



Turun yliopisto
University of Turku

**ASSESSMENT OF YOUNG LEARNERS'
ENGLISH PROFICIENCY
IN BILINGUAL CONTENT INSTRUCTION CLIL**

Taina Wewer

University of Turku

Faculty of Education

Department of Teacher Education

Educational Sciences

Doctoral Programme on Learning, Teaching and Learning Environments Research

Supervised by

Professor Annikki Koskensalo

Adjunct Professor Tuula Merisuo-Storm

Department of Teacher Education

University of Turku

Finland

Reviewed by

Professor Tarja Nikula

Professor Ari Huhta

Centre for Applied Language Studies

University of Jyväskylä

Finland

Opponents

Professor Tarja Nikula

Professor Ari Huhta

Centre for Applied Language Studies

University of Jyväskylä

Finland

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-951-29-5753-8 (PRINT)

ISBN 978-951-29-5754-5 (PDF)

ISSN 0082-6987

Painosalama – Turku, Finland 2014

UNIVERSITY OF TURKU
Faculty of Education/Department of Teacher Education

WEWER, TAINA: Assessment of Young Learners' English Proficiency in Bilingual Content Instruction CLIL

Doctoral Dissertation, 257 pages, 12 appendices
Applied Linguistics
June 2014

This applied linguistic study in the field of second language acquisition investigated the assessment practices of class teachers as well as the challenges and visions of language assessment in bilingual content instruction (CLIL) at primary level in Finnish basic education. Furthermore, pupils' and their parents' perceptions of language assessment and LangPerform computer simulations as an alternative, modern assessment method in CLIL contexts were examined. The study was conducted for descriptive and developmental purposes in three phases: 1) a CLIL assessment survey; 2) simulation 1; and 3) simulation 2. All phases had a varying number of participants.

The population of this mixed methods study were CLIL class teachers, their pupils and the pupils' parents. The sampling was multi-staged and based on probability and random sampling. The data were triangulated. Altogether 42 CLIL class teachers nationwide, 109 pupils from the 3rd, 4th and 5th grade as well as 99 parents from two research schools in South-Western Finland participated in the CLIL assessment survey followed by an audio-recorded theme interview of volunteers (10 teachers, 20 pupils and 7 parents). The simulation experimentations 1 and 2 produced 146 pupil and 39 parental questionnaires as well as video interviews of volunteered pupils. The data were analysed both quantitatively using percentages and numerical frequencies and qualitatively employing thematic content analysis.

Based on the data, language assessment in primary CLIL is not an established practice. It largely appears to be infrequent, incidental, implicit and based on impressions rather than evidence or the curriculum. The most used assessment methods were teacher observation, bilingual tests and dialogic interaction, and the least used were portfolios, simulations and peer assessment. Although language assessment was generally perceived as important by teachers, a fifth of them did not gather assessment information systematically, and 38% scarcely gave linguistic feedback to pupils.

Both pupils and parents wished to receive more information on CLIL language issues; 91% of pupils claimed to receive feedback rarely or occasionally, and 63% of them wished to get more information on their linguistic coping in CLIL subjects. Of the parents, 76% wished to receive more information on the English proficiency of their children and their linguistic development. This may be a response to indirect feedback practices identified in this study.

There are several challenges related to assessment; the most notable is the lack of a CLIL curriculum, language objectives and common ground principles of assessment. Three diverse approaches to language in CLIL that appear to affect teachers' views on language assessment were identified: instrumental (language as a tool), dual (language as a tool and object of learning) and eclectic (miscellaneous views, e.g. affective factors prioritised). LangPerform computer simulations seem to be perceived as an appropriate alternative assessment method in CLIL.

It is strongly recommended that the fundamentals for assessment (curricula and language objectives) and a mutual assessment scheme should be determined and stakeholders' knowledge base of CLIL strengthened. The principles of adequate assessment in primary CLIL are identified as well as several appropriate assessment methods suggested.

Keywords: language assessment, second language proficiency, bilingual education, content and language integrated learning, CLIL, English language, computer simulation

TURUN YLIOPISTO

Kasvatustieteiden tiedekunta/Opettajankoulutuslaitos

WEWER, TAINA: Nuorten oppijoiden englannin kielitaidon arviointi kaksikielisessä CLIL - sisällönopetuksessa

Väitöskirjatutkimus, 257 sivua, 12 liitettä

Soveltava kielitiede

Kesäkuu 2014

Tämä vieraan kielen oppimisen, opettamisen ja arvioinnin tutkimus tarkasteli luokanopettajien arviointikäytänteitä sekä kielitaidon arvioinnin haasteita ja visioita kaksikielisessä sisällönopetuksessa (CLIL) suomalaisen perusopetuksen alakoulussa. Lisäksi tutkittiin oppilaiden ja heidän vanhempiansa käsityksiä arvioinnista ja LangPerform-tietokonesimulaatioiden soveltuvuudesta CLIL-arviointiin uutena vaihtoehtoisena arviointimenetelmänä. Kuvailtava ja kehittävä tutkimus toteutettiin kolmessa vaiheessa: CLIL-arviointikysely, simulaatio 1 ja simulaatio 2. Kaikissa vaiheissa oli eri määrä osallistujia

Tutkimuksen populaation muodostivat CLIL-luokanopettajat, heidän oppilaansa sekä näiden huoltajat. Otos oli monivaiheinen ja perustui sekä harkinnan- että sattumanvaraisuuteen. Aineisto trianguloitiin. Yhteensä 42 opettajaa eri puolilta Suomea, 109 kolmas-, neljäs- ja viidesluokkalaista oppilasta sekä heidän 99 huoltajaansa kahdesta erityyppisestä lounaissuomalaisesta koulusta osallistui CLIL-arviointikyselyyn, jota syvennettiin nauhoitetuin teemahaastatteluin (10 opettajaa, 20 oppilasta ja 7 huoltajaa). Simulaatiokokeiluista 1 ja 2 kerättiin yhteensä 146 oppilaskyselykaavaketta ja 39 huoltajakaavaketta sekä vapaaehtoisten oppilaiden videohaastatteluita. Aineisto analysoitiin käyttäen sekä kvalitatiivisia (teemoitteleva sisältöanalyysi) että kvantitatiivisia (prosenttiosuudet, frekvenssit) menetelmiä.

Tutkimusaineistosta kävi ilmi, että kielitaidon arviointi alakoulun CLIL-opetuksessa ei ole vakiintunut käytäntö. Se näyttää usein olevan epäsäännöllistä, satunnaista, epäsuoraa sekä enemmän vaikutelmiin kuin näyttöön tai opetussuunnitelmaan perustuvaa. Eniten käytetyt arviointimenetelmät olivat opettajan observointi, kaksikieliset kokeet tai koeosuudet sekä opetuskeskustelu; vähiten käytettiin portfolioita, simulaatioita ja vertaisarviointia. Vaikka opettajat yleisesti pitivät kielitaidon arviointia tärkeänä, viidennes heistä ei kerännyt arviointitietoa systemaattisesti ja 38 % antoi oppilaille kielellistä palautetta harvoin.

Sekä huoltajat että oppilaat toivoivat saavansa enemmän palautetta CLIL-kielitaidosta. Oppilasta 91 % koki saavansa palautetta harvoin tai satunnaisesti, ja 63 % toivoi saavansa enemmän tietoa osaamisestaan englannin kielellä eri oppiaineissa. Vanhemmista 76 % toivoi saavansa enemmän tietoa lapsensa englannin kielitaidosta ja sen kehittymisestä. Tämä saattaa olla vastareaktiota tutkimuksessa havaitusta opettajien epäsuorasta palautekäytännöstä.

Arviointiin liittyy monia haasteita, joista merkittävimpiä ovat CLIL-opetussuunnitelman ja kielitaitotavoitteiden sekä yhteisten arviointikäytänteiden puute. Tutkimuksessa huomattiin kolme erilaista lähestymistapaa kieleen, jotka näyttävät vaikuttavan opettajien suhtautumiseen kielitaidon arviointiin CLIL-opetuksessa: instrumentaalinen (kieli välineenä), kaksoisfokus (kieli välineenä ja oppimisen kohteena) sekä valikoiva (sekalaisia näkemyksiä, affektiiviset tekijät keskiössä). Lisäksi LangPerform -tietokonesimulaatiokonsepti vaikuttaa toimivalta vaihtoehtoiselta arviointimenetelmältä alakoulun CLIL-opetuksessa.

Tutkimuksen perusteella suositellaan voimakkaasti arvioinnin perusteiden (opetussuunnitelma, kielitavoitteet) ja yhteisen arviointirakenteen määrittämistä sekä asianosaisten CLIL-tietopohjan vahvistamista. Tutkimuksessa identifioidaan alakoulun CLIL-opetuksen arvioinnin periaatteita sekä ehdotetaan CLIL-opetukseen soveltuvia arviointimetoodeja.

Asiasanat: arviointi, oppilasarviointi, kielitaito, englannin kieli, vieraskielinen opetus, simulointi

Acknowledgements

I express my sincerest gratitude to all of those who have supported me in my aspirations in any way, shown interest or participated in this research process or simply been there for me.

Members of Academia

Professor emerita Annikki Koskensalo

Professor Tarja Nikula

Professor Ari Huhta

Adjunct Professor Tuula Merisuo-Storm

Adjunct Professor Kim Haataja

Doctor Nancy Commins

Doctor David Marsh

Suzanne Collins

University of Turku: Faculty of Education

Turku University Foundation

Finnish National Board of Education

Principals and teacher colleagues - prior and present

Research participants: CLIL teachers, pupils and their parents

PROFICOM crew and everyone who participated in the simulation production

My dearest ones

Irma and Hannu

Jaana

Mom Maire

Erik

Teppo

I dedicate this thesis to the memory of my father Aimo and little brother Ari.

Everything is possible. Accomplishing the impossible only takes a bit longer. (Unknown)

Contents

List of Figures	8
List of Tables	9
List of Abbreviations	10
1 INTRODUCTION.....	11
1.1 Purpose of the study	13
1.2 Theoretical framework.....	14
1.3 Organisation of the study	15
2 BILINGUAL CONTENT INSTRUCTION CLIL	18
2.1 CLIL in the field of bilingual education	20
2.2 Theoretical rationale for CLIL.....	27
2.3 CLIL in Finland	43
3 SECOND LANGUAGE PROFICIENCY	51
3.1 Models of second language proficiency.....	52
3.1.1 Common European Framework of Reference CEFR.....	54
3.2 Language proficiency in CLIL.....	58
3.2.1 Academic language	59
3.2.2 Classroom interaction	66
3.2.3 Language development	68
4 SECOND LANGUAGE ASSESSMENT	73
4.1 Assessment in Finnish basic education	74
4.1.1 Principles of classroom assessment	79
4.1.2 The 21 st century education and assessment.....	83
4.2 Alternative assessment	85
4.2.1 Collaborative testing	88
4.2.2 Task-based performance assessment	90
4.2.3 Technology-based assessment: computer simulations	94
4.3 Assessment in primary CLIL	98

5 RESEARCH METHODS	107
5.1 Research framework	107
5.2 Research questions.....	110
5.3 Participants.....	112
5.4 Research phases.....	114
5.4.1 CLIL assessment survey.....	115
5.4.1.1 Questionnaires	116
5.4.1.2 Interviews	118
5.4.2 Computer simulations	120
5.4.2.1 Implementation.....	121
5.4.2.2 Piloting.....	123
5.5 Data analysis	128
6 CLIL ASSESSMENT SURVEY.....	130
6.1 Assessment practices in CLIL.....	131
6.1.1 Provision of feedback.....	143
6.1.1.1 Frequency of provided feedback.....	146
6.2 Challenges in CLIL assessment.....	151
6.3 Adequacy of CLIL assessment.....	159
6.3.1 Received feedback.....	159
6.3.3 Frequency and sufficiency of received feedback.....	164
6.4 Preferences and future visions.....	168
7 COMPUTER SIMULATIONS.....	177
7.1 Pupils	177
7.2 Parents	192
8 DISCUSSION	199
8.1 Fundamentals and central issues in CLIL assessment	200
8.2 CLIL assessment practices	213
8.3. Affordances of computer simulations	223
8.4 Validity and significance	228
8.5 Conclusion and implications to CLIL	233
References	241
Appendices.....	258

List of Figures

FIGURE 1. Theoretical framework of the study	15
FIGURE 2. The continuum of content and language integration	26
FIGURE 3. The model of learning objectives	30
FIGURE 4. The Cummins model of language proficiency	31
FIGURE 5. The dual-iceberg representation of bilingual proficiency	32
FIGURE 6. The Conceptual Reservoir	33
FIGURE 7. The Input Hypothesis Model of L2 learning and production	34
FIGURE 8. An observation tool for effective CLIL instruction	40
FIGURE 9. Output, noticing and second language learning	42
FIGURE 10. The facets of CLIL instruction	48
FIGURE 11. A view of academic proficiency	59
FIGURE 12. Finnish assessment convention at classroom and school level: grades 1–7	78
FIGURE 13. A model of oral test performance.....	91
FIGURE 14. The alignment of standards, assessment and instruction of English language learners	103
FIGURE 15. Empirical research framework.....	108
FIGURE 16. Assessment scheme in PROFICOM simulations	121
FIGURE 17. An example of a task on the screen: organising text	122
FIGURE 18. Example of the simulation introduction PowerPoint slide	124
FIGURE 19. Pupils performing the simulation	124
FIGURE 20. Means of collecting assessment information on pupils' language proficiency.....	132
FIGURE 21. Means of providing assessment information to pupils on their level of language proficiency	144
FIGURE 22. Frequency of teacher-provided assessment information.....	148
FIGURE 23. Importance of receiving information on English acquisition in CLIL subjects.....	160
FIGURE 24. Importance of receiving information on English acquisition in CLIL subjects according to gender..	161
FIGURE 25. Forms of feedback identified by pupils.....	162
FIGURE 26. Means of receiving information on child's English proficiency and its progress	163
FIGURE 27. Frequency of feedback on English proficiency experienced by pupils	165
FIGURE 28. Sufficiency of information on child's English proficiency and its level in parents' opinion	166
FIGURE 29. Means of feedback preferred by pupils.....	169
FIGURE 30. Parents' wish to receive more information on the English proficiency of their child	171
FIGURE 31. Parents' preferred means of receiving information on their child's English proficiency level	173
FIGURE 32. Pupils' anxiety level in the beginning of simulation 1.....	178
FIGURE 33. Pupils' estimation of their coping in English in both simulations	180
FIGURE 34. Perceived linguistic difficulty of tasks in simulation 1	183
FIGURE 35. Perceived linguistic difficulty of tasks in simulation 2	184
FIGURE 36. Appropriateness of a computer simulation in assessing aspects of language (pupils).....	185
FIGURE 37. Appropriateness of a computer simulation in assessing content in English (pupils).....	186
FIGURE 38. Pupils' recollection of simulation topics studied through English at school.....	191
FIGURE 39. Parents' estimate of the CEFR proficiency level of their children.....	192
FIGURE 40. Appropriateness of a computer simulation in assessing aspects of language (parents)	194
FIGURE 41. Appropriateness of a computer simulation in assessing content in English (parents)	195
FIGURE 42. Desired alignment of curriculum, instruction and assessment in CLIL	199
FIGURE 43. CLIL English as the sum of its constituent parts.....	208

List of Tables

TABLE 1. Models of bilingualism.....	22
TABLE 2. Primary focus of instruction in the three models of CBI	23
TABLE 3. Contrasting forms of content and language integrating language programmes and regular EFL.....	27
TABLE 4. Contrasting language acquisition and language learning	35
TABLE 5. Distinctions between focus on meaning, focus on form and focus on forms	39
TABLE 6. Six-levelled CEFR scale for processing text	55
TABLE 7. The roles of language in CLIL instruction	58
TABLE 8. Comparison of linguistic features in ordinary and academic English	61
TABLE 9. The interface of academic content and language	65
TABLE 10. Grading of L1 and L2 linguistic objectives in passive and active bilingualism	69
TABLE 11. Language competencies favourably affected or unaffected by CLIL	71
TABLE 12. Scope and dimensions of classroom-based assessment	79
TABLE 13. Feedback types	82
TABLE 14. Types of task-based performance assessments with examples	93
TABLE 15. The influence of task characteristics on performance areas	94
TABLE 16. Test method characteristics and CALT advantages and limitations	95
TABLE 17. Contrastive table of participant school features	113
TABLE 18. Main features of the research phases	115
TABLE 19. Quantity of pupils answering the CLIL assessment questionnaire	117
TABLE 20. Statistics of parents' questionnaires in the CLIL assessment survey	118
TABLE 21. Participants in the simulation piloting sessions.....	123
TABLE 22. Frequency of teacher feedback to pupils on language proficiency or its progress	147
TABLE 23. Teachers' descriptions of feedback on pupils' language proficiency or progress to parents.....	148
TABLE 24. Parents' comments on sufficiency of assessment information	166
TABLE 25. Parents' remarks on sufficiency of assessment information on English progress.....	167
TABLE 26. Examples of parents' comments on additional assessment information.....	171
TABLE 27. Examples of parents' remarks on additional, more detailed assessment information	172
TABLE 28. Pupils' effort and endeavour in the simulation sessions.....	179
TABLE 29. Percentages of pupils' general and gender-specific coping in both simulations.....	181
TABLE 30. Pupils' comments on their coping in English in simulations.....	181

List of Abbreviations

A1–C2	CEFR scale descriptors from the lowest to highest (A1, A2, B1, B2, C1 and C2)
BICS	basic interpersonal communication skills, casual everyday language
CAF	complexity, accuracy and fluency of language performance
CALL	computer-assisted language learning
CALLA	Cognitive Academic Language Learning Approach
CALP	cognitive academic language proficiency
CALT	computer-assisted language testing
CBA	classroom-based assessment
CBA	content-based assessment
CBI	content-based instruction
CEFR	Common European Framework of Reference for languages
CLIL	content and language integrated learning, bilingual content instruction
CS	city school, the other research school
CUP	common underlying proficiency
DA	dynamic assessment
EAP	English for academic purposes
ECML	European Centre for Modern Languages
EFL	English as a foreign language
ESL	English as a second language
ELP	European Language Portfolio
EU	European Union
IALC	integrated assessment of language and content
ICT	information and communication technology
IRF	initiation–response–feedback discourse pattern
L	learner, student, pupil
L1	mother tongue(s), first language(s)
L2	second language(s), any language(s) acquired after the L1(s)
LAD	language acquisition device
MM	mixed (research) methods
MYP	Middle Years Programme
NCC	National Core Curriculum for Basic Education (grades 1-9)
NLP	natural language processing
PBA	performance-based assessment
P	parent
P.E.	physical education
R	researcher
SLA	second language acquisition
TBA	task-based assessment
TBPA	task-based performance assessment
TL	target language
TOEFL	Test of English as a Foreign Language
TS	town school, the other research school
UG	universal grammar
U.S.	United States of America
YLL	young language learner
ZPD	zone of proximal development

1 INTRODUCTION

Two for the price of one! This persuasive argument (see Bonnet 2012, 66) has expedited the propagation of bilingual content instruction, also known by the acronym CLIL (content and language integrated learning). CLIL is a fairly recent 'newcomer' in the field of bilingual second language acquisition (SLA) in Finland and in Europe, but around the time of the millennium, CLIL experienced such a boom in popularity that it has expanded to other continents. This global interest in CLIL resides in the efficacy of it as a language programme. The ultimate aim of CLIL is to intensify language acquisition by teaching and learning subject matter (content) through an additional language which is foreign to the learner.

