskapsställning och förmåga (Bachman & Palmer, 2012). Messick (1989) beskriver att validitet består i argument som stödjer riktigheten och lämpligheten i dragna slutsatser utifrån provresultat eller annan bedömning. Därmed behöver provuppgifter mäta vad de är avsedda att mäta; inte irrelevanta aspekter, vilket Messick kallar *construct-irrelevant variance*, och inte heller för lite av vad som behöver mätas, av Messick kallat *construct under-representation*. I den gemensamma referensramen för språk,

SWEDISH SUMMARY

GERS, beskrivs validitet som ett mått på hur väl de bedömda kvaliteterna motsvarar vad uppgifterna avser att bedöma.

I föreliggande studie vilar det konceptuella ramverket på Messicks vidgade validitetsbegrepp (Bachman, 2005; Messick, 1989; 1996), vilket omfattar aspekter förknippade med relevans och användbarhet (McNamara, 2009; Shepard, 1993). I en CLIL-kontext är det viktigt att identifiera eventuella hot mot validiteteten som kan härledas till oklarheter i tillämpandet av en undervisningsmodell som medför andra krav på elevernas förmåga. Detta blir extra viktigt då otydlighet råder kring språkets roll och eventuella inverkan, som beskrivits ovan. Materialet i denna studie analyseras således med hjälp av en valideringsmodell framtagen och bearbetad av Crooks, Kane och Cohen (1996), med syfte att identifiera hot mot validiteten vid bedömning. Analysprocessen med hjälp av modellen beskrivs kortfattat under metodavsnittet nedan.

Bedöma språk eller innehåll?

CLIL är en undervisningsmodell där själva termen beskriver en i grunden oskiljaktig förening mellan innehåll och språk (Coyle, Hood & Marsh, 2010). Samtidigt har det konstaterats att integreringen av de två inte är komplikationsfri (Gajo, 2007), vilket kan förklaras av konkurrensförhållanden (Dalton-Puffer, 2007). Termen *innehåll* kan vara svår att definiera, ofta står den i relation till ämnes- och kursinnehåll, vilket t ex avses i rubriken ovan. Sålunda anser språklärare i föreliggande studie att de inte kan bedöma ämnesinnehåll, liksom flera av ämneslärarna, både i denna studie och i tidigare (Hönig, 2009), anser att de inte kan eller bör bedöma elevers språkliga förmåga. I en nationell svensk diskurs har det betonats att alla lärare är språklärare. Från Skolverkets sida har material publicerats som hävdar att språk och ämneskunskap hänger tätt ihop, samt att elevers språkkunskaper kan utvecklas i alla ämnen (Skolverket, 2012b).

För att sträcka sig bortom ämnesinnehåll och teman i relation till olika kurser, så kan innehåll inbegripa flera aspekter. Coyle, Hood och Marsh (2010) beskriver att innehåll kan inkludera såväl kunskap som förmågor och förståelse som vi önskar att elever ska omfatta. Vid en studie av ämnesplaner och angivna kunskapskrav i olika discipliner, framträder både ämnesspecifika innehållsliga mål och sådana som är ämnesöverskridande. På motsvarande sätt inkluderar elevers tillägnande av ämneskunskap och språk att de behöver använda sig av både kognitiva och språkliga förmågor av skiftande

komplexitet när de ska processa språk och innehåll. I samband med bedömning gäller även att eleverna har försetts med rätt språkliga verktyg, omfattande både ett ämnesspecifikt ordförråd och ämnesöverskridande skolrelaterade genrer (Llinares et al, 2012; Olander & Ingerman, 2011; Schleppegrell, 2004; Schleppegrell & O'Hallaron, 2011; Snow, Met & Genese, 1989) så eleverna är utrustade för de uppgifter de förmodas klara av. Biggs (2001, 2003) modell av "constructive alignment", beskriver hur effektiv undervisning är ett resultat av att undervisning och metod är samordnade med de tänkta lärandemålen, d v s målen bör finnas med från början. Detta blir på nytt problematiskt i en CLIL-kontext där målet med undervisningsspråket inte är tydligt definierat.

Syfte och frågeställningar

Syftet med föreliggande studie är att undersöka *om* och i så fall *hur* bedömningspraktiken och bedömningsdesignen skiljer sig åt, eller inte, beroende på om undervisningsspråket är svenska eller engelska, i det sistnämnda fallet så kallad CLIL. Frågeställningen inriktar sig på hur bedömningsformerna i dessa två sammanhang, CLIL och icke-CLIL, tar sig uttryck i biologi och historia: Finns det skillnader i bedömningsverktygens utformning beroende på vilket undervisningsspråk som har använts, och vad baseras dessa skillnader i så fall på? En del av studien ägnar sig även åt språkundervisningen i engelska på berörda skolor, för att se om bedömningsdesignen och kursinnehållet där påverkas av det faktum att andra kurser studeras på engelska. Utifrån den ämnesöverskridande och ämnesintegrerade karaktären i CLIL är ett mål även att jämföra några gemensamma drag mellan ämnena, för att om möjligt identifiera beröringspunkter. De tre specifika forskningsfrågorna är som följer:

- Finns det skillnader i bedömningspraktik beroende på om undervisningsspråket är engelska eller svenska, CLIL respektive icke-CLIL, i biologi och historia på gymnasiet? I så fall, *hur* skiljer den sig i så fall åt och *på vilka grunder*?
- Påverkas innehåll och bedömningsformer i engelskkursen i de fall där andra ämnen undervisas på engelska? I så fall, *hur* påverkas de?
- Hur ser bedömningsformerna ut i de olika disciplinerna med avseende på *språk, form* och *innehåll*? Finns det gemensamma drag?

Material och metod

Studien fokuserar lärares arbete och omfattar tolv lärare, varav tre undervisar i biologi, tre i historia och resten i engelska vid tre gymnasieskolor i en stor samt två mellanstora svenska städer. En av skolorna är helt engelskspråkig, men med svensk läroplan, de andra två erbjuder ett respektive två gymnasieprogram med internationell profil där undervisningsspråket i övervägande fall är engelska. Denna studie ingår i ett större projekt, finanserat av vetenskapsrådet, CLISS-projektet, Content and Language Integration in Swedish Schools (för en närmare beskrivning, se Sylvén & Ohlander, 2014). Därmed är skolorna i denna undersökning de samma som för projektet som helhet.

Materialet som insamlades under 2013-14, består av halvstrukturerade intervjuer, en enkät samt bedömningsmaterial i form av skriftliga provexempel. Intervjuerna omfattar totalt 8 timmar, och varierar från 16 till 58 minuter i längd. En intervjuguide användes med följande teman: lärares erfarenheter av CLIL, ämnessyn, syn på bedömning samt använda bedömningsinstrument, kursmaterial, kursplan och förekomsten av ämnesövergripande samarbete. Två av de deltagande engelsklärarna valde att lämna studien efter intervjuerna. En annan begränsning var att enbart en av de kvarvarande engelsklärarna valde att bidra med bedömningsexempel. En av historielärarna (CLIL) som meddelat att hon avsåg bidra med samtliga provexempel från en kurs, lämnade till slut enbart två av fyra prov. Övriga ämneslärare lämnade in samtliga bedömningsexempel.