In different parts of the world there are diverse labels for instruction that combine teaching of content and an additional language. CLIL shares roots with and bears resemblance to Canadian language immersion and the originally North-American content-based instruction, but regardless of many interfaces, the circumstances and emphases are not identical. CLIL-type language instruction aiming at additive, active bilingualism has wider significance and applicability to any educational context where children are being taught through a language that is not their mother tongue. Millions of children in the world, and thousands in Finland, learn subjects at school through an additional language, which is why CLIL perspectives and principles may help these children, who often have an immigrant background, in their learning.

This study at hand is situated in CLIL in which both the conventional language of instruction (Finnish) and foreign language (English) are used in teaching and learning. Instruction is thus bilingual. CLIL is primarily about language acquisition in various subject content learning contexts. Therefore, the language needed for learning different topics has to be subject-specific. This is an issue that has lately drawn growing interest, but needs more attention. Moreover, the question of implicit versus explicit language teaching is another issue to be clearly addressed. I will take a stand on these two matters in this research report. Language issues in CLIL, in other words the efficacy of CLIL, have been studied rather extensively, which is appropriate regarding the language emphasis. The research results are very favourable for CLIL as a language acquisition model (e.g. Marsh & Wolff 2007; Ruiz de Zarobe & Jiménez Catalan 2009). Also several comprehensive didactic handbooks (e.g. Bentley 2010; Coyle, Hood & Marsh 2010) have been generated with improved classroom pedagogics in mind.

The enthusiasm in instructional aspects is understandable from a practical point of view, but language education, as any kind of education, is based on three supporting elements which form the educational tripod on which education rests: 1) learning objectives, 2) instruction and 3) assessment. Consider a tripod without one leg: it is unable to stand at all. Even with a shorter, crooked or somehow defected leg, the tripod is dysfunctional. With one weak or missing educational element this education is not fully functional. Instruction and curricular aspects in CLIL research and literature have been covered more extensively than assessment issues. Therefore, this study is focussed on the assessment element in CLIL – an integral part of bilingual

content instruction (see also Llinares, Morton & Whittaker 2012, 280) and inseparable from the other two legs of the tripod.

Understanding the Finnish assessment context is relevant to understanding this study. Finland does not represent “assessment societies” (Broadfoot & Black 2004, 19) that rely on accountability policies in educational assessment, but it has rather adopted a ‘softer’, humanistic approach to assessment and education. Lately, however, especially PISA studies have drawn national attention to the efficacy of education and appropriateness of assessment. In Finland, assessment is not a vehicle of control, but a means to monitor and enhance learning. Following from this, the current Finnish National Core Curriculum for Basic Education (NCC 2004), which is the normative document framing and guiding the sketching process of local-level curricula (municipal, school-specific or programme-specific), mentions only continuous (formative) and final (summative) assessments of learning. Basic education refers to grades 1–9, pupils aged ca 7–12.

These normative curricular guidelines should then also be transferred to CLIL: the NCC posits inarguably that assessment in CLIL should give the teacher, pupils and their parents *adequate* information on the child’s progress in the target language in relation to the pre-set objectives. Does this happen? If it does, how? And what is adequate information? Adequate, according to a monolingual dictionary (MacMillan 2007), denotes “good enough for a particular purpose” or “satisfactory”. The stakeholders in assessment –, that is, teachers, pupils and their parents – are the most competent people to answer the question of adequate assessment investigated in this study.

Assessment schemes vary in Finland, because the NCC guidelines for CLIL are interpreted by local decision makers who have been invested with a considerable amount of freedom to apply national curricular guidelines in their own contexts. Following from this, there is no common model for bilingual content instruction either. Each education provider is allowed to decide the extent and model of CLIL implementation. Teachers can also decide the methods of continuous assessment they see fit to use within their own contexts, but formal reporting is normally regulated at the local level. Therefore, adequate assessment information in one CLL context may not be adequate in another.

There are global efforts to accelerate reformation in education so that it would better meet the requirements of the changing society, working life and technological advancements that have drastically shaped the ways we are producing, accessing and sharing information. This trend is inevitably also notable in the field of assessment. Finland is no exception to these developmental aspirations. In Finland, for instance, the first computerised baccalaureate exams will be administered in 2016. This is the same year that the renewed National Core Curriculum for Basic Education will be put into effect, which aims to modernise and update current educational practices – also including assessment.

Calls for modernised, alternative assessment as well as the need to document and analyse classroom-based assessment have been made by various influential scholars (e.g. Birenbaum & al. 2006, Cumming 2008, 5–6, 13). Similar demands have been made in respect to CLIL as well

(Asikainen & al. 2010, 4). The potential of technology-based assessment should be examined more closely in order to develop alternative, more informative ways to assess language proficiency. This is why this study, instead of being satisfied with merely describing prevalent assessment practices and gauging their adequacy, went deeper and also examined computer simulations as an alternative assessment method that might reveal more of the language proficiency of CLIL learners.

Although Finland was one of the first European countries launching CLIL, assessment in the Finnish context has not, so far, been investigated at all, and elsewhere it has barely been studied. The two existing studies I am aware of (Hönig 2010, Serragiotto 2007), as well as research or developmental projects (e.g. AECLIL 2013, CLILA 2013), have mainly focussed on secondary or tertiary students. There is no research on young learners' language assessment in primary CLIL, but scholarly texts are more commonplace although not abundant. As a primary CLIL teacher and language teacher (English and German), I became intrigued by the question of how assessment in CLIL is organised and what kinds of practices are implemented. The spark for assessment research was flared during my master's thesis in education pertaining to assessment in instruction without study books. I also saw the topic as utterly viable and far-reaching – one that potentially has an impact on practice as well as theory. My initial investigations of assessment in bilingual content instruction proved that there is indeed a marked research gap in this area (see also Pérez-Cañado 2012, 331).

1.1 Purpose of the study

My desire to find out how assessment of pupils' English language proficiency in bilingual content instruction was organised and administered at primary level for young language learners, if administered at all, thus laid the foundation for this research. McKay (2006, 65) differentiates four purposes for assessment research in young learners' language education:

- 1) to investigate and share information about current assessment practices,
- 2) to find ways to ensure valid and fair assessment tasks and procedures,
- 3) to find out more about the nature of young learner language proficiency and language growth and
- 4) to investigate and improve the impact of assessment on young language learners, their families, their teachers and their school.

The present study attempts to address purposes 1, 2 and 4 explicitly. The core areas of this study are current CLIL assessment practices as perceived by the stakeholders of assessment (teachers, pupils and their parents) as well as the usage of computer simulations in CLIL assessment. One substantial aim of this study was to create an overall description of assessment of pupils' English language proficiency in Finnish primary CLIL and to see how assessment in CLIL is organised. Another aim was to investigate whether these practices are 'adequate' as

stipulated by the NCC. Since CLIL provision in Finland varies remarkably, and there is no predefined description of the adequacy of assessment information, this was a challenging undertaking.

In order to achieve a valid description, the topic was triangulated from different stakeholder angles using mixed research methods that, in producing more representative research results, allow for a more diversified and multifaceted account. The overall description was achieved by 1) elucidating which assessment practices teachers use, 2) investigating 'adequate assessment' by asking the stakeholders (pupils and their parents) about the frequency and sufficiency of given and received assessment information, 3) finding out the challenges of and obstacles to assessment in CLIL and 4) outlining opinions, preferences and visions concerning improved future assessment in CLIL. My attempt was also to provide some modest, yet critical considerations and perspectives for the establishment of a theory base for assessment in CLIL.

Furthermore, to answer the call for modernised assessment methods, to orient towards future assessments and to explore the prospects of technology-based assessment in CLIL, LangPerform computer simulations (see Haataja 2010) were experimented with in profiling CLIL pupils' English language proficiency. Experimenting with computer simulations was an attempt to establish ways of evaluating and reporting on the language performances of primary CLIL pupils and examine the appropriateness of computer simulations as an alternative and additional assessment method in primary tuition. To avoid producing a technological innovation research report that makes a mere "I tried it and I liked it claim" (Chin, Dukes and Gamson 2009, 554), I am providing ample assessment evidence to both support and challenge the use of computer simulation as an alternative assessment tool in CLIL contexts.

1.2 Theoretical framework

This study at hand represents the discipline of applied linguistics, because 1) it involves a substantial language element; 2) it is based on theories from second language acquisition (SLA); 3) it is interdisciplinary in connecting various areas of applied linguistics: bilingual instruction, foreign/second language acquisition, language assessment as well as computer-assisted language learning (CALL); and 4) it seeks to investigate a practical problem from theoretical and empirical perspectives (see Cook & Wei 2009). The theoretical foundation comprises a research and literature review of the three most relevant SLA concepts forming the theoretical framework of this study, as portrayed in Figure 1: CLIL, second language proficiency and second language assessment.

Although the circles in Figure 1 are depicted as being of equal size, the assessment aspect carries more weight than second language proficiency and CLIL. An individual chapter has been devoted for each of the main concepts. The central pivot of the Venn diagram is where this research of language assessment of primary CLIL pupils' English proficiency resides.

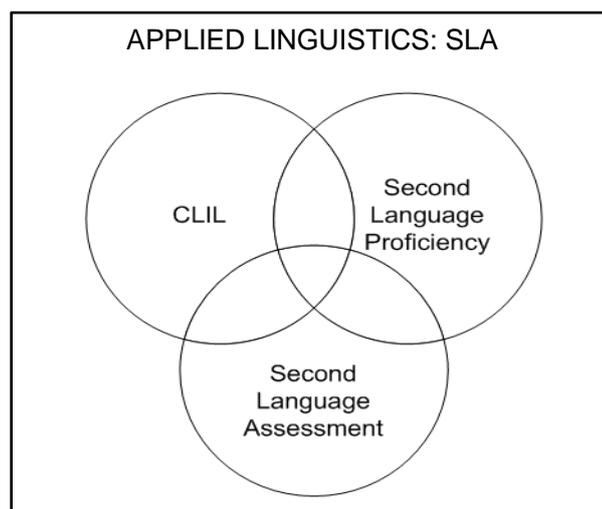


FIGURE 1. Theoretical framework of the study

In my attempt to define the domain of English language assessment in primary CLIL, I have heeded Cumming's (2008, 3) notions of aspects that are involved in defining any domain of language assessments:

- conceptualising what language is and is not in CLIL
- demarcating the purposes and scope of the language assessment in CLIL
- specifying relevant components and contexts of language use and knowledge in CLIL
- analysing empirically how people perform in such CLIL contexts
- establishing ways of evaluating and reporting
 - on these linguistic performances
 - on the effectiveness of the assessment instrument and procedures.

The first three points in respect to this particular study will be discussed in the theoretical foundation, and the final two points are dealt with in the computer simulation experiment section.

1.3 Organisation of the study

The study is organised into two main parts, the theoretical foundation (chapters 2–4) and empirical implementation including results, discussion and conclusion (chapters 5–8). Each theory chapter starts with a definition of the main concept and each section ends with a concise summary of the main content. This research report contains a relatively large number of figures and tables because every decent CLIL teacher is familiar with the power of visuals. These are listed, along with the abbreviations used in this report, after the contents on pages 8–10.

While chapter 1 works as an introduction to the research and helps to perceive the entity of the study, chapter 2 pertains to bilingual content instruction CLIL, the educational frame within which this research is placed. Chapter 2 thus specifies the context of the language use which will be assessed. I will situate CLIL within the field of bilingual instruction which is important, because

CLIL is not automatically identical to other forms of bilingual instruction with which it is often mistakenly associated. This is also why research results derived from other bilingual programmes are not as such always applicable in CLIL. Furthermore, chapter 2 looks into the theoretical tenets of CLIL which lend justification to instructional emphases and preference to some forms of assessment over others. This chapter also gives an overview of CLIL in Finland.

Chapter 3 centres on the characteristics of language in CLIL. It defines second language proficiency which is the target of assessment in this study and specifies the relevant components of language use and knowledge in CLIL. Language proficiency is seen as an ability to appropriately use language in specific situations. Out of many models of communicative second language proficiency, I have chosen the most relevant, the Common European Framework of Reference for Languages (CEFR 2001), for closer presentation. Since subject-specific language use is marked in CLIL, I will also address academic language needed in subject study as well as provide a literature review of notable studies on second language development in CLIL. I have delimited the theoretical basis of the current study so that second language learning is not covered, although it is very relevant for the comprehensive understanding of the CLIL learner language development. An excellent overview of second language learning for the interested reader is, for instance, provided by Dörnyei (2009) or Ortega (2009).

Chapter 4 demarcates the purposes and scope of assessment in CLIL and explores the features of CLIL-appropriate assessment. It thus sets the theoretical frame for assessment of young learners' English proficiency in bilingual content instruction. Assessment is approached from the Finnish context in order to give the reader a better understanding of how the Finnish assessment system deviates from, for example, mainstream accountability assessments, and why assessment in CLIL in this study is mainly confined to formative assessment. Also future trends and requirements in assessment will be addressed to frame the computer simulation experiments which is why technology-based language testing, in particular computer simulations, are also scrutinised along with other two alternative assessments (collaborative assessment and task-based performance assessment). I have chosen to submit these three alternative assessments due to their potential for CLIL assessment. Research in the field of CLIL assessment (a term I will use interchangeably for assessment in CLIL) is scarce; thus the main topic of assessment in CLIL is primarily tackled theoretically.

The empirical implementation of the research is reported in chapter 5 in which the empirical framework is also presented along with precise research questions. Chapter 5 gives a detailed account of how the three phases of the study (the CLIL assessment survey and two computer simulation experimentations) were conducted, when and who participated in them. The data analysis methods are also explained. The given questionnaires and materials connected to the data collection can be found in the Appendices section in their original language (Finnish).

The subsequent chapters 6 and 7 disentangle the results obtained in the three research phases. The results are expounded thematically instead of as disassembled outcomes, and participants' voices are markedly present in the form of quotations. Chapter 6 entails the CLIL

assessment survey findings answering the questions of how assessment of children's English language proficiency is conducted in Finnish CLIL classrooms and whether the pupils and parents in the two sample schools find it adequate. Also, the challenges impinging upon the administration of CLIL assessment as well as future visions are unravelled in this chapter. The findings of the two separate simulation experiments are, when possible, combined and presented in chapter 7 from the viewpoints of pupils and parents, respectively.

In chapter 8, I will discuss the obtained core findings and compare the outcomes to prior research. Because CLIL assessment research is still in its infancy and barely existing, I have relied, when applicable, on recent research in CLIL, SLA and second language assessment. The unique computer simulation experiments without any prior reference are gauged with the help of Chapelle's (2001) criteria for appropriate computer-assisted language learning (CALL) which I have modified to work as criteria for appropriate computer-assisted language assessment.

Chapter 8 is also critically concerned with the quality of this research; I will appraise the validity and significance of the study, the methods used and results obtained. I will reflect on the significance of the research results for CLIL implementations, instruction, teacher qualities and naturally assessment practices. The final section of the Discussion concludes this research report by disentangling the prerequisites for potentially successful assessment in CLIL in the form of an articulate list which is comprised of five sections: 1) fundamentals and central issues in CLIL assessment, 2) adequacy of CLIL assessment, 3) affordances of computer simulations in CLIL assessment, 4) validity and significance of the study and 5) conclusion in the form of recommended CLIL assessment practices and methods. In this chapter, I will also highlight a few potential areas for further research and encapsulate the research report with a dense *code of assessment practices for CLIL class teachers*.

2 BILINGUAL CONTENT INSTRUCTION CLIL

Contemporary *second language acquisition* (SLA) programmes can roughly be categorised into two mainstreams according to the study purpose (general or specific) and the number of languages used (monolingual or bilingual). Further specifications can be made on the grounds of the status of the language (target or tool), the role of the learner (student or user) and the focus of instruction (meaning, form or forms). This study is conducted within the context of *content and language integrated learning*, widely known by its acronym CLIL. CLIL is a specific purpose bilingual programme in which the target language (TL) is the tool of teaching and learning disciplinary content matter, but the TL is also the target of learning. The learner thus mainly acquires the TL while using it for studying, but the language can also be studied; thus the focus is mainly on meaning but also on form.

I have generated the following definition of CLIL by incorporating aspects from the classical CLIL definition (e.g. Maljers, Marsh & Wolff 2007, 8; Coyle, Hood & Marsh 2010, 1) and a slightly fine-tuned definition of Mehisto and Lucietto (2010):

CLIL is a dual-focussed teaching and learning approach in which the main language of schooling and an additional language or two are used for promoting both content mastery and language acquisition to pre-defined levels.

This definition emphasises that the role of language in CLIL is two-fold: it is the medium of study but also the end result of it. Additionally, the goal-orientation of the language study is clearly present and the bilingual nature of instruction is stated. It also recognises, unlike prior definitions, that the linguistic backgrounds of the CLIL learners may vary – the language of schooling is not automatically the mother tongue (L1, home language) of all learners; indeed, there may be many different L1s in the classroom. The additional TL is not necessarily the second language for the learner, who may have competences in several languages prior to the CLIL study. This often applies to multicultural environments. If the main language of schooling is not the mother tongue of the learners, they should receive additional support for the development of their mother tongue – a principle that is recognised by the classic definition of bilingual education in Andersson and Boyer (1978, 16–17) and supported by the theoretical underpinnings of CLIL (see section 2.2).

CLIL is translated into various models or programmes in many different ways. For example, the balance between the content and language shifts, the starting age may change, the number of subjects in CLIL implementations varies and even the terminology used is subject to some fluctuation. Most often CLIL is cautiously said to be an approach, but increasingly and in diverse connections, CLIL is alternately also referred to as a teaching methodology, model or programme (Järvinen 1999), method (e.g. Mehisto, Marsh & Frigols 2008, 12; Bentley 2010, 5), methodological approach (Coyle, Hood & Marsh 2010, 1), form of instruction (e.g. Nikula & Järvinen 2013, 144) and “system rather than a method because it essentially defies the principles

and criteria accepted in designing mainstream methods of foreign language teaching” (Dakowska 2013). CLIL might even be called a philosophy, because there are certain principles that are agreed upon, a shared CLIL worldview, although some discrepancies also exist.¹ The ultimate reason to implement CLIL is to cater for improved, communicatively functional, bilingual proficiency through non-linguistic content study which is why it is also a “Sneaky Way” of teaching languages (Bot 2007, 276). Functionality as an attribute of language proficiency refers to the ability to use the language purposefully in a specific situation as opposed to accurate, grammatically correct language use.

Snow, Met and Genesee (1989) have named several factors that advocate for integration of language and content learning instead of treating them as individual subjects. First, language development and general cognitive development are naturally intertwined in the first language acquisition; children learn things and facts about the real world through language. Second, they highlight that language is normally used in meaningful, social interactions and academic contexts to exchange information on real world issues rather than language per se; people tend to talk about things they know and have encountered. This means that language learning is purposeful – language is needed and used for a real reason, for negotiating and learning phenomena of the world instead of practising structures and communicative phrases, as in the formal, mainstream English instruction, referred to as EFL, *English as a foreign language*, in this study.

Third, the content-basis provides a motivational and cognitive springboard for language learning because “content is interesting and of some value to the learner and therefore worth learning” (ibid., 202). Hence, content adds value to the language learning, and language provides access to the desired information. Fourth, the variety of academic language is substantially different from the language used outside the school; it has to be mastered in order to succeed in higher educational levels. Furthermore, Snow, Met and Genesee (1989, 203) remark that the mastery of subject-specific vocabulary may be “a prerequisite to mastery of specific content or to academic development in general”. In order to study specific content, such as photosynthesis, topic-related vocabulary is essential and contributes to academic language proficiency. Following from this, academic language proficiency has a substantial role in CLIL and therefore also in the theoretical foundation of this research.

CLIL represents the currently prominent educational framework of socio-cultural constructivism which highlights the social interaction between the learners, learning through negotiations of meaning (Cook 1997, 224) and mutual construction of knowledge. It thus sees the learner as an active language user instead of passive recipient of input. Furthermore, according to the current perceptions, the goal in SLA in general and also in CLIL is not to become a near native-speaker: second language (L2) users should “be viewed as multicompetent language users rather than deficient native speakers” (Cook 1999, 185). In CLIL, the focus is primarily on

¹ See e.g. Gierlinger’s blog post discussion about CLIL models and designations <http://clilingmesoftly.wordpress.com/clil-models-3/>.

meaning (content), but as it will be shown in section 2.2, a paradigm shift from purely implicit language acquisition towards a focus on form (more explicit or accurate language instruction and learning) is grounded on research. Research has also mostly shown ample evidence of the benefits of CLIL which will be particularly addressed in the next chapter. As a whole, there is strong evidence that CLIL promotes better language proficiency and tolerance of incomplete language performance (Nikula & Järvinen 2013) and it enhances acceptance towards people representing other cultures and languages (García 2009).

The following section investigates the emergence of CLIL in the field of bilingual education and contrasted to the two partially reminiscent forms of bilingual education, content-based instruction (CBI) and immersion. It also provides some comparison of CLIL to EFL in which the language is taught as a subject. It is also necessary to examine the theoretical background of CLIL to better understand the theoretical underpinnings for assessment in CLIL (2.2), after which I will take a closer look at CLIL in Finland (2.3). The purpose of this chapter, therefore, is to give an overview of CLIL, relate it to other forms of bilingual education and provide a frame for the following theoretical discussion concerning the TL aspect in CLIL and its assessment.

2.1 CLIL in the field of bilingual education

It is important to understand how CLIL differs from and is similar to other bilingual, content and language integrating approaches to second language acquisition in order to better interpret research and literature in this realm. *Bilingual education is one in which two languages are used as the medium of instruction in any part of or across the curriculum*, and it stresses the importance of leaning to the learners' mother tongue as a resource (Andersson & Boyer 1978, 16–18). Bilingual education has experienced a renaissance over the past 50 years (e.g. Andersson & Boyer 1978; García 1996); especially Europe has witnessed a bilingual boom during the last quarter century. This is due to the general potential benefits of bilingual education that to some extent also apply to CLIL (Baker 2011, 249–250):

- The attained language proficiency in two (or more) languages is typically high.
- Bilingual education enhances enculturation in which the main traits of the target culture are learned and embraced.
- Strong, additive forms of bilingual education often lead to biliteracy.
- Student achievement is higher.
- Cognitive benefits are acquired.
- Learner self-perception is raised.
- Bilingual education can establish a more firm identity.
- Mastery of two (or more) languages brings economic benefits and better employment.
- Societal, ethnic group and community benefits are accrued.

Bilingual education, however, is by no means a new innovation, as the discussion in Genesee (1987, 1–2) and in Dalton-Puffer (2007a, 1–2) remarks: it dates back at least a couple of millennia predating even the Greek and Roman Empires where bilingual education was already a common practice. Bilingual education, in Baker’s (2011, 207) words, is “a simplistic label for a complex phenomenon”. This statement is based on the fact that it appears in so many forms and approaches. As Brisk (2010, 20) puts it, “[b]ilingual education programs have in common the use of two or more languages for instruction, but that is where the similarities end”. The reasons for the diversity of bilingual education and programmes are mostly political, but sometimes related to cost-efficiency, economics and pedagogy (Baker 2011, 208). The reasons for CLIL implementation are mostly empowerment of communication with people in the outside world and promotion of deeper understanding of language and culture (see the list by Ferguson, Houghton and Wells in Genesee 1987, 4 or Baker 2011, 208).

In bilingual education, there is variation and controversy, for example, in the following issues (Baker 2011; Brisk 2010, 17–20; García 2009):

- target groups (majorities or minorities),
- language goals (monolingualism, bilingualism, biliteracy or multilingualism),
- sociocultural integration (maintenance of the heritage culture),
- language use in the curriculum (language of literacy introduction, ratio of languages used in content instruction, separation of languages during instruction and scaffolding of mother tongue),
- incorporating culture (cultural assimilation or enrichment),
- measuring academic achievement (methods and language/s used),
- assessment of language proficiency (high stakes testing vs. formative practices) and
- language proficiency of teachers (native vs. non-native teachers, language fluency).

Bilingual programmes thus have different orientations and can therefore be broadly distinguished into three main categories according to the incorporation of the aspects mentioned above. Baker (2011, 208–219) in his typology avails himself of three categories of bilingual programmes: 1) monolingual forms of education for bilinguals aiming at monolingualism, 2) weak forms which strive for either relative monolingualism or limited bilingualism and 3) strong forms of education aspiring to bilingualism and biliteracy. There is also some fluctuation in classifications and labels of bilingualism. For example, García (2009a, 51–56) differentiates between four models of bilingualism: additive, subtractive, recursive and dynamic, densely presented in Table 1, in which the polarities subtractive and dynamic bilingualism represent totally different attitudes of and views on languages.

Subtractive bilingualism aims at suffocating the L1 of, for instance, an immigrant person, whereas additive bilingualism accepts the L1 as equal to L2 (García 2009b, 130–131). Recursive bilingualism refers to situational bilingualism in which the L1 becomes repeatedly rediscovered and used for new purposes; dynamic bilingualism denotes the “varying degrees of abilities and

uses of multiple language practices needed for people” to be linguistic, social agents in distinct areas of language use (ibid.). Dynamic bilinguals are able to use their languages, two or more, purposefully and functionally. These broad categories of both Baker and García can be further divided in subcategories (see Baker 2011; Brisk 2010, 8–11; García 1996, 2009a and Siegel 2003 for categorisations and further sub-models).

TABLE 1. Models of bilingualism (García 2009a, 51-55)

Subtractive	Additive	Recursive	Dynamic
moves towards monolingualism	attempts balanced bilingualism	accepts the flows of bilingualism	encourages communicative and dynamic bilingualism
$L1 + L2 - L1 = L2$	$L1 + L2 = L1 + L2$	language is reconstituted in new functions	plurilingualism in the global context, drawing from different sources

CLIL represents an additive, strong form of bilingual education aiming at dynamic bilingualism (see e.g. García 2009b, 136). Bilingual programmes are thus delivered in many ways for numerous purposes which is why there is, as García (2009a, 208) claims, more than thirty different designations for them. The term content and language integrated learning CLIL was coined in the mid-1990s by a group of experts in Europe (Marsh 2009, vii) to serve as a superordinate, “umbrella” term for “a dozen or more educational approaches” (Mehisto, Marsh & Frigols 2008, 12). Depending on the perspective, however, CLIL can also be considered as a hyponym instead of a hypernym to bilingual education, and therefore as *one form of bilingual instruction* (e.g. Baker 2011; Eurydice 2006, 8; García 2009a). CLIL will be perceived as such in this document as well: it is one form of bilingual instruction among many others. I will justify this perception in the following paragraphs that diversify CLIL in comparison to content-based instruction (CBI) and immersion, its main equivalents.

CLIL compared to CBI

CLIL as a term is widely used in Europe, although the terms vary according to the given language due to the words forming the abbreviation: CLIL in Spanish is AICLE, SPRINT in Swedish, EMILE in French, and in German-speaking countries the term Fremdsprachiger Sachfachunterricht is often used. The designation of CLIL has spread to Latin America, Asia and even Australia. In the United States (U.S.), the equivalent name commonly used for integration of non-linguistic content and language is *content-based instruction*, CBI (Lyster 2011, 611), which started to emerge in the 1980s (Stoller 2008; Stryker & Leaver 1997a, 285). CBI “implies the total integration of language learning and content learning” and can simultaneously articulate “a philosophical orientation, a methodological system, a syllabus design for a single course, or a framework for an entire program of instruction” (Stryker & Leaver 1997b, 5) – as in CLIL.

The diverse programmes integrating content and language all share the premise that language is efficiently learned through content study. However, as Weigle and Jensen (1997, 201) posit, there are differences in the primary focus and weight they attach to either language or

content mastery. They differentiate three models of CBI: theme-based, sheltered and adjunct, which all have different primary foci as demonstrated in Table 2.

TABLE 2. Primary focus of instruction in the three models of CBI (Weigle & Jensen 1997, 202)

	Language	Content
Theme-based	x	
Sheltered		x
Adjunct	x	x

The *theme-based CBI* model draws from specific topics and themes within the subjects that are taught in the target language, whereas *sheltered content instruction*, the most common CBI type in primary and secondary levels in the U.S., holds language as a secondary goal and aims at conveying the core skills and issues (Weigle & Jensen 1997, 201-202). The *adjunct CBI* model combines both content instruction intended for native speakers and the instruction of academic language skills for non-native students at the higher levels of education (ibid.). CLIL with low or medium exposure to language resembles theme-based instruction.

Moreover, the distinction between English as a foreign language (EFL) and English as a second language (ESL) makes the difference between CLIL and CBI even more noticeable. For example, McKay (2006) separates content-based *foreign* language instruction (CLIL) and content-based *second* language instruction (CBI) – this distinction is essential because bilingual education in the United States often has another bearing than that in Europe. In this study, however, I will not make any distinction between foreign and second language; these terms will be perceived as synonymous.

In Europe, bilingual programmes aim at instilling “high proficiency levels in both languages, even if one of them is not the official language of that country”, i.e. the goal is to learn a foreign language. In the U.S., bilingual education mainly refers to transitional programmes that educate minority children and attempt to assimilate them into the mainstream society and its language through schooling (Bialystok 2001); the goal is to learn a second language. Sheltered English CBI in the U.S. thus often represents subtractive bilingualism. CLIL is principally intended for majority-language speakers, whereas sheltered CBI is meant for minorities with monolingual, subtractive intention (García 2009a, 186). It follows, then, that when drawing from CBI research and literature in relation to CLIL, it is essential to keep in mind the fundamental distinction between the subtractive and additive forms of CBI because the contexts and aims are not automatically comparable to CLIL.

CBI has attracted attention outside the U.S. in the same manner as CLIL has spread to other continents. It is noteworthy that CBI in other than subtractive contexts can be perceived as a form of additive bilingualism aiming at similar goals as CLIL, as the definition by Stoller (2008, 59) indicates: “Content-based instruction is an umbrella term referring to instructional approaches that make dual, though not necessarily equal, commitment to language and content-learning objectives”. Her definition closely resembles the CLIL definition of Mehisto, Marsh and Frigols (2008, 12). In this sense, the terms CLIL and CBI can be perceived as ‘continental rivals’. CLIL

and CBI in its additive form are merely two sides of the same coin (see also Lyster 2011, 611). In this vein, it is not surprising that CLIL and CBI share the same theoretical rationale that has lent support to both innovations. The theoretical rationale underpinning both CLIL and CBI but also immersion programmes is discussed in section 2.2.

Origins of CLIL

The origins of both CBI and CLIL can be traced back to language immersion. The most prominent immersion programmes are the *Canadian immersion* models, although similar experiments in primary and secondary education were also conducted in the former Soviet Union from the 1960s to the mid-1980s (Stryker & Leaver 1997b, 15). The widely reported and influential Canadian model arose in the province of Quebec as a result of an initiative taken by a group of native English parents in 1965 who wished to fill the linguistic and cultural gap between English-speaking and French-speaking Quebecers by insisting on higher quality instruction in French for their children (Andersson & Boyer 1978, 185; Genesee 1987, 4). This is why the Canadian immersion is occasionally also referred to as French immersion.

The practice of teaching monolingual English children in French spread throughout the country during the next two decades for various reasons: 1) the social, economic and political benefits of bilingualism became recognised and more salient; 2) the government supported and funded the establishment of bilingual programmes; and 3) media and research reports disseminated the successful outcomes of the immersion programmes (Johnson & Swain 1997, 3). The Canadian success expedited the further expansion of immersion to the United States, Australia and other parts of the world such as Finland (Johnson & Swain 1997, 4), where the Swedish immersion programme started in the late 1980s (Björklund 1997, 85).

Immersion programmes can be executed in several ways according to the extent (*total and partial immersion*) and the age of start (*early immersion* starting from the preschool or the first grade, *delayed immersion* starting from the 4th or 5th grade and *late immersion* postponed to the end of primary or the beginning of secondary school) which can be combined into various versions such as *early total immersion* signifying 100% instruction in the foreign language, with literacy training included (Genesee 1987, 19–21). International schools typically represent such early total or near-total types of immersion. Other forms of immersion include heritage language immersion revitalising indigenous languages, double immersion incorporating two foreign languages, and activity-centred immersion which appears to be equivalent to task-based learning (ibid.). A common factor is that the students receive “part of their instruction through the medium of a second language and part through their first language” (ibid., 1).