Intervjuer liksom enkäter analyserades tematiskt med utgångspunkt i svenska kursmål och med hjälp av intervjuguiden. De skriftliga proven analyserades utifrån faktorer såsom frågetyp, t ex flervalsfrågor, matchningsfrågor, kortsvarsfrågor eller uppsatsfrågor (Brown & Hudson, 2002; Levin & Marton, 1973; Wedman, 1988; Wikström, 2013) samt vilka krav proven ställer på elevers förväntade kognitiva och språkliga förmåga. För att bedöma det sistnämnda användes beskrivningar som återfinns både i CLIL-litteratur samt studier av bedömning i CLIL (Hönig, 2009; Wewer, 2014), samt i tidigare studier av bedömningsmaterial i andra ämnen (Lindmark, 2013; Odenstad, 2010; Rosenlund, 2011). De verktyg som ansågs användbara består i Blooms' reviderade taxonomi (Anderson & Krathwool, 2001) och så kallade akademiska funktionsord i provfrågor som ställer olika krav på kognitiv förmåga, t ex *ange*, *beskriv*, *analysera*, eller frågeord, t ex *vad*, *hur*. Likaså användes en tillämpning av Cummins' kvadrant som kombinerar kognitiv

komplexitet med språklig komplexitet (Coyle, 1999). Eftersom det inte finns några tidigare gjorda studier av det här slaget inom CLIL, var det till att börja med svårt att hitta beskrivningsvariabler eller en lämplig metod för att kunna enhetligt beskriva och jämföra provuppgifter från skilda ämnesdiscipliner. I arbetet med Cummins kvadrant och Blooms taxonomi, liksom granskningen av ovan nämnda litteratur och tidigare gjorda studier, framkom flera gemensamma jämförbara kriterier, och en egen modell uppstod.

En valideringsmodell användes för att identifiera möjliga hot mot validiteten i bedömning i en CLIL kontext. Crooks, Kane och Cohens kedjemodell från 1996 ansågs tydlig och användbar för ändamålet. Den är utformad i åtta steg och styrkan i varje steg eller länk beror på om själva syftet med bedömningen är tydlig och lämpligheten i bedömningsuppgifterna i relation till dessa syften. Den andra länken består exempelvis i poängsättning och ett hot mot denna länk kan utgöras av om för mycket vikt läggs vid någon parameter, i CLIL kontext skulle ett sådant hot kunna utgöras av att för stor vikt fästs vid bristande stavning, då fokus i uppgiften är på historisk analysförmåga.

Resultat och diskussion

Resultaten i studien står i första hand i relation till de tre forskningsfrågorna. För det första förefaller det inte som att CLIL och användandet av engelska som undervisningsspråk har någon inverkan på lärares bedömningspraktik. De skillnader som kan spåras, framför allt i historia, verkar snarare bero på individuella preferenser eller ämnessyn. Just historia beskrivs av Pace (2011) som en luddig disciplin, syftande på att det inte funnits tydliga riktlinjer för bedömning. De två CLIL-lärarna föredrar uppsatser, men av olika karaktär. Den ena förordar en narrativ genre, då detta bäst anses avspegla den historiska ämneskaraktären, medan den andra föredrar en mer akademisk form, vilken anses vara mer tydlig och strukturerad. En kollega (icke-CLIL) föredrar frågeprov, men vissa av frågorna, framför allt mot slutet av kursen, har uppsatskaraktär. Det som är viktigt med avseende på validiteten i bedömningsformerna, är att syftet med uppgiften går att redogöra för samt att bedömningskriterierna är transparenta. Alderson och Banerjee (2002) hävdar att ju mer strukturerad en uppgift är, desto mer reliabel är bedömningen. I samband med användandet av uppsatser som bedömningsform anger CLIL-lärarna att de ser vissa problem med språkets negativa inverkan i vissa fall, vilket man löser med samtal med berörda elever. Ingen av CLIL-lärarna har

SWEDISH SUMMARY

någon färdig bedömningsmatris att presentera. En av dem tvärtom värjer sig mot detta och menar att bedömningskriterierna blir tydliga under processens gång i dialog med eleverna.

I biologi används både frågeprov och andra skrivuppgifter såsom laborationsrapporter och fältanteckningar. Två kollegor, en CLIL och en icke-CLIL, samarbetar kring utformningen och flera provfrågor och uppgifter är närmast identiska. En skillnad som framkommer i intervjuerna, är att CLIL-läraren fokuserar ämnesspecifik vokabulär, vilket delvis kan härledas till användandet av engelska, men även läraren i icke-CLIL förser eleverna med ordlistor över ämnesspecifika begrepp. CLIL-läraren medger att det generella akademiska utfyllnadsspråket inte ägnats någon särskild fokus, men påpekar dock att det skulle vara en intressant aspekt att ha i åtanke. Icke-CLIL kollegan däremot nämner behovet att tillgodose utvecklandet av ett mer generellt akademiskt språk. Sammanfattningsvis kan vissa modifieringar i bedömningspraktiken på grund av CLIL spåras; frekvensen av frågeord på en lägre kognitiv nivå kan förefalla något högre i dessa prov och eleverna har i vissa fall möjlighet att välja vilket språk de vill använda, ha tillgång till ordböcker eller få längre tid på sig vid prov. Dessutom finns möjligheten att göra förtydliganden i efterhand om språket tycks utgöra ett hinder. Alla CLIL-lärare ger uttryck för att det enda som bör skilja sig åt är språket, och hänvisar till att de ju lyder under samma nationella ämnesplaner och kursmål. I enkäten anger alla lärare utom en, att nationella kursmål är mycket viktiga vid betygssättning.

Svaret på den andra forskningsfrågan; om innehåll och bedömningsformer i engelskkurserna påverkas, är i huvudsak nej. I CLIL-litteratur beskrivs framgången i CLIL som undervisningsmetod vara avhängigt av graden av samarbete lärare emellan (Coyle, 1999; Haataja, 2013). Interdisciplinärt samarbete tycks dock mindre frekvent bland de deltagande lärarna i denna studie än i andra icke-CLIL kontexter. Endast en engelsklärare (icke-CLIL), som lämnade studien efter intervjuerna, angav att hon samarbetade med en historiekollega. Frågan som infinner sig är vad bristen på avspeglning av innehåll från ämneskurserna beror på. Enligt lärarna själva beror det på brist på tid att hinna med något annat än det kursinnehåll de är tvungna att klara av i relation till sina ämnesmål. Engelsklärarna fäster stor vikt vid de nationella proven. Dessa används som referens för vad som behöver ingå i kursen, och gamla prov används som bedömningsinstrument och för föreberedelse för de ”riktiga” proven. En annan faktor som hämmar ämnesintegrering beskrivs

som brist på tid för samarbete och samplanering. Slutsatsen är att CLIL verkar hålla sig inom ämneskurserna, vilket bekräftas av andra (Massler et al, 2014). Det som bör påpekas är dock att i avsaknaden av nationella riktlinjer för CLIL, saknas även beskrivningar för hur integrering och samarbete ska gå till.