The quite extensive body of the Canadian immersion research with several longitudinal evaluations of immersion programmes suggests that students make significant gains in linguistic, academic and cognitive realms (Lazaruk 2007). The impact of immersion on second language proficiency in particular has been the central pivot in the immersion research indicating the

efficacy of the programme. The research has generally showed, according to Genesee (1987, 1991), the following outcomes:

- No significant delay occurs in the development of mother tongue.
- Academic achievement is not hampered.
- Functional language proficiency is markedly better.
- Native-like levels are not reached nor pursued.
- Early total immersion is more efficient than other immersion alternatives.
- The more exposure to language and the earlier the immersion starts, the higher the second language proficiency.
- Older students outperform younger ones in the speed of acquisition.
- With older students, the intensity of the immersion study may compensate for the gradual accumulation of language exposure.
- The continuation from one level to another (e.g. from primary to secondary) within the programme needs to be considered.

These results obtained during the two first decades of French immersion programmes have been reinforced by more contemporary studies (e.g. Turnbull, Hart & Lapkin 2003) which also introduced new elements such as the programme of intensive French immersion and the achievement of immigrant children, which was found to be comparable to their Anglophone classmates (Lazaruk 2007). Recent studies of bilingualism and immersion interested in metalinguistic awareness, sociolinguistic abilities, cognitive benefits and thinking processes of bilinguals have concluded that bilinguals 1) are linguistically more analytical, and able to direct their attentional focus; 2) distinguish sophisticated differences in vocabulary; 3) exhibit sensitivity to communicative cues; 4) are able to adopt two viewpoints leading to different linguistic processes, the complexity of which will be affected by the balance of bilingualism; and 5) enjoy socio-economic benefits due to the high appreciation of bilingualism (Lazaruk 2007).

These encouraging conclusions naturally accelerated the propagation of immersion programs resulting in the emergence of CBI and CLIL. CLIL has, in fact, especially in Spanish CLIL literature, been termed as “semi-immersion” (e.g. Pérez-Vidal 2007) which as such is not a very descriptive term because there are quite a number of diverse implementations of immersion. In relation to CLIL, the excellent immersion research results are problematic: similar results are expected of CLIL as well, although the language exposure is often considerably narrower.

Met (1999) depicts this variation in language exposure as a line of continuum which illustrates the positions of various content and language integrating models both in relation to each other and according to the extent of language and content matter, as in Figure 2. The place of CLIL in the continuum is difficult to determine due to its many implementations. Its location in the continuum could be, depending on the language emphasis, anywhere between the two poles. CLIL and immersion seem to be, at least in contrastive studies examining various aspects of these types of language programmes most often contrasted with regular mainstream EFL, which

represents the general purpose of language instruction and typically displays communicative aims in language use and promotes accuracy.

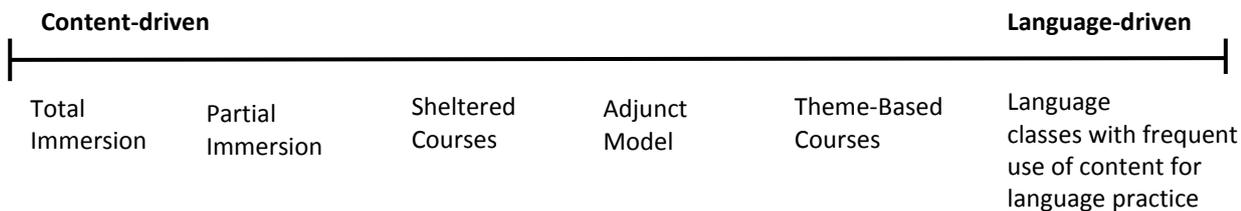


FIGURE 2. The continuum of content and language integration (Met 1999)

To conclude this chapter, below I present Table 3 based on García (2009a), Lasagabaster and Sierra (2010), Somers and Surmont (2012), who commented and criticised the article of Lasagabaster and Sierra, as well as Swain and Lapkin (2005) who revised the prominent immersion characteristics drafted by Swain and Johnson (1997). The table attempts to summarise the core content of this section. It is noteworthy, however, that the pieces of information given in the table are generalisations and therefore models deviating from these archetypes will exist.

SUMMARY

CLIL, content and language integrated learning, has its origins in immersion and content-based instruction CBI. As a phenomenon it is not one-dimensional and easy to grasp, because it has, like its precursors, many appearances. CLIL can thus be regarded as a generic term for various forms of bilingual education or as an educational programme combining both the study of a foreign, additional language and disciplinary content. CLIL, from the standpoint of this study, is treated as one form of bilingual education which is perceived as a superordinate term for CLIL. The term CLIL is mainly used in the European context whereas CBI is related to the North-American settings where it has many realisations that are different from CLIL. For instance, CBI in its sheltered form is subtractive, aiming at assimilating speakers of other languages to the mainstream monolingual education.

Immersion is linked with Canada, where it started in the 1960s as an experiment with the objective to educate minority English-speakers to able French-English bilinguals in French schools. Both CLIL and immersion represent additive forms of bilingual education, although CBI may also promote additive bilingualism. The positive research findings in immersion have created high expectations for the efficiency of CLIL. The three main content and language integrating approaches thus, basically, have similar goals, although the balance between language and content, among other aspects, may vary significantly. Section 2.3 looks into the theoretical tenets that all these three approaches share to some extent.

TABLE 3. Contrasting forms of content and language integrating language programmes and regular EFL

LANGUAGE PROGRAMME	Immersion Programmes	Content-Based Instruction CBI	Content and Language Integrated Learning CLIL	Regular Language Education (EFL)
Linguistic target group	either minorities or majorities depending on the context	subtractive: minorities OR additive: mainly majorities	mainly majorities without exclusion of minorities	all groups
Target level	all levels	all levels	all levels	all levels
Area of prominence	predominantly Canada but also elsewhere	predominantly U.S. but also elsewhere	predominantly Europe but also elsewhere	global
Overarching goal	additive: meaningful education and bilingualism	subtractive: sheltered education, transit education OR additive: meaningful education and some type of bilingualism	additive: cognitively challenging, meaningful education and some type of bilingualism	additive: communicative competence in an additional language
Academic goal	academic achievement, high second language proficiency and increased cultural awareness	acquisition of academic language and content learning; linguistic and cultural assimilation to majority OR higher language proficiency	higher functional proficiency in a second prestige language, academic achievement and familiarisation with the additional culture	learning an additional language and becoming familiar with an additional culture
Language use	language(s) used as medium of instruction	language used as medium and target of instruction	language(s) used as medium and target of instruction	additional language taught as subject
Instructional use of language	uses target language mostly	uses target language mostly	uses two or more languages	uses target language mostly
Support for mother tongue	overt support, instruction also in mother tongue	some support for mother tongue	overt support, instruction also in mother tongue	support not relevant
Pedagogical emphasis	integration of language and content, fluency	integration of language and content, functional fluency	integration of language and content, functional fluency	explicit language instruction, accuracy

2.2 Theoretical rationale for CLIL

In order to fully understand CLIL and to orient towards language assessment in CLIL, it is important to be familiar with the theoretical rationale underpinning it. Immersion research and theoretical assumptions directly or implicitly connected to it have influenced the construction of the theoretical rationale for other models of bilingual education such as CLIL and CBI which draw – or rather used to draw – on the same research base. In many cases, the educational baseline

in immersion and CBI is different to that in CLIL, and therefore, the theoretical rationale for CLIL in this study has to be interpreted through the lens of the European and specifically Finnish context.

The current foreign language education can be depicted, according to Harjanne and Tella (2008, 56), as “socio-culturally oriented, communicative and transcultural”, so also CLIL. The theoretical basis of CLIL derives from the field of SLA, the Vygotskian sociocultural theory and especially the theory of communicative competence, educational and cognitive psychology, and studies within bilingual education (see Coyle, Hood & Marsh 2010 27–47; Dalton-Puffer 2007a, 193–196; Dalton-Puffer, Nikula & Smit 2010a, 6–8; Järvinen 2007, 254; Stoller 2008, 60–61). Llinares, Morton and Whittaker (2012, 13) add to the above mentioned perspectives Halliday’s systemic functional linguistics, which emphasises language as the conveyor of meaning.

The most basic tenets for CLIL that are underpinning or have shaped it are listed in the following with the supporting theories in parenthesis:

- 1) **language and knowledge is co-constructed in social interaction and with the help of others** (Vygotski’s Zone of Proximal Development),
- 2) **the linguistic and cognitive level of difficulty gradually increases** (Bloom’s taxonomy),
- 3) **CLIL represents academic-type of language rather than casual language** (Cummins’ distinction of BICS and CALP),
- 4) **the mother tongue and prior languages have a bearing on learning of succeeding languages** (Cummins’ Common Underlying Proficiency and Conceptual Reservoir of Miramontes, Nadeau & Commis),
- 5) **language acquisition occurs better with rich input in relaxed atmosphere** (Krashen’s SLA theory, known as the Input hypothesis or the Monitor model),
- 6) **linguistic output is equally important** (Swain’s Output hypothesis),
- 7) **language learning is both implicit and explicit** (Krashen’s SLA theory, the model of second language learning by Bialystok, focus on form approach and Schmidt’s Noticing hypothesis), and
- 8) **language is used for a meaningful purpose** (Long’s (1996) Interaction hypothesis).

Especially early CLIL has also been informed by Krashen and Terrel’s (1983) Natural Approach to language acquisition. Due to limitations in space, I will not describe all theoretical rationale for CLIL here, and the ones I do will be treated concisely with references to CLIL and assessment.

Vygotski’s Zone of Proximal Development

Vygotski’s *Zone of Proximal Development* (ZPD), generally influential in education, is leaning on social-constructivist views on learning. Vygotski (1978, 86) himself defines ZPD as follows:

It is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or collaboration with more capable peers.

Vygotski (ibid.) distinguishes at least two developmental levels: the already completed actual developmental level and the potential development accomplished when responding to externally mediated assistance of peers or adults. The ZPD, then, is a “space” where students “perform beyond their current abilities” (Lantolf 2000, 13) with external help. The ZPD is especially empowering in collaborative work which is one of the basic working methods in CLIL; students are negotiating of meaning (i.e. discussing or working with the content through the TL) either with their classmates, the teacher or other TL users and thereby also learning language from each other.

This also applies to assessment. Assessment in the ZPD is termed *Dynamic Assessment* (DA). The intention of DA is to simultaneously make diagnoses of skills and to provide a specific form of support, mediation, which is “to bring to light underlying problems and help learners to overcome them” (Lantolf & Poehner 2008, 273). Such mediations used during the assessment process are, for example, “performance prompts, hints, leading questions etc.” (Lantolf & Poehner 2004, 49). In other words, DA aims at pinpointing the existing (language) skills of an individual and finding the potential for ad hoc and future acquisition; it is therefore process- rather than product-oriented.

Bloom’s taxonomy

When creating tasks for CLIL instruction or deliberating assessment tasks, *Bloom’s taxonomy* is worth consideration. For example, the taxonomy was applied in the second language simulation experimented with in this study (see 5.4.2 and Appendix 12). Bloom’s prominent taxonomy of cognitive objectives, both the original from the 1950s and the revised version from the turn of the millennium (see Anderson, Kraftwohl & al. 2001), is helpful in all stages of instruction: objective setting, planning of instruction and assessment tasks. Furthermore, it is useful in exploring and defining what kind of language (structures etc.) is needed in different cognitive processes. The revised Bloom’s taxonomy distinguishes two levels of cognitive processes in instructional settings: lower order thinking skills LOTs (remembering, understanding and applying) and higher order thinking skills HOTs (analysing, evaluating and creating). When combining the dimension of cognitive processes with the knowledge dimension (factual knowledge, conceptual knowledge, procedural knowledge and metacognitive knowledge), a cross table of educational objectives is created.

The practical and illustrative 3D model portrayed in Figure 3 shows how these two dimensions interact when learning objectives are defined in various levels of the taxonomy.² The intersection of the cognitive process and knowledge dimensions provide examples of objective statements which are not to be confused with learning activities (Heer 2011) but can be used as such. The statement contains a verb in red and a noun phrase in blue. The verb usually

² See <http://www.celt.iastate.edu/teaching/RevisedBlooms1.html> for an interactive realisation of the model.

designates the intended cognitive process and the noun phrase the object of the knowledge dimension (Anderson, Kraftwohl & al. 2001, 4–5), which can be content-related, language-based or both. The higher the step in the model, the more demanding the cognitive process related to the knowledge dimension.

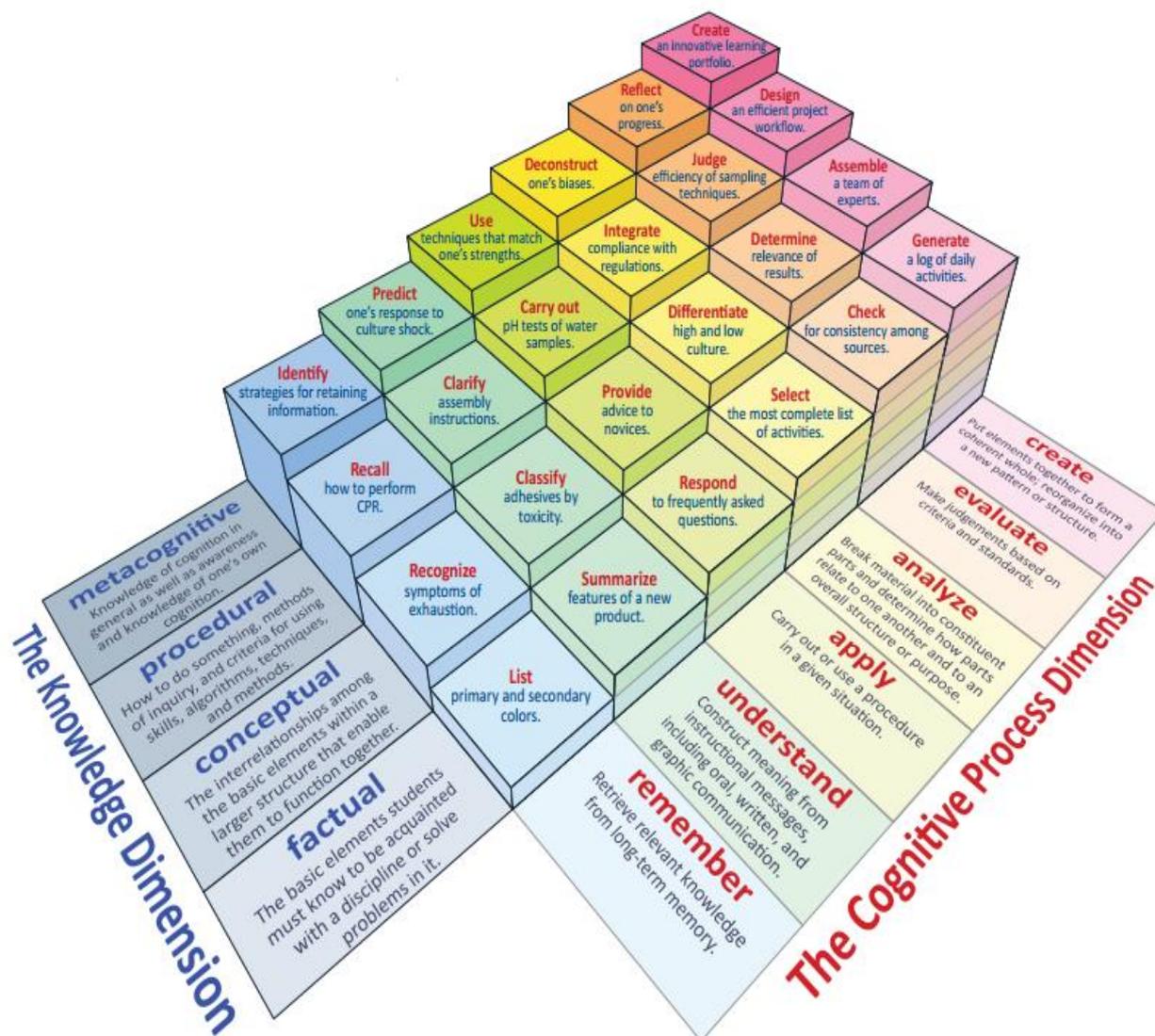


FIGURE 3. The model of learning objectives (Heer 2011)

CLIL scholars Mehisto, Marsh and Frigols (2008, 155) suggest that the revised taxonomy is “particularly useful as a checklist” to ensure that most tasks during the lesson are based on applying, analysing, evaluating and creating instead of understanding and remembering. Ideally, the objectives are planned to advance from LOTs to HOTs in a syllabus thus offering students cognitive challenges intertwined with linguistic and content goals. The same naturally applies to the design of assessment tasks which should not render memorisation and understanding only but the widest possible range of dimensions of cognitive processes and knowledge to avoid one-dimensional tasks.

Cummins' framework of language proficiency

Cummins has contributed significantly to the research and theories of bilingual programmes through his work concerning language minorities. He introduced the distinction of *Basic Interpersonal Communication Skills* (BICS) and *Cognitive Academic Language Proficiency* (CALP) in the late 1970s and early 1980s in order to draw teachers' attention to the linguistic struggle that bilingual immigrant children were facing at school (Cummins 2008, 71). This dichotomy is of fundamental relevance to the CLIL context because it refers to different kinds of language proficiencies of which CALP is more characteristic of CLIL study and BICS to instruction of EFL.