Den sista forskningsfrågan berör gemensamma nämnare mellan ämnena avseende språk, innehåll och form i bedömnings-sammanhang. En skillnad som framkommer, till att börja med, berör lärarnas uppfattning av elevernas språkliga kompetens. Medan engelsklärarna ser CLIL-eleverna som avancerade språkinlärare, anser ämneslärarna att de måste stötta eleverna språkligt. Detta kan sägas återspegla att två olika språkliga register är i fokus i de två kontexterna, vilket kan jämföras med Cummins BICS, Basic Interpersonal Communicative Skills och CALP, Cognitive Academic Language Proficiency, där det första är mer frekvent i engelskklassrummet, och det sistnämnda i ämneskurserna. Analysen visar att det finns vissa skillnader mellan disciplinerna avseende vad som kan beskrivas som innehåll. En beröringspunkt relaterar till det gemensamma akademiska språket och användandet av skriftliga genrer. Det ämnesspecifika innehållet skulle dock kunna användas för att berika det språkliga uttrycket i engelskurserna. Begreppet innehåll, vilket tidigare konstaterats, visar sig vara kopplat till ämnets karaktär. I biologi anses ämnesbegreppen vara en del av innehållet (Llinares et al, 2012), men detta är inte lika tydligt i historia. I den engelska ämnesplanen framstår begreppet innehåll som mest mångfacetterat. Den skriftliga genren framstår som den samlande faktorn, där språk och innehåll möts, oberoende av hur dessa definieras. Nedan följer några sammanfattande förslag på pedagogiska implikationer för bedömning i en språk- och ämnesintegrerad CLIL-kontext.

Pedagogiska implikationer

Studien bekräftar den redan konstaterade avsaknaden av tydliga definierade riktlinjer för CLIL som undervisningsmetod (Sylvén, 2013; Yoxsimer Paulsrud, 2014). I likhet med tidigare forskning framträder ämneslärares osäkerhet hur språket ska hanteras vid bedömning. Ett syfte med CLIL skulle kunna vara det parallella utvecklandet av ämnesrelaterad och generell språklig kompetens. För validiteten och värdet i CLIL skulle ämneskurser och engelskurser behöva samarbeta för att bidra till främjandet av CALP, (cognitive academic language proficiency), och ett generellt akademiskt språk. Om syftet med CLIL är att samtidigt kunna testa faktakunskap, analytisk

SWEDISH SUMMARY

förmåga och språklig repertoar kan användandet och undervisningen av skriftliga genrer (Hyland, 2003; 2007) och portfolio för bedömning ses som en framkomlig väg. I punkterna nedan sammanfattas några möjliga riktlinjer för bedömning i CLIL-kontext:

- Identifiera, definiera och undervisa relevanta ämnesrelaterade skriftliga genrer. Exemplifiera med ”måltexter” och ”målspråk”
- Lyfta fram och undervisa både ämnesspecifika begrepp och generellt akademiskt språk
- Utarbeta och använda transparenta bedömningsmatriser som inkluderar både innehåll och språk, gärna i samarbete med kollegor.

REFERENCES

- AECIL-EACEA, (2012). *Assessment and evaluation in CLIL*. Ibis, Como – Pavia.
www.aecil.net
- Airey, J. (2012). I don't teach language. The linguistic attitudes of physics lecturers in Sweden. *AILA Review*. Integrating Content and Language in Higher Education, Vol.25 John Amsterdam/Philadelphia: Benjamin Publishing Company
- Airey, J. (2013). Plenary speaker at "Temadag om CLIL" in Stockholm, Sweden, March 2012.
- Airey, J. (2011). Talking about teaching in English: Swedish university lecturers' experiences of changing teaching language. *Ibérica* 22 (2011): 35-54.
- Alderson, J. C. & Banerjee, J. (2002). *Language testing and assessment*. (Part 2). I: Language teaching 35. S. 79-113.
- ALTE, (2011). *Language Test Development and Examining*. European Council: Language policy division
- Anderson, L. & Krathwool, D. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, Longman.
- Atkinson, P. & Coffey, A. (2011). Analyzing Documentary Realities. In Silverman, D. (2011). *Qualitative Research. Issues of Theory, Method and Practice*. 3rd ed. 77-92. Los Angeles, SAGE.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford [u.a.]: Oxford Univ. Press.
- Bachman, L. F. (2005). Building and Supporting a Case for Test Use. *Language Assessment Quarterly* 2(1), 1-34.
- Bachman, L. F. & Palmer, A. S.(2010). *Language assessment in Practice*. Oxford: Oxford University Press.
- Bailey, K. M., (1999). *Washback in language testing*. TOEFL Monograph Series. New Jersey: Educational Testing Service.
- Bernstein, M. (1999). Vertical and horizontal discourse: An essay. *British Journal of Sociology Education* 20:157-173.
- Biggs, J. (2001). On Constructive Alignment. Background notes to support a seminar given by Professor John Biggs. jbiggs@bigpond.com

ASSESSING LANGUAGE OR CONTENT?

- Biggs, J. (2003). Aligning teaching and assessment to curriculum objectives, LTSN Generic Centre
- Black, P. & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* 80(2), 139-148.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36, 81–109.
- Brewster, J. (2004). Content-based language teaching: a way to keep students motivated and challenged? CATS: The IATEFL Young Learners SIG Publication
- Broadfoot, P. & Black, P. (2004). Redefining assessment? The first ten years of assessment in education, *Assessment in Education: Principles, Policy and Practice* 11 (1), 7-26.
- Brown. H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Longman.
- Brown, J. D. & Hudson, T. (2002). *Criterion-referenced Language Testing*. Cambridge University Press.
- Cambridge ESOL. (2008). *Teaching Science through English – a CLIL approach*. University of Cambridge, ESOL Examinations.
www.CambridgeESOL.org
- Canale, M. & Swain, M. (1980). *Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing*, Applied Linguistics, Vol.1, No1
- CoBaLTT Project, <http://www.carla.umn.edu/cobaltt/>
- Cohen, L, Manion, L. & Morrison, K. (2011). *Research Methods in Education*. 7 ed. London & New York, Routledge
- Corrigan, D., Gunstone, R., Jones, A. (eds.) (2013). *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*. Springer.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*. Council of Europe
Retrieved from http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp.
- Costa, F. (2009). ICLHE/CLIL at the Tertiary Level of Education. State-of-the-Art. *Studies About Languages (Kalbu Stuijos)*, issue 15/2009, pp. 85-89, on www.ceeol.com
- Coyle, D. (1999). Supporting students in content and language integrated learning contexts: planning for effective classrooms. Masih, John (ed), *Learning through a foreign language. Models, methods and outcomes*. Lancaster, CiLT.
- Coyle, D. (2010). *Content and Language Integrated Learning. Motivating Learners and Teachers*. blocs.xtec.cat/clilpractiques1/files/.../slrcoyle.pdf
- Coyle, D, Hood P, & Marsh, D. (2010). CLIL: Content and Language

REFERENCES

- Integrated Learning, Cambridge University Press.
- Creese, A. & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal* 94, 103-115.
- Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H.I. Braun (Eds.), *Test validity* (pp.3-17). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychology tests. First published in *Psychological Bulletin*, 52, 281-302.
<http://psychclassics.yorku.ca/Cronbach/construct.htm>
- Crooks, T. J., Kane, Michael T. & Cohen, A. S. (1996). Threats to the Valid Use of Assessments. *Assessment in Education: Principles, Policy & Practice*, 3(3), 265-285.
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon: Multilingual Matters
- Cummins, J. (2000). Language, Power and Pedagogy. *Bilingual Children in the Crossfire*. Multilingual matters Ltd.
- Cummins, J. & Swain, M. (1986). *Bilingualism and social education: aspects of theory research and practice*. London, Longman.
- Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual review of Applied Linguistics*, 31, 182-204.
- Dalton-Puffer, C. (2007). Discourse in Content and Language Integrated Learning (CLIL) Classrooms. John Benjamins B.V.
- Dalton-Puffer, C. (2008). *Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe*. Werner Delanohy and Laurentz Volkman, (eds.) *Future Perspectives for English Language Teaching*. Heidelberg, Carl Winter.
- Dentler, S. (2007). *Windows on CLIL. Content and Language Integrated Learning in the European Spotlight*, s. 166-171, Alkmaar: European Platform for Education
- Edlund, A. (2011). *Undervisning på engelska i den svenska gymnasieskolan, ett experiment med potential? En studie av tre elevgruppers engelska texter i två register* (Licentiate thesis). Studier i språkdidaktik – Studies in Language Education, 4. Stockholm University.
- Ercikan, K & Roth, W.-M. (2006). What Good Is Polarizing Research Into Qualitative and Quantitative? *Educational Researcher*. Pdf.
- Erickson, G. (2010). Good Practice in Language Testing and Assessment – A matter of Responsibility and Respect. In Kao, T. and Lin Y. (Eds), *A New Look at Teaching and Testing: English as Subject and Vehicle* (s.237-258). Taipei,

ASSESSING LANGUAGE OR CONTENT?

- Taiwan: Bookman Books Ltd
- Euroclit. (2010). <http://www.euroclit.net>
- European Commission. (2006). *Europeans and other languages. Special Eurobarometer, 243*.
- European Commission. (1996). *White paper on teaching and learning: Towards the learning society*. <http://www.ise.ee/docs/whitepaper.htm>
- European Commission. (2014). Improving the effectiveness of language learning: CLIL and computer assisted language learning. http://ec.europa.eu/languages/library/.../clil-call_en.pdf
- European Commission. (2012). Languages 2010 and beyond. http://ec.europa.eu/languages/languages-of-euopre/languages-2010-and-beyond_en.htm
- European Survey on Language Competences (ESLC). (2012). http://ec.europa.eu/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf.
- Eurydice. (2006). *Content and language integrated learning (CLIL) at school in Europe*. Brussels: European Commission. retrieved from http://eacea.ec.europa.eu/education/eurydice/thematic_studies_archives_enphp#2006
- Gablasova, D. (2014). Issues in the assessment of bilingually educated students: expressing subject knowledge through L1 and L2. *The Language Learning Journal*, 42:2, 151-164.
- Gajo, L. (2007). Linguistic Knowledge and Subject Knowledge: How Does Bilingualism Contribute to Subject Development? *International Journal of Bilingual Education and Bilingualism*. 10:5, 563-581.
- García, O. (2012). Theorizing translanguaging for educators. In C.Celic & K. Seltzer, *Translanguaging: ACUNY-NYSIEB guide for educators*, 1-6
- García, O. (2009). *Bilingual Education in the 21st Century: A Global Perspective*. Chichester, UK: Wiley-Blackwell.
- Gatehouse, K. (2001) Key Issues in English for Specific Purposes. Curriculum Development. *The Internet TESL Journal*. Vol. VII, No. 10, October 2001. <http://iteslj.org/Articles/Gatehouse-ESP.html>
- Gibbons, P. (2003). Mediating language learning: Teacher interactions with ESL students in content-based classroom. *TESOL Quarterly*, 27, 247-272.
- Gibson, J. J. et al. (1979). *Notes on affordances*. Retrieved August, 2014 from www.alamut.com/notebooks/a/affordances.html
- Gipps, C. (1999). *Socio-Cultural Aspects of Assessment*. Review of Research in

REFERENCES

- Education 1999 24:355.
- Haataja, K. (2013). Plenary speaker at “Temadag om CLIL” in Stockholm, Sweden, March 2013.
- Halliday, M.A.K, (1994). *An Introduction to Functional Grammar* (2nd edition). London: Arnold.
- Harlen, W. (2007). *Assessment of Learning*. London: SAGE publications.
- Hattie, J. (2009). *Visible Learning*. Routledge.
- Hofer, B. K. & Pintrich, P.R (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning, *Review of Educational Research* 67(1), 88-
- Hogan, T. (2013) SAGE Handbook of Research on Classroom Assessment: *Constructed-Response Approaches for Classroom Assessment*.
- Hogan, T. P. (2007). *Educational assessment: A practical introduction* . New York: John Wiley.
- Hughes, A. (1994). *Backwash and TOEFL 2000*. Commissioned by educational Testing Service (ETS). University of Reading.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing* 12 (2003) 17-29.
- Hyland, K. (2007). Genre pedagogy: Language literacy and L2 writing instruction. *Journal of Second Language Writing* 16 (2007) 148-164.
- Hyltenstam, K. (2004). Engelskan, skolans språkundervisning och svensk språkpolitik. In *Engelskan i Sverige: Språkval i utbildning, arbete och kulturliv*, 36-110. Stockholm: Norstedts Förlag.
- Hymes, D. (1971). *On Communicative Competence*. Philadelphia: University of Pennsylvania Press.
- Hönig, I. (2009). *Assessment in CLIL. A case study*. Universität Wien: Magistra der Philosophie
- Inbar-Lourie, O. (2008). Language assessment culture. Shohamy E & Hornberger N. H (eds.), *Language testing and Assessment. Encyclopedia of Language and education, vol 7*. Springer. 285-299.
- Jönsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity. *Educational research Review*, 2, s. 130-144.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.
- Kiely, R. (2012). CLIL – The question of assessment. *Developing Teachers.com*
- Kjellén Simes, M. (2008). *Room for improvement? A comparative study of Swedish learners' free written production in English in the foreign language classroom and in immersion education*. Diss. Karlstad : Karlstads universitet, 2008. Karlstad.
- Krashen, S. (2008). Language education, Past, present and future. *RELC Journal* 2008 39: 178

ASSESSING LANGUAGE OR CONTENT?