While BICS (conversational fluency, informal playground language, buddy language) refers to language in everyday face-to-face context-embedded situations with “meaningful supportive paralinguistic and situational cues such as gestures and intonation”, CALP (formal, academic language) is the context-reduced language register needed for success in schooling (Cummins 1982, 5–6; 2008, 72). Both are acquired in social contexts, but CALP gradually becomes diverged from language proficiency during the first grades at school and is the vital determinant of educational progress (Cummins 2008, 72; 1980a, 177–178). CALP in the CLIL context will be further discussed in section 3.2.1.

In order to refine the distinction between BICS and CALP both in L1 and L2 as related to the academic achievement of bilingual learners, Cummins (1982, 6) introduced a framework consisting of two continuums presented as horizontal and vertical axes, as in Figure 4.

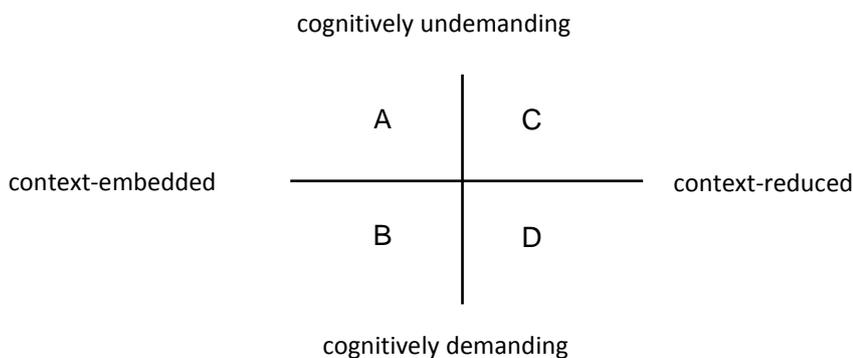


FIGURE 4. The Cummins model of language proficiency (Cummins 1982, 6)

Academic language tasks are typically context-reduced (quadrants C and D), whereas casual everyday language interactions are more context-embedded (quadrants A and B), but the dimensions in the models are rather interactive and fluid than polarities of a dichotomy (Cummins 1984, 13). The framework has inspired development of further SLA models such as Gass's Input-Interaction-Output (IIO) model (see Block 2003), but it has also been criticised, for example, for the choice of terminology and not considering social and cultural dimensions (Cummins 1984; 2008). The framework was later further elaborated accordingly. Hence, Cummins and Man (2007,

799–801) differentiate three distinct developmental elements in language proficiency: 1) conversational fluency (i.e. BICS), 2) discrete language skills referring to the learning of rules of language and 3) academic language proficiency (i.e. CALP) including “knowledge of the less frequent vocabulary of English as well as the ability to interpret and produce increasingly complex written and oral language”.

Additionally, Cummins’ model has given the foundation for the *CLIL Matrix* (see e.g. Coyle, Hood & Marsh 2010, 43, 67–68), a specific application for CLIL contexts, which incorporates the dimensions of cognitive and linguistic demands with low and high parameters forming a two-by-two analysis grid. The systematic construction of academic skills in the CLIL environment should advance from simple to complex, starting with low cognitive and low linguistic demands and ending with high cognitive and high linguistic demands. As a result, when distilling the essence of the Cummins’ framework and CLIL Matrix, the initial CLIL occurs in context-embedded situations featuring more BICS-type language, and the embedding is gradually reduced, simultaneously increasing the cognitive, content-based and linguistic demands. This principle can also be applied in assessment in CLIL.

Cummins’ Interdependence Hypothesis: Common Underlying Proficiency

Cummins’ (1980b) assumption that bilingual or multilingual people would have only one source of proficiency to draw from challenged the traditional perception of people having separate linguistic sources for each language. Cummins’ notion was portrayed in the hypothesis of *Common Underlying Proficiency* (CUP) which encompasses the skills and knowledge a bilingual person has in two or more languages and allows for transference from one language to another (transfer). It thus discusses the linguistic interdependence of both BICS and CALP across different languages: “In the CUP model, experience with either language can, theoretically, promote the development of proficiency underlying both languages, given adequate motivation and exposure to both, either in school or wider environment” (Cummins 1980b, 95).

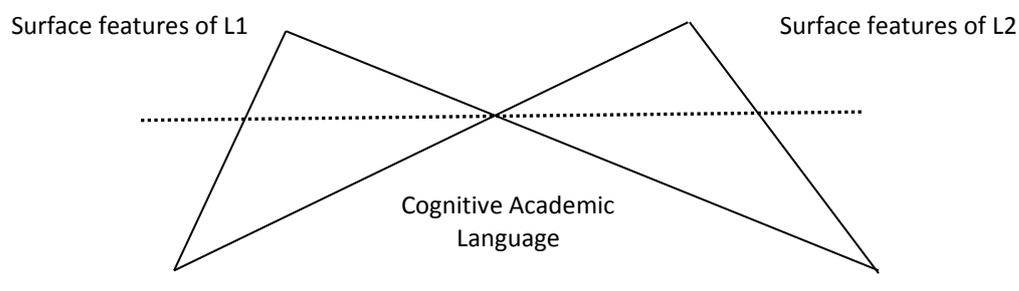


FIGURE 5. The dual-iceberg representation of bilingual proficiency (Cummins 1980b, 87)

Figure 5 illustrates CUP, also known as the *Dual Iceberg Model*. Instead of having two (or more) separate linguistic resources, the person has one to draw upon regardless of the number of languages. The surface features of these languages are depicted as tops of icebergs while most

of the language resource remains hidden. This is a focal notion which CLIL teachers and language teachers need to be aware of both in instruction and assessment. It is worthwhile to contemplate which strategies and tasks help pupils to produce as much and as nuanced language as possible so that they can demonstrate the variety of their language knowledge and skills hidden below the surface.

Miramontes, Nadeau and Commins (2011) have represented the CUP model visually as a *Conceptual Reservoir* that can be added to and accessed through any language a person speaks. At birth, individuals begin to fill the conceptual reservoir with concepts in their first language obtained through listening, observing, doing (Cf. the silent period in Krashen's Natural Approach), imitating and reading. These absorbed concepts are then used in production such as speaking, writing, artistic expression and physical movements. The reservoir is deepened and pathways strengthened by input and output. The second language is built on the L1 using the already existing conceptual reservoir. When the production in L2 starts, the utterances are usually something that the learners already know in their L1.

The reservoir can also be added to and deepened through the second language. Miramontes, Nadeau and Commins (2011) argue that once the concepts are acquired and restored in the L1 (or language/s learned after that), anything can be expressed through an additional language (positive transfer). As a result, the support given for the development of L1 always bears fruit in form of enhanced second language acquisition. In the bilingual CLIL classroom, it is the teacher's role to add to the reservoir using any means possible. Students can also learn to represent concepts learned in the L2 through the L1 as a result of transfer.

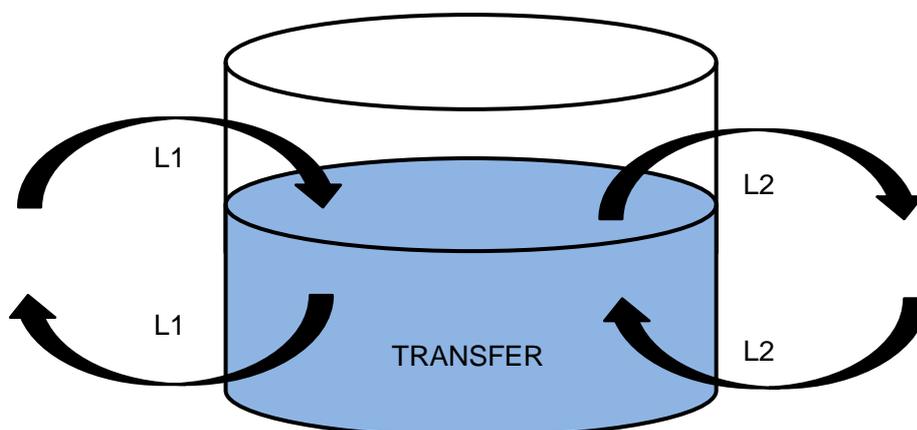


FIGURE 6. The Conceptual Reservoir (Miramontes, Nadeau & Commins 2011, 27)

This model can be depicted in each of its four stages: 1) L1 intake, 2) L1 intake + L1 output, 3) L1 intake + L1 output + L2 intake and 4) L1 intake + L1 output + L2 intake + L2 output. Figure 6 illustrates the complete final stage in which both or all languages are in use; the blue colour represents the part filled with concepts and the arrows input and output in L1 or L2. There is no intake – or output for that matter – without input. This principle represents the core component in Krashen's well-known theory of second language acquisition introduced next.

Krashen's Input Hypothesis

The *Input Hypothesis* (Krashen 1985), both influential and controversial, dates from the late 1970s, and was the “first broad-scope theory of SLA” (Block 2003, 19). It has had a long withstanding influence in bilingual content instruction; the echoes of this pack of theories, also known as the *Monitor Model*, are audible even in this research. The Monitor Model is embedded in five hypotheses discussed, among others, by Krashen and Terrell (1983):

- **the Acquisition/Learning Hypothesis:** the distinction of acquisition (“subconscious picking” in natural communicative situations such as learning a mother tongue) and learning (conscious activity, knowledge of language),
- **the Natural Order Hypothesis** stating that the rules of language (morphemes) are absorbed in a certain fixed order,
- **the Monitor Hypothesis** explaining how learning in adults is operating as a language “editor” affecting the formulation of utterances,
- **the Input Hypothesis** claiming that understanding the input slightly above the current level of language proficiency (“i+1” principle) is essential for the acquisition of language (including the notions of a silent period observed especially in children acquiring languages, i.e. gradually emerging spoken language) as well as
- **the Affective Filter Hypothesis** highlighting the fact that a learner’s “mental block” may prevent the acquisition or learning of a language regardless of rich, comprehensible input (Krashen 1985; Krashen & Terrell 1983).

Figure 7 illustrates the entity of Krashen’s Input Hypothesis model which represents implicit teaching and learning. It demonstrates how comprehensible input, preferably slightly above the proficiency level of the learner, is filtered through the learner’s affective filter (e.g. attitudes, feelings).

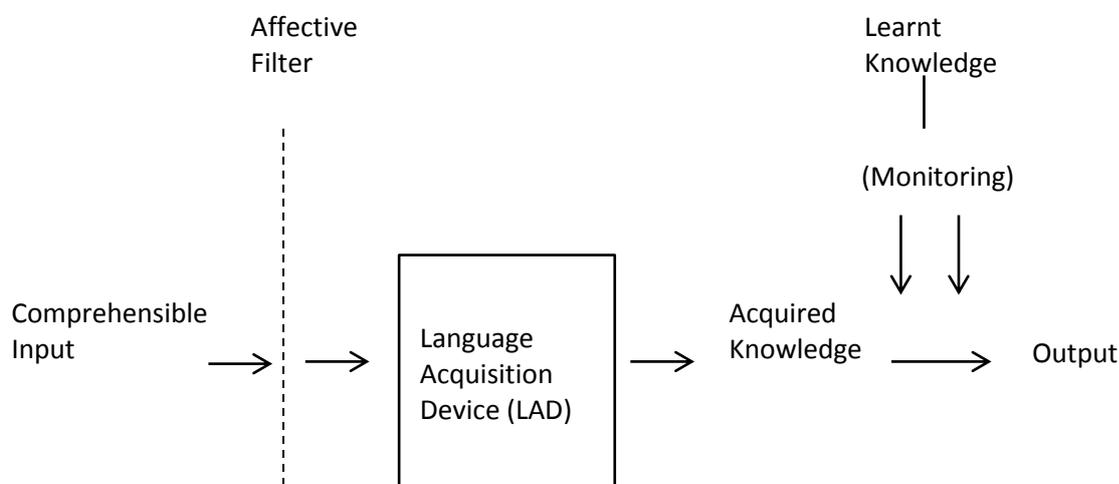


FIGURE 7. The Input Hypothesis Model of L2 learning and production (Cook 1993, 54)

The *Language Acquisition Device* (LAD) refers to Chomsky's notion of the innate capacity of children to acquire languages (later replaced by the concept of *Universal Grammar* UG, see e.g. Chomsky 1965). The figure makes the distinction between acquisition and learning salient – they are not synonyms in Krashen's theory. Table 4 clarifies this distinction further. The learned knowledge allows the (adult) learner to consciously monitor, revise, rephrase and correct the language which is then finally produced.

The differences between the language acquisition and language learning presented in Table 4 are often perceived as crucial features differentiating between CLIL and EFL. Explicit, conscious language learning in formal situations is associated with EFL, whereas implicit, subconscious language acquisition in informal situations is believed to be more emblematic of CLIL, although this conception has been challenged by the notion of focus on form (see p. 37). Furthermore, content study through an additional TL cannot be perceived as a totally informal learning environment because the context is instructional. Another polarisation is that EFL is in pursuit of accuracy and uses grammatical rules as the basis of linguistic production, while CLIL is keener on functionality. In this study, acquisition and learning are perceived as near-synonyms.

TABLE 4. Contrasting language acquisition and language learning (Cook 2013)

Acquisition	Learning
implicit, subconscious	explicit, conscious
informal situations	formal situations
uses grammatical 'feel'	uses grammatical rules
depends on attitude	depends on aptitude
stable order of acquisition	simple to complex order of learning

Krashen's model of SLA has been criticised for a number of reasons (see Block 2003 for a summary of this criticism and an account of other SLA models) but also embraced, particularly by teacher practitioners who recognised it as pragmatic; indeed, Krashen originally was an English teacher. In the light of contemporary, communicative and socio-constructivist views on language acquisition or learning, which are seen as complementary in this study instead of exclusive, the theory has lost some of its influence although not totally abandoned – it is rather viewed in light of more recent theories. One of such theories is Swain's Output Hypothesis.

Swain's Output Hypothesis

Swain's Output Hypothesis (1993) arose, among other rival hypotheses, as a response to the Input Hypothesis and drew attention to the possibility that acquisition and learning can also take place while producing language (spoken and written output); production may force "learners to recognize what they do not know or know only partially" (ibid., 159). The Output Hypothesis was premised on the observation of Swain that, regardless of the massive amount of input the Canadian immersion students received, their productive skills were weak. Swain (ibid.) differentiates four manners in which output may enhance language development:

- 1) output is meaningful practice contributing to automatisisation and fluency;
- 2) output may push the learner from semantic to syntactic processing especially when s/he notices an information gap and will therefore pay closer attention to the input;
- 3) output is a means to test communicative hypotheses (e.g. expressions, structures) and see if they work; and
- 4) output generates responses in form of feedback (e.g. clarification requests, corrections) which, in turn, helps the learner to modify the output.

In today's understanding of language acquisition it is fair to say that the Input and Output Hypotheses are interrelated and inseparable, forming a loop of acquisition and learning. In terms of the CLIL classroom, comprehensive i+1 input alone does not suffice to enhance language acquisition; the learners need to be activated to produce language and they have to have opportunities to practice TL production. As to CLIL assessment, tasks concentrating on comprehension alone do not reveal the language potential. For that reason, productive test items that encourage using the language as meaning conveyor, exhibiting content knowledge and incorporating the various levels of Bloom's taxonomy (see p.29) are a necessity in CLIL assessment. Output is equally important as input in CLIL. Also, the question of implicit acquisition and explicit learning as differentiated in Krashen's model is of interest to CLIL and has intrigued language theorists and, at least until recently, remained an issue for debate.

Implicit (incidental) and explicit (intentional) learning

The dichotomies of implicit and explicit L2 learning as well as incidental and intended L2 learning in the area of cognitive psychology are of interest to all language teaching approaches. *Implicit learning* takes place without conscious noticing, while *explicit learning* follows from learning with rules (see Table 4). In that sense, it is natural learning, because pieces of information are absorbed and acquired from the rich learning environment for later use and application (Ellis 1994, 1). In simple terms, implicit learning is learning without rules or without awareness of rules (Ortega 2009, 99–100), which explains why the distinction explicit/implicit is often made in terms of grammar instruction (DeKeyser 1998, 56).

The question of whether second language learning is “possible without intention, without attention, without awareness and without rules” in the first place has puzzled SLA researchers (Ortega 2009, 94). In fact, it is one of the key questions in CLIL: should language be taught or should learners be allowed to acquire it on their own without any specific focus on it? The general consensus is, according to Ortega (ibid.), that incidental learning is possible (for example L2 vocabulary acquisition during pleasure reading as in Huljstin 2003, 363), but learning with intention results in faster, better and more extensive learning than without it (Ortega 2009, 95, 107).

There is controversy in the CLIL field over whether the language aspect should be treated overtly (learning), tacitly (acquisition) or a halfway approach adopted. For instance, Järvinen (2004) states that the key factors in the CLIL method are: the prioritisation of content and

meaning, less emphasis on language forms and rules in teaching, meaningful communication, increase in the time dedicated for learning and more opportunities for learning. Dalton-Puffer and Smit (2007) represent the opposite view according to which implicit language acquisition is commonly – but wrongly – associated with CLIL.

Dalton-Puffer and Smit (2007, 8–9) list several beliefs and pseudo-attributes of CLIL based on Krashen's Input Hypothesis theories. Such beliefs, according to them, are based on pure intuition without solid research background. For example, CLIL is believed to provide a naturalistic language learning environment, 'a language bath', accenting implicit language acquisition over explicit language learning. Another belief is that meaning (content) precedes form (language) in CLIL classrooms, thus enabling meaningful communication and reduction of language anxiety. Dalton-Puffer and Smit (*ibid.*) explain that these common beliefs are the aftermath of theoretical models that have had an impact on language learning research, in particular immersion and CBI, and thereby also on the foundations of CLIL. In fact, Dalton-Puffer (2007a, 194) describes the move away from Krashen's theoretical base as "fundamental" for CLIL.