- Krashen, S. (1989). We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis. *The Modern Language Journal*. 73:iv. 440-464.
- Krashen, S. & Brown, C. L. (2007). What is academic language proficiency? *Research Papers*. Singapore Tertiary English Teachers Society, 1-4. Retrieved from <http://www.sdkrashen.com>
- Krashen, S. & Terrell, T. D. (1983). *The Natural Approach. Language Acquisition in the Classroom*. Oxford: Pergamon.
- Lantolf, J. P. & Poehner, M. E. (2008). Dynamic assessment. *Encyclopedia of Language and education, 2nd edition, Volume 7: Language testing and assessment*, 273-284. Springer Science & Business Media LLC.
- Levin, L. & Marton, F. (1973). *Provt teori och provkonstruktion*. Stockholm Almqvist & Wiksell.
- Liberg, C. (2009). *Genrepedagogik i ett didaktiskt perspektiv*. In Symposium 2009. Genrer och funktionellt språk i teori och praktik. Red. Olofsson, M. Nationellt centrum för andra språk. Stockholm, Stockholms universitets förlag.
- Lim Falk, M. (2008). *Svenska i engelskspråkig skolmiljö: ämnesrelaterat språkbruk i två gymnasieklasser*. Doctoral thesis. Stockholm: Stockholms universitet,
- Lévesque, S. (2008), *Thinking historically: educating students for the twenty-first century*, Toronto 2008.
- Lindberg, I. (2009). Conceptualizing school-related, academic language – theoretical and empirical approaches. In P. Juvonen (Ed), *Papers from the ASLA Symposium in Stockholm, 7-8 November, 2008, Association suédoise de linguistique appliquée (ASLA). language and Learning*, 27-42. (ASLA:s skriftserie nr 22). Stockholm: Stockholms universitet.
- Lindberg, I. (2011). Fostering multilingualism in Swedish schools – intentions and realities. In R. Källström and I. Lindberg (Eds.), *Young urban Swedish: Variation and change in multilingual settings*, 149-172. University of Gothenburg, Sweden.
- Lindmark, T. (2013). *Samhällskunskapslärares ämneskonceptioner*. (Doctoral thesis.) Umeå: Pedagogiska institutionen, Umeå universitet. <http://umu.diva-portal.org/>
- Little, D. (2009). *The European Language Portfolio: where pedagogy and assessment meet*. Strasbourg: Council of Europe Retrieved from http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Publications/ELP_pedagogy_assessment_Little_EN.pdf.
- Little, D. & Erickson, G. (2015) Learner Identity, Learner Agency, and the Assessment of Language Proficiency: Some Reflections Prompted by the *Common European Framework of Reference for Languages. Annual Review of*

REFERENCES

- Applied Linguistics*, **35** (2015), pp. 120–139. © Cambridge University Press, 2015
- Llinares, A., Morton, T. & Whittaker, R. (2012). *The Roles of Language in CLIL*. Cambridge, Cambridge University Press.
- Lundahl, C. (2007). Kunskapsbedömning och kunskapssyn. *I kunskapens namn – en antologi om kunskap, makt och kreativitet*.
- Lyster, R. & Balinger, S. (2011). Content-based language teaching: Convergent concerns across divergent contexts. *Language Teaching Research on content-based language teaching (CBLT)*. Sage publications.
<http://ltr.sagepub.com/content/15/3/279.abstract>
- Maiworm, F. & Wächter, B. (2008). *English-Taught Programmes in European Higher Education. The Picture in 2008*. Bonn:Lemmens.
- Malone, M. E. (2008). Training in language assessment. E. shohamy and N. H. Hornberger (eds). *Encyclopedia of Language and Education*. 2nd Edition, Vol. 7: Language Testing and Assessment, 225-239. © 2010 Springer Science+Business Media LLC.
- Masih, J. ed (1999). *Learning through a foreign language. Models, methods and outcomes*. London, CiLT.
- Massler, U. , Stotz, D. & Quessier, C. (2014). Assessment instruments for primary CLIL: the conceptualisation and evaluation of test tasks. *The Language Learning Journal*, 42:2, 137-150.
- McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction* (5th ed). Upper Saddle River, NJ: Pearson.
- McNamara, T. (2006). Validity in Language Testing: The Challenge of Sam Messick's legacy. *Language assessment quarterly*, 3:1, 31-51.
- Mehisto, P, Marsh, D. & Frigols, J. M. (2008). *Uncovering CLIL: Content and Language Integrated Learning in Bilingual and Multilingual Education*. Oxford, Macmillan Education.
- Messick, S. (1989). *Validity*. I R. L. Linn (Ed.), Educational measurement, Vol. 3 (pp.13-103). New York: American Council on Education
- Messick, S. (1996). *Validity and Washback in Language Testing*. Educational Testing Service, Princeton: New Jersey.
- Mohan, B. (1986). *Language and content*. Reading, MA: Addison-Wesley.
- Morgan, C. (2006). Appropriate language assessment in content and language integrated learning. *Language Learning Journal*, No 33, 59-67.
- Morgan, C. (1999). Teaching history in a foreign language: what language? Masih, John (ed), *Learning through a foreign language. Models, methods and outcomes*. Lancaster, CiLT.

ASSESSING LANGUAGE OR CONTENT?

- Moss, P. et al. (2006). Validity in Educational Assessment. In *Review of research in Education*, 30(1), 109-162. London: Sage Publications.
- Nikula, T. (2005). English as an object and tool of study in classrooms: International effects and pragmatic implications. *Linguistics and Education* 16, 27-58.
- Nikula, T. (2007). Speaking English in Finnish content-based classrooms. *World Englishes*, Vol 26, No2, 206-223.
- Nixon, J. (2001). *Quality in SPRINT. Towards quality assessment and assurance in content and language integrated education. A field study report.* Stockholm: Skolverket.
- Nixon, J. (2000). *SPRINT: Språk- och innehållsintegrerad inlärnin och undervisning. Rapport på uppdrag av Skolverket.* Stockholm: Skolverket.
- Odenstad, C. (2010). *Prov och bedömning i samhällskunskap. En analys av gymnasielärares skriftliga prov.* (Licentiate thesis. Karlstad University Studies, 2010:41.) Karlstad: Karlstad University Press
- OECD. (2011). *Reviews of Evaluation and Assessment in Education: Sweden.* OECD Publishing.
- OECD. (2012). "Sweden" in *International Migration Outlook 2012*, OECD Publishing
- Olander, C. (2014). Ppt from conference, personal communication.
- Olander, C. & Ingerman, Å. (2011). Towards an inter-language of talking science: exploring students' argumentation in relation to authentic language. *Journal of Biological Education*, 45:3, 158-164.
- Olsson, E. (2011) "Everything I read on the Internet is in English" On the impact of extramural English on Swedish 16-year-old pupils' writing proficiency, (Licentiate thesis.) Institutet för svenska som andraspråk, Institutionen för svenska språket, Gothenburg University.
- Oscarson M, & Apelgren B M. (2010). Mapping language teachers' conceptions of student assessment procedures in relation to grading: A two-stage empirical inquiry. Science Direct.
- Pace, D. (2011). Assessment in History: The case for "decoding" the discipline. *Journal of the Scholarship of Teaching and Learning*, Vol. 11, No. 3, August 2011, pp. 107 – 119.
- Pérez Vidal, C. (2007). The need of focus on form (FoF) in content and language integrated approaches: an exploratory study. *Volumen Monográfico*:39-54.
- Popham, J. W. (2006). All About Accountability/Needed: A Dose of Assessment Literacy *Improving Professional Practice*, **63** 84-85
- Popham, J. W. (2009). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48:4–11, 2009 Copyright © The College of Education and Human Ecology, The Ohio State University