Krashen (1985) is the most notable proponent of abandoning explicit grammar instruction although, as stated by DeKeyser (1998, 56), Krashen's arguments in favour of implicit learning are "mostly indirect". The Input Hypothesis presented earlier in this section is based on large amounts of comprehensible input, whereby the students are to "induce the rules...without any conscious learning", similarly to L1 acquisition (DeKeyser 1998, 56). One may query whether this leads to 'sink or swim' situations where the learner has to cope independently and self-sufficiently with the comprehensible input slightly above his/her current abilities (the *i+1* principle).

From such concerns follows the discussion of whether or not CLIL should rest on implicit teaching and learning of language. Allen, Swain, Harley and Cummins (1990, 75) claim that "not all content teaching is necessarily good language teaching" because the focus is on grasping content (meaning) rather than on language (form). They argue that learners need to be assisted to focus on relationships between form and meaning, i.e. language and content. This is the topic of the next passage.

Focus on form

In CLIL, the primary focus is traditionally on meaning and the secondary on language (see e.g. Järvinen 2004, Lorenzo 2007). This view has been challenged by several scholars. For example, Lightbown and Spada (1994, 576–577) claim that "[t]here is evidence that not all language features can be acquired when learners' attention is focused exclusively on meaning". This evidence comes mainly from immersion studies finding that the target language skills do not develop as fully as expected (see e.g. Genesee 1987; Swain 1993), presumably due to the heavy emphasis laid on content learning which leaves language acquisition as incidental (Xanthou 2011, 118). Lightbown and Spada (2008, 184) note that experience with communicative language teaching and CBI shows that "meaning-based exposure to language allows L2 learners to

develop comprehension skills, oral fluency, self-confidence, and communicative abilities, but they continue to have difficulties with pronunciation as well as with morphological, syntactic and pragmatic features of the L2". Language is the vehicle for learning which is why *focus on form* is needed. As a result, Swain (1988, 81) argues that second language learners need to be guided to notice and learn the relationships between meaning (content) and form (language) so that language production is facilitated and the initial struggle with language is alleviated.

According to current principles, instruction is most powerful and efficient when it delivers both content and language related goals, and contains attention to both form and meaning (Spada & Lightbown 2008). The integration of form-focused instruction is thus no longer questionable – rather questionable is the deployment: an integrated or isolated focus on form (how, when, during or outside of the lesson and the content context)? What the simultaneous driving of both content and language objectives ultimately requires – along with finding the right balance between these two – is a focus on form and choosing between implicit and explicit approaches to language teaching and learning. The choice between implicit and explicit teaching and learning is “one of the most critical” because they are related to two kinds of approaches to language: synthetic and analytic (Long and Robinson 1998, 15).

A synthetic, structural approach is analogous to *focus on forms* methodology; that is, EFL standard language teaching with methods such as Grammar-Translation and Total Physical Response (Long & Robinson 1998, 16). The analytic approach can be divided into two sub-categories: *focus on meaning* and *focus on form* (ibid.). Focus on meaning approaches do not treat “the languages as an object of study, but by experiencing them as a medium of communication”, while focus on form is “helping learners to use features that they have already partially acquired with greater accuracy rather than entirely new forms” (Ellis 2012, 18).

The term focus on form refers to approaches involving “an attempt to induce incidental acquisition through instruction by drawing learners’ attention to linguistic forms while they are communicating” (ibid., 272). In comparison to focus on meaning, focus on form incorporates some instruction of linguistic features when the occasion arises while performing “a series of pedagogic tasks” in a meaning-focused lesson; a shift to focus on form may coincide with content instruction, drawing students’ attention to linguistic codes, i.e. form (Long & Robinson 1998, 23). Table 5, synthesised from Long and Robinson (1998), Graaff and Housen (2009), Lightbown and Spada (2008) and Ellis (2012, 272), elucidates the distinctions between focus on meaning, focus on form and focus on forms.

One could claim, then, that focus on meaning refers to approaches such as CLIL and CBI perceived narrowly, while focus on form concerns CLIL and CBI putting the dual focus completely into effect. Underrating the language focus is not in the spirit of CLIL. Language in CLIL is certainly not “just a tool”, a statement which, according to Tella (1999, 26), seems to “be rooted in an innocent and somewhat naïve belief that foreign languages can be regarded as simple tools in Teaching [t]hrough a Foreign Language”. The orientation of CLIL should, instead of seeing language as tool, regard language as a tool *and* an object of study (dual focus).

TABLE 5. Distinctions between focus on meaning, focus on form and focus on forms

Aspect	Focus on meaning	Focus on form	Focus on forms
Orientation	language-as-tool, analytic	language-as-tool, analytic	language-as-object, synthetic
Type of learning	incidental	incidental/intentional	intentional
Options in language teaching (e.g.)	Natural Approach Immersion	Task-based language teaching, Content-based language instruction	Regular EFL, Grammar Translation, Total Physical Response
Primary focus of attention	meaning (content)	meaning (content)	form (language)
Secondary focus of attention	form	form	content
Acquisitional processes	interpsychological mediation; intrapsychological mediation	interpsychological mediation; intrapsychological mediation; noticing; noticing-the-gap; modified output	conscious rule-formation; proceduralisation; automatisation; monitoring
Syllabus type	task-based	task-based; linguistic functions	structural, rule-based
Instructional processes (e.g.)	tasks; scaffolded production; comprehensible input; negotiation of meaning;	tasks; scaffolded production; dynamic assessment; input-priming; negotiation of meaning; corrective feedback; linguistic consciousness- raising through tasks	exercises; consciousness- raising through the provision of explicit rules; structured input; controlled production practice; free production practice; corrective feedback

Dalton-Puffer (2011) explicates that the perception of relating only implicit language teaching and learning to CLIL is a relic from the early beginnings of CLIL in Europe; the principle of implicit content-based language instruction can be derived from the Canadian immersion. Furthermore, she states that “we can detect a clear move away from relying solely on the idea of self-propelled, implicit language learner” (ibid, 194).

There is an impressive record of research showing that focus on form instruction has a positive impact on the development of students’ interlanguage and linguistic production, especially accuracy, in immersion, CBI and CLIL environments (e.g. Cormier & Turnbull 2009; DeKeyser 1998, 56; Housen & Pierrard 2005; Loewen 2005; Lyster 2004; Pérez-Vidal 2007; Pica 2002; Rodgers 2006; Schleppegrell, Achugar & Orteiza 2004; Xanthou 2011; Zuengler & Brinton 1997) and as a result, there is call for focus on form in CLIL (Pérez-Vidal 2007). This provides support to Bialystok’s (1978) versatile model of second language learning which acknowledges both input and output, incorporates implicit, explicit and other kind of language knowledge for inferencing as well as both formal (EFL-type) and functional (CLIL-type) practising.

Graaff, Koopman, Anikina and Westhoff (2007, 608, 620) recognise form-focussed processing, implicit or explicit, as beneficial to language learning in CLIL settings and list it as one factor in effective language teaching in CLIL. They also posit that effective instruction for language acquisition in CLIL encompasses, within the wider SLA framework of language exposure, use and motivation, the following features: functional communication, simultaneous

attention to form and meaning as well as corrective feedback (ibid., 607). They propose a language-sensitive observation tool for effective language pedagogy in CLIL contexts based on Westhoff's SLA Pentapie Model. The tool was designed to observe teacher classroom performance which facilitates the acquisition of learner language in CLIL contexts. Form-focussed input is placed centrally in the observation tool graph (Figure 8). However, the study of Graaff and colleagues (2007) detected no instances of form-focussed practices in Dutch secondary CLIL classrooms – a parallel finding with Pica (2002) from tertiary CBI. It seems that the theory of effective CLIL instruction has not yet been embraced in the educational field.

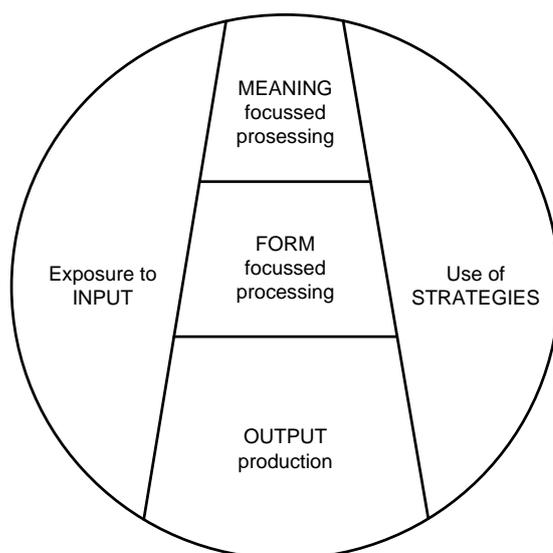


FIGURE 8. An observation tool for effective CLIL instruction (Graaff & al. 2007, 610)

The promotion of form-focussed instruction in CLIL is not a subject of total agreement. Lorenzo (2007, 266) notes that selected language items should be a part of a “hidden curriculum” by which he posits that language should not be explicitly instructed but rather “absorbed unintentionally”. Functional language use in realistic, ‘real purpose’ situations has also faced criticism. Cook (1997) judges focus on meaning and use of authentic, naturalistic language: he calls learner-centred, meaning-driven approaches such as task-based instruction and CBI “fashions”. He denotes that “[t]he belief in a focus on meaning is the dogma of our time. It derives from an uncritical acceptance of theories of language and language acquisition developed without reference to what learners want or need”.

Instead of focusing on meaning and authentic or natural language only, which Cook also challenges as concepts, he advocates language play as a resource in language teaching. He argues that it is wrong to think that authentic language use (if such a thing exists, he disputes) entails (Cook 1997, 230):

using language only for task-solving, for social action, or for talking about the real world. Authentic, natural language both for children and for adults can also be preparatory, repetitive, artificial, removed from reality, and focused upon the rules of the game, including the rules of grammar and phonology.

He concludes that language learning should be perceived as a complex phenomenon that embodies every possible type of language use – including language play that includes, for instance, rhymes, songs, stories and literature.

CLIL does not exclude language play, and I consider language play and focus on form rather supplementary than exclusive. As DeKeyser (2003, 321) states, “a considerable amount of work suggests there is a positive role of some kind of attention to form, that is, either through the explicit teaching of grammar and explicit error correction, or at least through more indirect means such as input enhancement”. This brings us to Long’s Interaction Hypothesis and Schmidt’s Noticing Hypothesis.

Long’s Interaction Hypothesis

In the 1980s, Long proposed an earlier version of the *Interaction Hypothesis* which was later updated (Long 1996). As the name implies, conversational interaction is perceived to have a critical, facilitative role in language acquisition, because it alternates with output, production, input and comprehension, especially when the discourse concerns negotiation of meaning (Long 1996, 449). Furthermore, meaning-focussed interaction is productive and meaningful, and it conflates “internal learner capacities, particularly selective attention” (ibid., 452). In order for the negotiation of meaning to lead to language acquisition, semantically contingent, versatile speech (such as “repetitions, extensions, reformulations, rephrasings, expansions and recasts”) produced by a competent speaker must be frequently present because the discourse should recycle the related target forms so that they become more salient and easier for the learner to notice (ibid., Cf. the Noticing Hypothesis).

The input should therefore be modified, for instance, by increasing the frequency of target forms, stressing the key words, repeating core parts, switching lexis, moving the target forms in a sentence to initial or final positions and pausing before or after them (ibid.). What this suggests for CLIL classrooms and language acquisition is that teachers should engage in arranging tasks that require and stimulate the negotiation of meaning, for “they may be one of the easiest ways to facilitate a learner’s focus on form without losing sight of a lesson’s (or conversation’s) focus on meaning” (ibid., 454). Thus, students should interact rather than teachers lecture.

Schmidt’s Noticing Hypothesis

In 1990, Schmidt proposed that noticing, which is a subjective experience, is the “necessary and sufficient condition” for the linguistic input to converse into linguistic intake (Schmidt 1993, 209). That is, if the linguistic feature under scrutiny is not noticed and “consciously registered” (Schmidt 2010, para 1), no learning or acquisition takes place. This notion, suggested to clarify eclectic and misconceiving terminology circling the dichotomies of conscious/unconscious, incidental/intentional and implicit/explicit in language learning, became known as the *Noticing Hypothesis*. As with any theory or hypothesis, Schmidt’s Noticing Hypothesis faced strong criticism, but it also generated enthusiasm and a number of empirical studies, most of which have

been supportive towards the theory (see Schmidt 2010 for a summary). For example, Swain and Lapkin (1995) examined and provided data to support their notion that it is actually the awareness (i.e. noticing) of a linguistic problem in output that generates cognitive processing leading to modified (i.e. enhanced) output. This process is depicted in Figure 9.

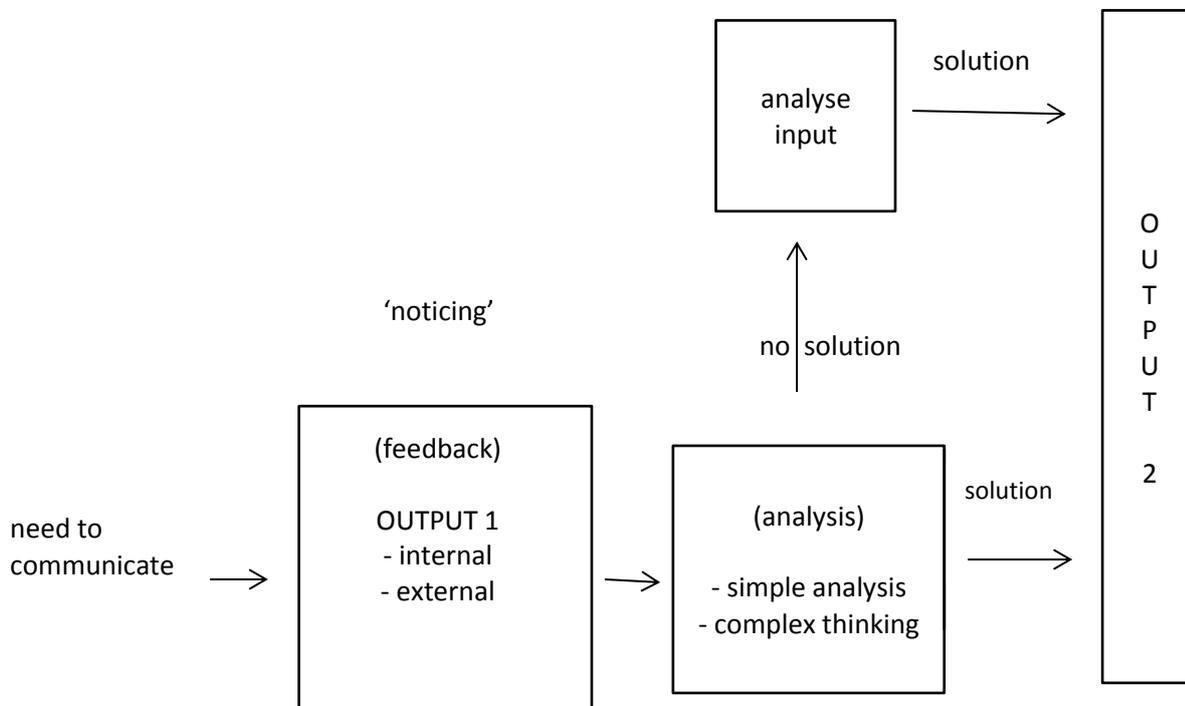


FIGURE 9. Output, noticing and second language learning (Swain & Lapkin 1995, 388)

In order to facilitate learning, teachers need to draw learners' attention towards relevant or problematic language features (feedback) when encountered either during instruction or production. This may occur by giving the exact rule or assisting the learners to infer the rule by inductive reasoning which gives them the joy of detection and builds ownership of learning (external feedback and analysis). As evident in the illustration, this may also occur intrapersonally. This is, the learner notices problems on their own, after which follows an investigation of the problem (internal feedback and analysis). If the problem is not solved, a more intensive analysis of the input (e.g. science texts, grammar book, teacher or peer scaffolding) is needed. Either way, the solution found for the problem results in modified output and enhanced learning. Comprehensible input, focus on form, scaffolding in the zone of proximal development, feedback (in form of assessment, for example) and ample possibilities for output are hence vital elements of SLA in general and CLIL in particular.

SUMMARY

The rationale for CLIL is based on theories of second language acquisition that emphasise collaboration, working with challenging materials and linguistically more capable language users. The distinction between communicative language and the language of schooling is important in

content learning and therefore also assessment. The underpinning theories stress the quantity and quality of input, output and interaction in CLIL as well as drawing from the previously learned languages as a linguistic resource and gradual abandonment of context-embeddedness. There is not, however, complete agreement on whether the language learning should be implicit or explicit in CLIL, but the tendency is strongly towards focus on form-type instruction which is informed by research pertaining to intentional/incidental learning and the Noticing Hypothesis. The traditional notion of CLIL being a form of implicit teaching and learning has thus been challenged by the notion of focus on form which to some extent necessitates explicit language instruction in order to enhance more accurate learner language development in relation to the target content.