REFERENCES

- Popham, J. W. (2011). *Classroom assessment: What teachers need to know*. (6th ed.). Upper Saddle River, NJ: Pearson.
- Porter, A. J. (2004). Curriculum Assessment. for Research in Education. Washington, DC: American Educational Research Association, 2004. 1. 1. <https://datacenter.spps.org/uploads/curricassess.pdf>
- Rapley, T. (2008) Silverman, D. (2011). Qualitative Research. Issues of Theory, *Method and Practice*. 3rd ed. Los Angeles, SAGE.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing* 2004 21: 249. DOI: 10.1191/0265532204lt283ed
- Rosenlund, D. (2011) *Att hantera historia med ett öga stängt: Samstämmighet mellan historia A och lärares prov och uppgifter*. Lund: Forskarskolan i historia och historiedidaktik. Lunds Universitet och Malmö högskola
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18:119-144. © Kluwer Academic Publishers, Dordrecht.
- Schleppegrell, M. J. (2004). *The Language of Schooling. A Functional Linguistics Perspective*. New Jersey: Lawrence Erlbaum Associates.
- Schleppegrell, M. & O'Hallaron, C. (2011). Teaching Academic Language in L2 Secondary Settings. *Annual Review of Applied Linguistics* (2011), 31, 3–18. Cambridge University Press, 2011, 0267-1905/11 doi: 10.1017/S0267190511000067
- Selghed, B. (2011). *Betygen I skolan. Kunskapssyn, bedömningsprinciper och lärarpraxis*. Stockholm, Liber.
- Serve Center, (2006). *How to Assess Student Performance in History: Going Beyond Multiple Choice Tests*. University of North Carolina at Greensboro.
- Shadish, W.R., Cook, T.D., Campbell, D.T. (2002) *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin Company
- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*. 19, pp. 405-450.
- Shepard, L. A. (2000). The role of assessment in a learning culture, *Educational Researcher* 29(7), 4-14.
- Sherris, A. (2008). Integrated Content and Language Instruction. *CAL digest*. September 2008.
- Shohamy, E. (2008). E. Shohamy and N. H. Hornberger (eds). *Encyclopedia of Language and Education*. 2nd Edition, Vol. 7: Language Testing and Assessment, 225-239. © 2010 Springer Science+Business Media LLC.
- Shulman, L.S. (1987). Knowledge and Teaching: Foundations of the New

ASSESSING LANGUAGE OR CONTENT?

- Reform. *Harvard Educational Review* (1987). 57, 1-22.
- Siegel, M., Menon, D., Sinha, S., Promyod, N., Wissehr, C. & Halverson, K. L. (2014). Equitable Written Assessments for English Language Learners: How Scaffolding Helps. *J Sci Teacher Educ.* (2014) 25:681-708.
- Siegel, M. & Surian, L. (eds.) (2012). *Access to language and Cognitive Development.* Oxford.
- Silverman, D. (2011). *Qualitative Research. Issues of Theory, Method and Practice.* 3rd ed. Los Angeles, SAGE.
- Snow, M. A., Met, M., & Genesee, F. (1989). A conceptual framework for the integration of language and content instruction. *TESOL Quarterly*, 23, 201-217.
- Socrates-Comenius 2.1 project (2009). *CLIL across contexts: A scaffolding framework for teacher education.* (<http://clil.uni.lu>)
- SOU 2008: 26. Värna språken – förslag till språklag.
www.sweden.gov.se/content/1/c6/10/09/59/4ad5deaa.pdf
- Stiggins, R. J. (1995). Assessment Literacy for the 21st Century. *The Phi Delta Kappan*, 11/1995, Vol 77, Issue 3
- Stiggins, R. J. (2014). Improve assessment literacy outside of schools too. *The Phi Delta Kappan*, Vol 96, (2) 67-72
- Stolare, M. *Att fånga historieundervisning: Ett förslag till analysmodell.* Sid 143-154. Linköping: Institutionen för kultur och kommunikation, Linköpings universitet, i Martinsson, G & Parmenius Swärd, S (2011) Ämnesdidaktik – dåtid, nutid och framtid: Bidrag från femte rikskonferensen i ämnesdidaktik vid Linköpings universitet 26-27 maj 2010.
- Sundqvist, P., & Sylvén, L. K. (2012). World of VocCraft: Computer games and Swedish learners' L2 vocabulary. In H. Reinders (Ed.), *Digital games in language learning and teaching* (pp. 189-208). Basingstoke: Palgrave Macmillan.
- Swedish National Agency for Education. (2013). *Curriculum for the upper secondary school.* <http://www.skolverket.se/publikationer?id=2975>
- Swedish National Agency for Education. (2012b). *Greppa språket. Ämnesdidaktiska perspektiv på flerspråkighet.*
- Swedish National Agency for Education. (1996). *Grundskola för bildning – kommentarer till läroplan, kursplaner och betygskriterier.* Stockholm: Skolverket/Liber Distribution.
- Swedish National Agency for Education. (2011a). Hur ser bedömningspraxis ut i Sverige?

REFERENCES

- <http://www.skolverket.se/skolutveckling/forskning/bedomning/tema-bedomning/hur-ser-bedomningspraxis-ut-i-sverige-1.157700>
- Swedish National Agency for Education. (2005). *National Assessment and Grading in the Swedish School System*. Stockholm: Skolverket.
- Swedish National Agency for Education. (2012a). (National course goals in history, biology and English)
<http://www.skolverket.se/om-skolverket/andra-sprak-och-lattlast/english/the-swedish-education-system/upper-secondary-school/about/subjects-available-in-english-1.209453>
- Swedish National Agency for Education. (2011b). Vilka typer av bedömningsformer finns?
<http://www.skolverket.se/skolutveckling/forskning/bedomning/tema-bedomning/vilka-typer-av-bedomningsformer-finns-1.157725>
- Sylvén, L.K. (2013). CLIL in Sweden – why does it not work? A metaperspective on CLIL across contexts in Europe. *International Journal of Bilingual Education and Bilingualism*, 16 (3), 301-320.
- Sylvén, L. K. (2006). Extramural exposure to English among Swedish school students. In J. Einarsson, E. Larsson Ringqvist & M. Lindgren (Eds.), *Approaches to teaching and learning in linguistic research: Papers from the ASLA symposium in Växjö, 10-11 November 2005* (pp. 252-260). Uppsala: ASLA, Svenska föreningen för tillämpad språkvetenskap.
- Sylvén, L.K. (2004). *Teaching in English or English Teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*. University of Gothenburg
- Sylvén, L. K. & Ohlander, S. (2014) The CLISS Project: Receptive Vocabulary in CLIL versus non-CLIL Groups. *Moderna Språk* 108(2): 80-114.
- Sylvén, L.K. & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24, pp.302-321.
- Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, ISSN0007-1005 Vol. 53, No. 4, December 2005, s. 466–478.
- Tardy, C. M. (2006). Researching first and second language genre learning: A comparative review and a look ahead. *Journal of Second Language Writing* 15, 79-101.
- Teachinenglish.
<http://www.teachingenglish.org.uk/article/test-question-types>
- The SERC portal for Educators.

ASSESSING LANGUAGE OR CONTENT?