2.3 CLIL in Finland

CLIL instruction in Finland has long roots dating back to the turn of the 1990s when changes in the Finnish legislation concerning basic education, upper secondary education and vocational education allowed the use of a foreign language in instruction (Takala, Marsh & Nikula 1998, 139). A few authors even credit Finland with the notion of being the birth country of CLIL (Fortanet-Gómez & Ruiz-Garrido 2009, 50; Graddol 2006, 86) or the “primus motor for CLIL in Europe” (Marsh 2013, 132–133). Reasons behind the implementation of CLIL are many, but as Baker (2011, 246) argues, “CLIL cannot be understood from a purely linguistic or educational perspective. As with all forms of bilingual education, there is a political ideology underneath.” Finland as a relatively small, officially bilingual country (Finnish and Swedish) and as a member state of the European Union (EU) is in a position where promoting bi- or multilingualism has been, and still is, essential for competent activity in international and European markets of labour, economy, politics and so forth.

The strong advocacy of multilingualism by the EU was stated for the first time in the White Paper on Education and Training (1995) which pronounced the so called “1+2 objective” according to which EU citizens should attain a proficiency in three community languages: their mother tongue and at least two other languages. It is thus not surprising that many European countries have embraced this initiative and intensified their language instruction by introducing CLIL. The political underpinning for CLIL in the European context (see also Eurydice 2006, 8–9) has steadily increased and is especially salient in the Luxemburg Presidency Position Document on Plurilingual Education (2005) drawing the conclusions that, among other things, CLIL instruction should be expanded, good CLIL practices investigated and public awareness of CLIL benefits increased. This call for action has been heard in the field of CLIL research since the number of published studies and interest in CLIL has steadily risen since the millennium and particularly the 2010s.

The rise of CLIL among the European language programmes thus reflects the EU policies, but there are a large number of other factors contributing to the increase of CLIL implementations in Finland as well as in other countries. One of them is the impetus brought by the strong evidence from immersion studies. Additionally, the traditional EFL instruction does not provide enough challenges for Finnish learners.³ The obvious general benefits of CLIL to the individual learner as discussed by Wolff (2007) add to the popularity of CLIL. He names a few advantages for the CLIL learner: 1) they learn languages better and faster due to a larger exposure to language; 2) they succeed in grasping more content because the cognitive processes are more complex due to the presence of an additional language; 3) they simultaneously become better prepared for the working life due to the academic language registers they need to use and 4) the co-operative learning methods give them an additional advantage in the labour market (Wolff 2007, 21–22). Moreover, the general reasons expediting bilingual education (see p. 20) have an impact on the Finnish CLIL proliferation.

In CLIL, language is used for meaningful purposes. The effect of shifting the focus from language forms to content is also one potential source of growing popularity (Dalton-Puffer, Nikula & Smit 2010b, 286). Additionally, the so called Social Turn in SLA influenced by sociolinguistics and socio-constructivist learning theories have contributed to the emphasis placed on the *use* of foreign languages instead of language knowledge obtained through traditional study (see e.g. Block 2003). Therefore, especially in the Finnish context, the modifier ‘functional’ or ‘operational’ seems to be often attached to language proficiency – the term *functional language proficiency (toiminnallinen kielitaito)* thus refers to language command which is sufficient for coping with various specific everyday situations (e.g. Pöyhönen & al. 2009). In CLIL, functional proficiency thus denotes a language command sufficient enough for following instruction in the TL, studying through the TL (e.g. solving mathematical problems) and linguistic expertise in subject-specific language.

Finnish CLIL Implementation and the CLIL curriculum

The CLIL trajectory in Finland can be roughly divided into three decades. The first decade of CLIL in the 1990s were the years of experimentation, launch and enthusiastic expansion supported by national agencies as well as the institutions of the EU. The initial CLIL model was based on Canadian immersion (see 2.1). The second decade, the 2000s, was an era of establishing scientific CLIL study (see especially 3.2) concerning mainly the language development in CLIL: i.e. the efficacy of the approach, students’ motivation, affective factors and classroom discourse. The third decade, the 2010s, will hopefully witness increasing professionalism through experience and expansion of the body of research also into areas such as assessment, cognition, content knowledge expressed through the TL, academic language and curriculum development.

³ See the news article reported by the news of national public service broadcasting company YLE from 2011 in http://yle.fi/uutiset/english_in_schools_too_easy_for_pupils/5089579 (April 30, 2013).

CLIL provision in Finland is part of mainstream education and available at all levels, and its “central objective is that the pupils be able to acquire a firmer language proficiency than in lessons reserved for the language in normal instruction” (NCC 2004, 270). Especially CLIL instruction in basic education (classes 1–9, ages 7–15) and upper secondary level has been popular since the 1990s. The surveys conducted in 1996 (Nikula & Marsh 1996) and 2005 (Lehti & al. 2006) show, however, that bilingual education is a downward trend. In 1996, 11.7% of schools at primary, lower and upper secondary levels provided CLIL instruction, whereas in 2005 the provision of CLIL had decreased to 5.7%. These two surveys were conducted at school level. The most recent follow-up survey from 2011 at municipal level (Kangasvieri & al. 2012), however, reveals that slightly over half of the municipals currently offering CLIL instruction estimate that the demand for CLIL will grow in the future, and the rest of municipals state that it will remain constant.

English is the predominant language in the European CLIL provision (Eurydice 2006, 56). This is also the case in Finland, although particularly in larger municipalities, other languages such as German, French and Russian are also options for CLIL (Marsh, Järvinen & Haataja 2007, 70). In Helsinki, in addition to the above mentioned languages, CLIL provision in Mandarin Chinese–Finnish, Spanish–Finnish and Estonian–Finnish is also available.⁴ Swedish immersion is provided particularly in the Helsinki Metropolitan area and Ostrobothnia, the Swedish-speaking western coastal area. Yet, despite parental interest in bilingual Swedish–Finnish schooling, such schools do not exist due to the ‘Finnish only’ or ‘Swedish only’ school systems ordained by law (Sundman 2013).

Finland represents a decentralised educational system which denotes that external control and steering is minimal (see Shohamy 2001, 29 for qualities of centralised and decentralised systems). The diversity of the Finnish CLIL environment is derived from the pedagogical freedom at the municipal, school and individual teacher levels. The National Core Curriculum for Basic Education (NCC 2004) defines the very broad guidelines, objectives and contents of CLIL instruction in Finland, providing the framework according to which each municipality and educational establishment should draw upon and compose a local level curriculum. The local curriculum determines its own pedagogical emphases such as the provision of bilingual instruction and the form of it. Consequently, there may be several curricula in a municipality: the general municipal curriculum and school-specific and CLIL (or other specific-purpose) curricula. In Finland, curricula are always open-access documents that are almost without exception available online, which is why their specifications are easily accessible for examination for anyone interested in educational and curricular matters.

In the NCC (2004), CLIL is referred to using the general term ‘instruction in a foreign language’ as opposed to ‘foreign language instruction’ (EFL). The NCC states that the instruction

⁴ State of art July 28, 2013. See Helsinki City Web page in English for current information: <http://www.hel.fi/hki/opev/en/Services/Education+for+foreigners/Teaching+in+foreign+languages>

provider decides which name will be used for CLIL in the local context. Typical for the Finnish context is the spectrum of designations which indicates how splintered the field of CLIL is. Kangasvieri and colleagues (2012) found 12 different labels for bilingual content instruction in Finland; the terms instruction in a foreign language was the most frequent and CLIL the second. Other variations were, for example, bilingual teaching, language-enriched and English-emphasised instruction.

The NCC is currently being renewed in a transparent, participatory process involving several drafting phases, students, parents, teachers, experts and any interested citizen, and it will come into effect in the beginning of the autumn term 2016. The final NCC draft⁵ appears to adopt the term bilingual instruction as a superordinate term, but differentiates between extensive and limited bilingual instruction. Instruction is defined as extensive when at least 25% of the instruction occurs in the TL or when it pertains to immersion of national languages (mainly Finnish or Swedish, but Sami, Romany or sign language may also come into this category). When less than 25% of tuition occurs in the TL, it is limited and the term language-enriched instruction is used. If these characterisations persist in the final document, the minimum TL exposure in CLIL is 25% in Finland.

The instruction provider is allowed to decide to which extent it implements CLIL instruction: the current NCC (2004, 270) states that the instruction provider is to “specify what subjects, and how much of their instruction, are to be taught in the foreign language”. Mother tongue and literature are excluded from CLL. It is noteworthy that the NCC does not impose which content within a subject has to be taught through the two languages: “The [local] curriculum specifies which subject areas that support the study of different subjects will be taught in the foreign language” (NCC 2004, 271). This principle remains untouched in the NCC draft (2014) for NCC 2016. Furthermore, and most importantly, the NCC (2004, 270) posits that the objectives of the foreign language acquisition in the given CLIL context have to be premeditated:

As a minimum, the objectives specify what sort of level is sought, in the course of basic education, in listening- and reading-comprehension skills, speaking, writing, and cultural skills.

As a result, the language specifications, the language objectives and the roles of L1 and L2 need to be determined along with the definition of the desired level of language proficiency. This naturally stipulates that the CLIL curriculum contains the specifications of at least linguistic aims, subjects of CLIL instruction and the proportion of foreign language. In the new NCC draft (2014), issues to be determined locally are distinctly listed.

However, regarding linguistic objectives, the NCC draft (2014) is less imperative; it suggests that in specifying the linguistic objectives, the scales of the Common European Framework of Reference for Languages (CEFR 2001, see also 3.1.1) may be used. The draft emphasises that the multifaceted development of the TL has to be taken care of, but no longer mentions the four basic skills and nor does it mention the desired level of language proficiency as

⁵ See the draft (in Finnish language) in http://www.oph.fi/download/156716_opsluonnos_luku_10.pdf.

an objective. Instead, subject-specific language accumulation is noted in addition to enhancing the correctness of language. This can be interpreted as suggesting that a focus on form approach should be favoured, especially since the pupils' role as *both* language learner and user is mentioned.

A teachers' role in bilingual instruction is not specifically described in the current NCC. In the future bilingual classroom, however, teaching should display language-awareness and good language pedagogy according to the NCC draft (2014, 75). The draft also states that instruction should take place in an "authentic language use environment" (*ibid.*) which implies that the language proficiency of a teacher in bilingual instruction must be high. The language demands for CLIL teachers are specifically raised with reference to the corresponding decree given by the Ministry of Education (2005). According to this decree, the competence of a CLIL teacher in spoken and written TL should be excellent. The Ministry of Education posits that such proficiency is exhibited by a) a certificate of at least the second highest level (5/6) in a National Certificates of Language Proficiency⁶, b) a minimum of 80 credits of academic university-level TL studies which are equivalent to advanced studies or c) an acknowledged certificate of teacher education abroad in a country in which the TL is an official language.

Due to these high language demands, officially qualified CLIL teachers are rarely recruited (see also Miettinen, Kangasvieri & Saarinen 2013, 80–83). The general minimum requirement for a Finnish teacher is a master's degree in education, behavioural sciences or in the subject they teach. Because of their academic studies, teachers stand in high esteem and their professional competence is trusted. The survey of Kangasvieri and colleagues (2012) reveals that the most decisive criterion for the recruitment of CLIL class teachers in Finland is their basic education instead of language proficiency or linguistic studies. This results in situations where the linguistic quality of CLIL instruction may vary enormously.

These teacher-related issues pose a true challenge to CLIL implementation. CLIL has influenced the emergence of a new teacher variety in Finland: one who teaches a language without being a language teacher, as Nikula and Järvinen (2013) point out. In some countries (e.g. Italy), CLIL teaching is executed by subject and language teachers in co-teaching pairs (Serragiotto 2007), whereas in Finland, the class teachers are, in most cases, solely responsible for both content and language instruction in their classrooms. The language proficiency of the CLIL teacher thus carries huge weight in the success of CLIL instruction but is not the sole linguistic input factor, because Finnish pupils in primary education also attend conventional EFL instruction parallel to their CLIL lessons. To what extent the CLIL class teacher and English teacher work together has not been studied yet in Finland.

⁶ An introduction to the National Certificates of Language Proficiency and skill level descriptors are available at <https://www.jyu.fi/hum/laitokset/solki/yki/english>.

Facets of Finnish CLIL instruction

The aspects that need to be considered when establishing – and maintaining – CLIL instruction are portrayed in Figure 10 which practically summarises the content of this section. The figure is slightly edited from the original by Mustaparta and Tella (1999, 46) to be compatible with the current NCC (2004) especially in the facet of objectives. The pedagogical trinity of the curriculum (objectives), implementation (instruction) and assessment (reflection) form a visible cycle.

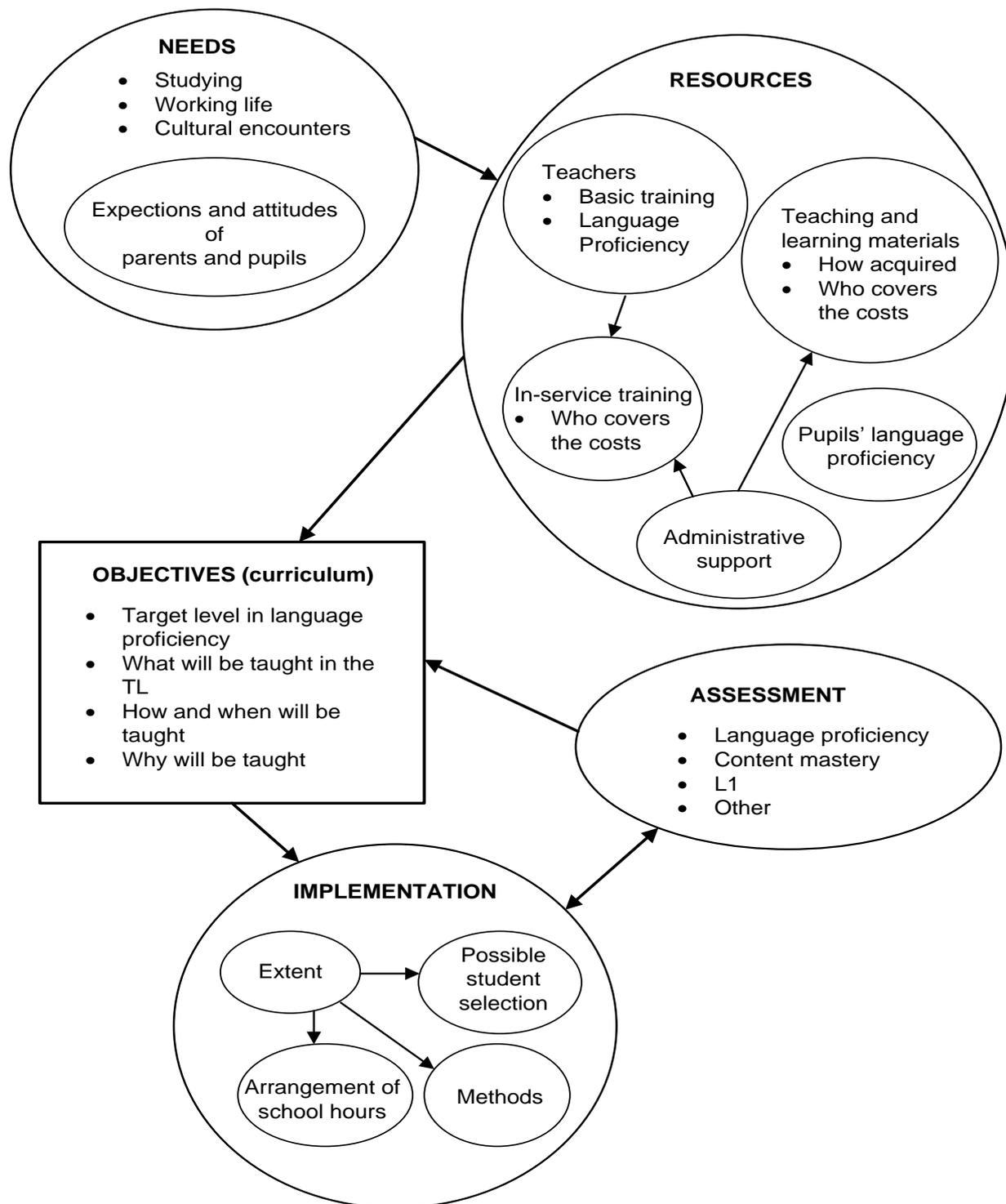


FIGURE 10. The facets of CLIL instruction (modified from Mustaparta & Tella 1999, 46)

Assessment in its own right is thereby presented (also in the original) as an integral, inseparable part of CLIL instruction to be considered prior to the launch of CLIL. The initial language proficiency of the learners is an obvious factor that affects the determination of learning objectives and actual instruction. Further, teachers' language proficiency has been valued as a significant resource contributing to the successful CLIL implementation. Assessment has an influence on both the instruction and the further design of instructional objectives either in syllabus or curricular planning.

The NCC (2004, 273) also gives specifications regarding assessment in CLIL. The document points out that

[a]ssessment must give the teacher, pupil, and parents or guardians adequate information about the pupil's language proficiency in relation to the given objectives. Growth in comprehension of the foreign [...] language is to be monitored, especially when instruction in the foreign language [...] begins in other subjects.

This is an unambiguous statement: target language assessment must be practiced in the CLIL context. What means or methods should be used is, again, an issue for the education provider and the individual CLIL teacher to consider, as is the frequency of assessment, the practice of conveying the gathered information to the parties involved, the quantity of information and what qualities that information should have in order to be adequate. In this respect, research is urgently needed and this study attempts to answer that call. In the final NCC draft (2014), these principles remain untouched, but the word 'adequate' has been replaced by 'versatile', and the monitoring of language development in each CLIL subject is highlighted as well as the use of self- and peer assessment and assessment as a collaborative act among teachers. The draft thus entails specifying language assessment in several ways.