- <http://serc.carleton.edu/sp/library/interdisciplinary/index.html>
- TIMSS. (2011). *Assessment Frameworks*. TIMSS & PIRLS International Study Center. Lynch School of Education, Boston College.
- Tittle, C. (1994). Toward an educational psychology of assessment for teaching and learning – Theories, contexts, and validation arguments. *Educational Psychologist* 29, 149-162.
- Trost, J. (2001). *Enkätboken*. Lund, Studentlitteratur.
- University of Cambridge. (2008). Teaching Knowledge Test (TKT) Content and Language Integrated Learning (CLIL) Handbook for teachers.
- Van Lier, L. (2004). *The ecology of language learning and classroom Research*. Retrieved August 2014 from www.utc.fr/~untele/2004ppt/vanlier_2004.ppt
- Vetenskapsrådet. (2011). (The Swedish Research Council) God forskningsred. <http://vr.se/etik/publikationerochriktlinjer.4.45a6e939122880e7d8e80001820.html>
- Wang, Y., & Wen, Q. F. (2002). L1 use in the L2 composing process: an explanatory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, 11(1) 225-246.
- Wedman, I. (1988). *Prov och provkonstruktion*. Stockholm, Utbildningsförlaget
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wewer, T. (2014). *Assessment of Young Learners' English Proficiency in Bilingual Content Instruction CLIL*. University of Turku, Faculty of Education/Department of Teacher Education, Doctoral Dissertation.
- White, E. (2009). *Are You Assessment Literate? Some Fundamental Questions Regarding Classroom-based Assessment*. Feature article. <http://www.slideshare.net/ewhite/are-you-assessment-literate?related=1>
- Whittaker, R., Llinares, A. & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research* 2011 15:343-362.
- Wikström, C. (2014). *Konsten att göra bra prov*. Stockholm, Natur och Kultur.
- Yoxsimer Paulsrud, B. (2014). *English-Medium Instruction in Sweden. Perspectives and practices in two upper secondary schools*. Stockholm University, Doctoral thesis in Education and languages and language development.

LIST OF APPENDICES

Appendix 1: Glossary

Appendix 2: Questionnaire in Swedish

Appendix 3: Questionnaire in English

Appendix 4: Overview teachers' responses to questionnaire

Appendix 1

Glossary

Academic function words/function words: More or less the same as cognitive descriptor words. Describes cognitive skills needed to perform a task, manifest through language use, e.g. “analyze”, “describe”

Accommodations: Alternations in the way tasks are presented

BICS/Basic Interpersonal Communicative Skills: Cognitively undemanding surface skills (Cummins, 1979)

CALP/Cognitive Academic Language Proficiency: Cognitively demanding language skills, needed to understand and produce academic language (Cummins, 1979)

CBI/Content based instruction: Content/subject matter used as a vehicle for foreign language learning

CBLT/Content based language teaching: see CBI

CC/Content compatible language: General academic language

CEFR/Common European Framework of Reference

CLIL/Content and Language Integrated Learning: Umbrella term for bilingual teaching, most often by the use of English

CO/Content obligatory language: Disciplinary specific language, including disciplinary concepts

Content subjects: as opposed to language courses, e.g. biology and history

Course goals: Comparable to course objectives or learning aims

Coyle’s four Cs: Conceptual framework for CLIL, comprising Content (subject matter), Communication (language learning and use), Cognition (thinking processes), Culture (intercultural understanding)

Cross curricular/disciplinary: Features or processes valid in several disciplines (at least two)

EFL/English as a foreign language: Signaling the status of English, not being an official language used in the community

EMI/ English medium instruction: bilingual teaching in English

ESL/English as a second language: Signaling the status of English, as an additional language used in the community, but not the learner’s L1

FL/Foreign language: Language other than the first language (L1), generally not used in the community

APPENDICES

Formative assessment: Also referred to as assessment for learning; observations during a course to help feed the students forward or make adjustments in the instructional process

Genre: Text types used in different academic subjects

Higher order thinking skills/HOTS: Advanced cognitive abilities; e.g. analysis, problem solving, creating.

Hybrid language: A type of interlanguage (see below)

ILO/Intended learning objective: Part of constructive alignment (Biggs 1999); formulated first, from which the assessment criteria and assessment design are derived

Interdisciplinary: Merging or integrating themes or features in two or more disciplines

Interlanguage: Denotes intermediate variants of a learner's language, developing towards the target language (Selinker 1992)

Lower order thinking skills/LOTS: Basic simpler cognitive abilities; e.g. remembering, recalling, describing.

Production test: Test involving students' own linguistic production, in the present study as in writing assignments and essays. Could be compared to performance assessment.

Question test: Test involving multiple tasks, test items or questions, in the present study as in written paper pencil tests

Scaffolding: Temporary interventions to perform a task and support learning

Summative assessment: Assessment leading up to a grade, achievement is measured at the end of a theme or a course

Translanguaging: Shifting and mixing of languages and registers

Writing assignments: Assignments purposed to be in the written mode to assess writing skills, preceded by a writing prompt, instructions and sometimes knowledge requirements

Written assignments: As opposed to oral assignments

Appendix 2

Lärarenkät

Bakgrundsfrågor

1. Den här enkäten svarar jag på i egenskap av:
engelsklärare
historielärare
biologilärare
2. Har du lärarexamen i detta ämne? Ja Nej
Om ja, vilket år tog du din examen?
3. Hur många år har du undervisat i ämnet?
4. Ämneslärare (i historia eller biologi), undervisar du något ämne *på* engelska (CLIL/SPRINT)?
Ja Nej
5. Om du undervisar på engelska
 - a) Hur många år har du gjort det?
 - b) Undervisar du samtidigt samma ämne på svenska?
Ja Nej
6. Skulle du vilja ha utbildning/fortbildning i engelska?
Ja Nej
7. Om du är språklärare, arbetar du ämnesintegrerat i någon form?
Ja Nej
8. Om du arbetar ämnesintegrerat, hur länge har du gjort det?
9. Med ämnesintegrerat avser jag:

Frågor som rör bedömning och kursinnehåll

10. Hur anser du att det är att sätta betyg i ditt ämne?
Lätt Ganska lätt Ganska svårt Svårt
11. Vilken betydelse har följande faktorer för din undervisning när du bedömer/sätter betyg?
 - a) Ämnesplanens målpunkter
Mycket stor Ganska stor Ganska liten Mycket liten

APPENDICES

- b) Nationella kunskapskrav
Mycket stor Ganska stor Ganska liten Mycket liten
- c) Nationella kursprov, i de fall det finns
Mycket stor Ganska stor Ganska liten Mycket liten
12. Min personliga uppfattning överensstämmer med Skolverkets bedömningsanvisningar i ämnet
Överensstämmer helt Överensstämmer till viss del Överensstämmer inte alls
Kommentar till föregående fråga:
13. Vilka typer av underlag använder du dig av
a) vid den summativa bedömningen som leder fram till ett betyg?
b) vid den kontinuerliga formativa bedömningen?
14. Vid betygsättning, vilka bedömningsgrunder anser du vara mest användbara?
De muntliga
De skriftliga
Båda i lika stor utsträckning
Kommentar till föregående fråga:
15. Vad inkluderar du i ditt bedömningsunderlag?
Bara innehåll
Bara språk
Både språk och innehåll
Det beror på uppgiftens utformning
Kommentar till föregående fråga:
16. Vilken betydelse har elevens provresultat, i jämförelse med andra bedömningsunderlag, när du sätter slutbetyg i kursen?
Betyder allt ganska mycket ganska lite
17. Vilka faktorer väger tyngst vid bedömning av elevers kunnande i ditt ämne?
Färdigheter i form av skriftlig och muntlig kompetens
Innehållsliga områden/moment (t ex kunskap om olika epoker, system, och teorier)
Användandet av ett för ämnet specifikt ordförråd
Användandet av ett mer generellt ordförråd och språkligt flyt
Behärskning olika uttrycksformer och modala hjälpmedel
Språklig korrekthet
Språklig komplexitet
Kunna behärska en ämnesrelaterad skriftlig genre
Analysförmåga

ASSESSING LANGUAGE OR CONTENT?

18. Var finns den viktigaste källan till stoff till de olika bedömningsmomenten?
Kursbok
Lärarens klassrumsgenomgångar
Aktiviteter i klassrummet/under kursen
Annat:
19. Om du är ämneslärare och undervisar i ditt ämne på engelska, tycker du att det finns någon skillnad mellan att utforma ett prov på svenska jämfört med engelska, förutom språket?
Ja Nej
Om ja, vari består den skillnaden?
20. Vilket eller vilka områden skulle du vilja kunna mer om när det gäller betyg och bedömning?
Tolkning av nationella kunskapskrav
Hur man konstruerar prov
Bedömning av muntlig produktion
Bedömning av skriftlig produktion
Alternativa bedömningsformer (t ex portfolio, självbedömning)
Alternativa bedömningsverktyg (t ex datoriserade)
Känner inget behov av fortbildning när det gäller betyg och bedömning
Annat
Om du svarade "annat" på föregående fråga, utveckla det här:

Frågor om ämnesintegrering och sambedömning

21. Har du arbetat med sambedömning tillsammans med annan ämneskollega
Ja Nej
Om ja, i vilket ämne?
22. Har ni använt prov där en av er har bedömt innehåll och den andre språk?
Ja Nej
23. Om nej, skulle du vilja samarbeta kring bedömning med en kollega?
Ja Nej
24. Vad är tänkbara hinder för tvärvetenskaplig sambedömning tror du?
Brist på tid
Brist på pedagogisk samsyn kollegor emellan
Brist på organisatoriska förutsättningar (t ex schema)
Ämnenas olika karaktär

APPENDICES

Appendix 3

Teacher questionnaire

Background information

1. I respond to this questionnaire as:
EFL teacher
History content teacher
Biology content teacher

2. Are you certified to teach in your subject? Yes No
If yes, which year did you receive your certification?

3. How many years have you taught your subject?

4. Content teachers (in history or biology), do you teach a subject using English as a medium of instruction (CLIL/SPRINT)?
Yes No

5. If you teach in English:
c) How many years have you done so?
d) Do you at the same time teach the same subject in Swedish?
Yes No

6. Would you like education/more training in English?
Yes No

7. If you are an EFL teacher, do you work interdisciplinary with a colleague?
Yes No

8. If you work interdisciplinary, how long have you been doing so?

9. By interdisciplinary I mean:

Questions concerning assessment and course content

10. How do you perceive of grading in your subject?
Easy Fairly easy Fairly difficult Difficult

11. How much do the following impact your teaching for assessment/grading purposes?
a) The course goals
Very much Pretty much Fairly little Very little

ASSESSING LANGUAGE OR CONTENT?

- b) National knowledge requirements
Very much Pretty much Fairly little Very little
- c) National test, when applicable
Very much Pretty much Fairly little Very little
12. My personal view agrees with the assessment guidelines of the National Agency of Education in my discipline
Fully agree Partly agree Disagree
Commentary to the previous question:
13. Which types of basis for assessment do you use
b) for the summative assessment leading up to a grade?
c) for the continuous formative assessment?
14. When grading, which basis for assessment do you find most useful?
The oral
The written
Both to the same extent
Commentary to the previous question:
15. What is included in the basis for assessment?
Content only
Language only
Both language and content
It depends on the type of task
Commentary to the previous question:
16. How important are test scores, compared with other basis for assessment, at the final grading in a course?
Means everything Means a lot Means fairly little
17. Which factors weigh the most when assessing students' knowledge in your subject?
Oral and written skills
Content areas/themes (e.g. knowledge about eras, systems, theories)
Use of content obligatory/subject specific vocabulary
Use of content compatible/general academic vocabulary and fluency
Mastery of various forms of expression and multimodal instruments
Linguistic accuracy
Linguistic complexity
Mastery of a disciplinary written genre
Analytical skills

APPENDICES

18. Where do you find the most important source of material for the different areas of assessment?
Course book
Teacher's own classroom lectures
Activities in the classroom/during the course
Other:
19. If you are content teacher and teach a subject in English, do you feel there is a difference designing the test in Swedish compared with English, other than the language?
Yes No
If yes, what difference?
20. Which area(s) would you like to know more about in relation to grading and assessment? Interpretation of national knowledge requirements
Test design
Oral assessment
Written assessment
Alternative assessment modes (e.g. portfolio, self-assessment)
Alternative assessment tools (e.g. computer based)
Do not feel a need for training regarding grading and assessment
Other
If you responded "other" in the previous question, please elaborate here:

Questions concerning interdisciplinary integration and collaborative assessment

21. Have you collaborated with a colleague around assessment?
Yes No
If yes, in which subject?
22. Have you used a test where one of you has assessed content and the other assessed language?
Yes No
23. If no, would you like to collaborate with a colleague around assessment?
Yes No
24. What may hinder interdisciplinary collaboration around assessment do you think?
Lack of time
Lack of pedagogical agreement between colleagues
Lack of organizational prerequisites
The different character of the disciplines

ASSESSING LANGUAGE OR CONTENT?

Appendix 4 Summary of CLIL and EFL teachers' experiences

Teacher and discipline	Teachers' views on CLIL	Work load/teacher training	Interdisciplinary collaboration
Biology CLIL	<ul style="list-style-type: none"> • It is stimulating, fun, developing • I'm aware I'm not bilingual, but do my utmost to give the students an adequate instruction • Provides students with a bigger disciplinary vocabulary 	<ul style="list-style-type: none"> • Takes a lot of time to produce material, to align material with national course goals. English course books focus more on details, less on • Had to find a course myself, subscribes to BBC wildlife • Have asked the school leaders for training, not received any yet 	<ul style="list-style-type: none"> • I don't care about language, that is up to the English teachers • I can borrow dictionaries, or ask for certain terms • Have received some good ideas, word games, how to work with vocabulary • Done in social science/English but not in biology
History CLIL	<ul style="list-style-type: none"> • Cool to acquire a language • Hard to find teachers willing to teach in English • Not the first choice to teach in English • Have to point out that you are not a native speaker • Very positive for the students , opens opportunities to study abroad 	<ul style="list-style-type: none"> • A lot of extra work • Extra strain since you doubt your own language skills • Not from the school • Would have been interesting to visit some other CLIL-schools go get some inspiration for course material and methods 	<ul style="list-style-type: none"> • Organizational issues prevent teachers from finding the time • Enriching, but very time consuming • Did more in a previous non-CLIL school
English	<ul style="list-style-type: none"> • It is a good teaching environment, motivated students • CLIL-students have a higher proficiency level than non-CLIL • Good for the students' future studies • Questionable to have non-native content teachers use an L2 as the language of instruction • Quality and level of language in both L1 and L2? 		<ul style="list-style-type: none"> • Too time consuming • Have to cover the English course content, no room for other themes • A bit afraid of unfamiliar subject themes • Would be ideal, but doesn't seem like there is time • One assignment two teachers, could alleviate some stress for the students