Assessment is always connected with the objectives of instruction, the actual implementation of it and CLIL pedagogy, i.e. principles derived from the underpinning theory. Assessment in CLIL has hardly been studied so far, and the manifold possibilities to arrange it need to be explored (see e.g. Barbero & Järvinen 2009; Byrnes 2008; Johnstone 2000; Langé 2007). This issue will be further addressed in section 4.3.

CLIL pedagogy is coherently informed by the methodologies used in immersion and CBI, although it has been necessary to modify those methods so that they would better suit the European context in which the aim is to add to the multilingualism of linguistic majorities. Nikula and Järvinen (2013, 145) have synthesised the basic features of CLIL instruction (see also Dalton-Puffer 2011, 183–184; Lasagabaster & Sierra 2010). These characteristics should give the reader a general idea of CLIL in Finland:

- The foreign language used in the instruction is not commonly used in the surrounding society but it is used as a lingua franca.
- The foreign target language is a prestigious language, most often English.
- CLIL teachers are most often not native language speakers, nor are they language teachers, but often content teachers.

- CLIL does not aim at replacing formal language instruction: CLIL lessons are separate from formal language lessons.
- Typically the ratio of foreign language is less than 50% of total instruction, although there are exceptions to this.
- CLIL normally starts after the acquisition of literacy in the language of instruction.
- The often superficial objectives of CLIL vary according to the extent of implementation.
- The overarching goal is to help the learners to achieve a functional language proficiency which enables them to use the language appropriately in diverse situations.

Dalton-Puffer (2011, 184) argues that under such conditions in CLIL could be conceived as “a foreign language enrichment measure packaged into content teaching”. The NCC reform may thus present itself as a promise for a more professional, linguistically aware, objective-oriented, assessment-active and form-focussed future CLIL in Finland.

SUMMARY

Finland was one of the first European countries to launch the CLIL approach. The National Core Curriculum for Basic Education (NCC 2004) gives the framing guidelines for CLIL provision, but leaves many decisions to be met and stated in the local CLIL curriculum at the municipal level. Language assessment is distinctly mentioned in the current NCC as well as in the draft for the renewed NCC 2016. Assessment should give the stakeholders in bilingual content instruction adequate and multifaceted information on the level and progress of the foreign target language. The NCC reform in 2016 will guide towards more precise CLIL implementation, because CLIL in Finland is translated into practice in multiple, often incomparable ways – implementation thus varies from school to school and even from one classroom to another. Finnish teachers are highly educated and enjoy pedagogical freedom in their work. The official language qualifications for CLIL teachers are supremely high, and therefore not always met. The provision of CLIL has decreased since the first years of implementation at all levels of education, but the popularity has been rather steady in recent years. English is by far the most common CLIL language in Finland.

3 SECOND LANGUAGE PROFICIENCY

A *second language* (L2) is any language learned or acquired after the mother tongue(s) (L1), and it is perceived as synonymous to *foreign language* in this study. When facilitating progress in learner language and assessing second (or foreign) language proficiency, it is necessary to grasp the principles behind second language acquisition and define which constituents form language proficiency as well as contexts of language use and knowledge (Alderson 2005, 1; Cumming 2008, 3). The context of the language use, CLIL, was discussed in the previous chapter. This chapter defines the constituents of second language proficiency in CLIL, examines relevant research and literature in that field and suggests what this all signals for CLIL instruction and assessment. The CLIL teacher, as Jäppinen (2004, 200) remarks, should be aware of the main differences between child and adult proficiency in both L1 and L2 in order to adequately teach and assess the TL. Within the scope of this study the principles of second language acquisition are not discussed (see Dörnyei 2009 or Ortega 2009 for that).

The concept *language proficiency* is defined in various ways in the literature. Simply put, language proficiency is “a general term denoting the degree of skill with which a person can use a language” (McKay 2006, 3). Language proficiency is a complex concept to define, because it is sometimes assimilated or more often contrasted with language ability or communicative competence (Nunan 1986; Martin-Beltrán 2010; McNamara 1996; Vollmer 1983). The initial distinction between competence and performance was made by Chomsky (1965, 4) who argued that competence is “the speaker-hearer’s knowledge of his own language” (i.e. grammar, vocabulary and other qualities language) and performance is “the actual use of language in concrete situations”. He perceives language use and performance as parallels: “the theory of language use – the theory of performance” (Chomsky 1965, 9).

Performance and proficiency can be conceived synonymously (Nunan 1986) just as well as competence and ability are synonyms (Vollmer 1983). However, I am more inclined to shortly define *proficiency as the use of language (performance) while manifesting language knowledge (competence)*. The following characteristics of language proficiency are synthesised from Vollmer (1983), Nunan (1986), Brindley (1992), Cohen (1994) and Martin-Beltrán (2010). In this study, language proficiency pertains to:

- 1) the extent and adequacy of the learner’s **control** of the (foreign) language,
- 2) the **ability to use language** in particular communicative situations with the help of several interrelated sub-skills such as syntax or socio-cultural competence,
- 3) the **functional application** of one’s **linguistic knowledge** and
- 4) the subjective **understanding of that language use** in different social situations.

CLIL, as pointed out in the preceding chapter, is more about using the language meaningfully than explicit learning. Functionality is thus the key concept in understanding language proficiency in CLIL contexts. The attainment of functional language proficiency is sometimes declared to be

the overall language objective which denotes that the learner is able, in different subject contexts, to use the TL satisfactorily and accurately enough for both acquiring and manifesting content knowledge. This implies that learners' language proficiency does not have to be perfect in order to be functional. Even inadequate language proficiency may be adequate in certain situations if the language user is able to compensate for communicative gaps with compensation strategies (e.g. paraphrases) or other resources available (e.g. electronic dictionaries, consulting others) in the environment (Sjöberg 2004, 147). Sjöberg remarks that the more language is used, the more functional the language user estimates it to be (ibid.). This is a crucial notion in terms of self-assessment.

Since this study is interested in the various ways in which language proficiency can be displayed for assessment purposes in CLIL settings and how these situations are utilised for assessment, I am adopting the terms proficiency and language performance rather than language competence and ability. I will also use the term functional language proficiency to emphasise the use of language for a specific purpose: content study. In other words and within the scope of this study, language proficiency represents and articulates a person's abilities or competencies across the language domains that become salient in communicative, meaning making settings.

This view of proficiency constituting of enabling skills is in alignment with the Common European Framework of Reference CEFR (2001) which will be introduced in the sub-section 3.1.1 subsequent to a brief, retrospective overview of developments in defining communicative, functional language proficiency (3.1). In section 3.2, I will concentrate on the manifestations of second language proficiency in CLIL. Since CLIL accumulates language needed for studying topics of various school subjects, I will examine academic language CALP, classroom interaction and second language development in CLIL settings by reviewing research and literature in those areas in the sub-sections of 3.2.

3.1 Models of second language proficiency

Modern foreign or second language education has adopted a holistic view on language in general: language is viewed as the "primary mediator of learning" which acknowledges its intellectual and social significance in all human intercultural and transcultural communication (Harjanne & Tella 2008, 57). Language also includes "the language learner and user with his/her [background] knowledge, skills, awareness, qualities and personality factors" (ibid., 59). The role of language thus is crucial in learning (see also Halliday 1993 for his considerations of language-based theory of learning).

Since the ultimate aim in CLIL is that the learner develops functional language proficiency, it is purposeful to turn to a model of language *use* rather than language knowledge. In order to better comprehend – and assess – the nature of language proficiency, and the entity and constituents of it, scholars have attempted to depict models of language ability (for summaries

see Bachman 1990; Cumming 2008; Fulcher & Davidson 2007 or McNamara 1996). Each decade has provided new insights into language proficiency so that the sophisticated nature of language and language use has gained more recognition.

Early models of language proficiency

Prior to the 1960's, linguistic structuralism emphasising the structure (grammar) and systems of a language prevailed, and language teaching and learning was practically based on automatism (Vollmer 1983, 6). The early models of both Lado (1961, 1964) and Carroll (1968) adhere to the *four basic language skills* of listening, reading, speaking and writing that are viewed as separate skills and independent components distinguished from the knowledge of, for example, grammar and vocabulary. The division of the four language skills is widely used (Cf. CEFR 2001; NCC 2004) and is still valid today due to its practicality.

In the 1970s, both Hymes (1967, 1972) and Halliday (1979) broadened the four skills perspective by introducing the aspect of interaction in social contexts. Hymes (1967) embarked upon a generation of models of *communicative competence* and became known as the voice of the sociolinguistic movement due to the stress placed on language as a functional device in social interaction and communication. Halliday's theory of language, known as the *Systemic Functional Linguistics*, derives from the functions of language: language constructs and conveys meanings and is interpreted as a social system.

The essential in Halliday's perception of language is that he sees it as a "meaning potential" which denotes the semantic or linguistic options a person has in interaction (Halliday 1979, 27). For example, the initial language functions of a pre-schooler may involve the following (Webster 2009, 223):

1. instrumental ('I want'): satisfying material needs
2. regulatory ('do as I tell you'): controlling the behaviour of others
3. interactional ('me and you'): getting along with other people
4. personal ('here I come'): identifying and expressing the self
5. heuristic ('tell me why'): exploring the world around (and inside one)
6. imaginative ('let's pretend'): creating a world of one's own and
7. informative ('I've got something to tell you'): communicating new information.

In short, Halliday explored, instead of language constructs, the ways it was used. This is why the theory has been seen as fundamental in CLIL circles (see Llinares, Morton & Whittaker 2012, 13).

Models of communicative language proficiency

Halliday's functions of language as well as Hymes's communicative competence, in stressing the language use in social settings, paved the way for many subsequent models that also recognise those dimensions. These include the model of Canale and Swain (1980) later modified by Canale (1983), the model of Bachman (1990) which was modified by Bachman and Palmer (1996, 2010),

the model of Celce-Murcia, Dörnyei and Thurrell (1995), the psycholinguistic model of Skehan (1998) and the Common European Framework of Reference for Languages (CEFR 2001), which as such does not represent a pure theoretical model but rather a referential, comprehensive taxonomy framework of language use in various aspects and contexts. Each model has naturally been shaped and influenced by its predecessors.

Each model of language proficiency includes three dimensions: 1) the models of knowledge, 2) underlying factors affecting the language performance, as well as 3) the actual language use (McNamara 1996, 48). The first refers to linguistic knowledge, for instance grammatical knowledge; the second includes all general and individual factors, such as personality, that may or may not influence the actual language use (ibid., 61–76). Fulcher and Davidson (2007, 37) state that the first two dimensions of knowledge and performance together form the factors that influence the communicative competence or the communicative language ability of a person which in turn enables the third dimension, the actual language use. In the following, I will introduce the Common European Framework of Reference for Languages (CEFR 2001) which is more inclined to describe the third dimension of a language proficiency model, the language use.

3.1.1 Common European Framework of Reference CEFR

The *Common European Framework of Reference for Languages* is a language policy document published by the Council of Europe with the aim to provide a mutual basis for European language education in designing “language syllabuses, curriculum guidelines, examinations, textbooks etc. across Europe” (CEFR 2001, 1). The CEFR underpins the aspirations of the European Council to promote European plurilingualism, and due to its prominent background, it is widely accepted and applied for learning, teaching and assessment purposes. For example, the Finnish National Core Curriculum for Basic Education (NCC 2004) has adopted the CEFR taxonomy as the basis for communicative language assessment.

The CEFR (2001, 1) describes “what language learners have to learn in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively”. The CEFR attempts to provide “objective criteria for describing language proficiency” for various purposes (ibid., 1). Yet, it does not explicitly define the concept proficiency but describes the constituents of communicative language competence. The CEFR (2001, 108–130) differentiates three basic components of communicative competence:

- 1) **linguistic competences** including lexical, grammatical, semantic, phonological, orthographic and orthoepic competencies;
- 2) **sociolinguistic competences** involving the knowledge of linguistic markers of social relations, conventions of politeness, expressions of folk wisdom, register differences as well as dialect and accent; and

- 3) **pragmatic competences** adhering to discourse competence, functional competence and interaction schemata.

These components of communicative language ability are characterised in fairly detailed ‘can do’ statements at three levels from A to C which are further divided into total of six subcategories forming a verbal scale taxonomy which serves also as assessment reference for ‘*how well* can do’. It defines performance levels in listening, speaking, reading and writing from the ‘Basic user’ (A1 and A2) through the medium level of ‘Independent user’ (B1 and B2) to the ‘Proficient user’ (C1 and C2), but it also provides scales for specific language aspects such as flexibility, sociolinguistic appropriateness and grammatical accuracy.

The approach of the CEFR to language is “action-oriented” and “it views users and learners of language primarily as ‘social agents’, i.e. members of society who have tasks (not exclusively language-related) to accomplish in a given set of circumstances, in a specific environment and within a particular field of action” (CEFR 2001, 9). These circumstances may thus also refer to CLIL classroom environments where the pupils are using the language for accomplishing content-related tasks and where the focus is more on language use than in explicit learning of language features. The six-levelled scales, as in the following scale of text processing in Table 6, can therefore be used for designing instruction but also assessing CLIL language use at general level.

TABLE 6. Six-levelled CEFR scale for processing text (CEFR 2001, 96)

PROCESSING TEXT	
C2	Can summarise information from different sources, reconstructing arguments and accounts in a coherent presentation of the overall result.
C1	Can summarise long, demanding texts.
	Can summarise a wide range of factual and imaginative texts, commenting on and discussing contrasting points of view and the main themes.
B2	Can summarise extracts from news items, interviews or documentaries containing opinions, argument and discussion.
	Can summarise the plot and sequence of events in a film or play.
	Can collate short pieces of information from several sources and summarise them for somebody else.
B1	Can paraphrase short written passages in a simple fashion, using the original text wording and ordering.
	Can pick out and reproduce key words and phrases or short sentences from a short text within the learner’s limited competence or experience.
A2	Can copy out short texts in printed or clearly handwritten format.
A1	Can copy out single words and short texts presented in standard printed format.

In adopting the CEFR, several benefits can be gained. Firstly, the taxonomy is widely available and established in Europe. Secondly, the validity of assessment increases since the description levels are identical for every CLIL student regardless of teacher, municipality and country of residence or the level or form of education. CEFR facilitates the comparison between the proficiency levels of students regardless of the implementation of bilingual education – the learning outcome matters, not the instructional methodology. Thirdly, the taxonomy coding system from A1 to C2 is simple enough for the stakeholders to follow, and the criteria provided by

the CEFR for proceeding to the next level are effortlessly verified and used even for self-assessment purposes.

Fourthly, the CEFR promotes learners' language awareness by encouraging self-reflection especially in form of its offspring, the European Language Portfolio (ELP). Additionally, the framework does not bind its users to any specific philosophy or assessment method such as tests, portfolios or interactive simulations; the scales and descriptions are applicable to any form of assessment. Finally, in providing a variety of scales, it is detailed enough to ensure, according to the need, either a general-level or more detailed mapping of the CLIL learner's language proficiency in a specific linguistic area (Cf. Table 6). A framework, such as the CEFR, is an interphase in instruction and assessment design.

Fulcher and Davidson (2007), in discussing the hierarchy of models, frameworks and test specifications, note that where models are overarching, frameworks are specifications of models and test items are designed leaning on frameworks. The theoretical background of the CEFR, the model of language proficiency it is grounded on, remains obscure. Järvinen (2012a, 224) remarks that the CEFR is based on the proficiency model of Bachman (1990), whereas Alderson (2005, 28) describes it as a synthesis of work by the Council of Europe stretching more than three decades back "from the notional-functional syllabus to the Threshold Level for English" which were elaborated through several versions to the current six-levelled CEFR. The CEFR document does not specify its theoretical foundation; it rather settles for stating that "it is the result of over ten years' research by a number of leading applied linguists and pedagogical specialists from the 41 member states of the Council of Europe" (CEFR 2001, back cover), but the identities of the specialists remain unknown for the large public.

Johnstone (2000, 133) is concerned with frameworks arising from experts' insights rather than research because "they may reflect an idealized rather than a real notion of what proficiency is and how it develops". He argues, for example, that the most basic level of spoken production in the CEFR disregards "the songs, poems, games and aspects of mathematics, science, history, geography and drama" that primary school language learners "soon experience through their foreign language, and which enables them to pull chunks of language from their long-term memory store that can go well beyond 'simple phrases or sentences'" mentioned in the basic user level description (*ibid.*, 132–133). This is a valid argument, especially in relation to CLIL instruction. The CEFR does not take academic language specifically into consideration.

The project work on describing aspects of academic language (subject-specific language) currently in progress in the European Centre of Modern Languages (ECML) is valuable and much needed (see ECML 2013 or p. 63). The project of the ECML, however, is not intended for the primary level leaving a gap to be tackled in the future. Mohan and Slater (2005) have paved the way for such work in their research report on the adaptability of Halliday's Systemic Functional Linguistics in the language integrated science study in primary contexts. They conclude with an urgent need for "linguistic definition and analysis of content" (*ibid.* 169).

The CEFR has also been criticised for being improper for the needs of young language learners (e.g. McKay 2006, 307-310). Little (2007, 651) raises the question of “how far [the CEFR] can accommodate two of the most important growth-points in school-based language learning across Europe: early-start and content and language integrated learning (CLIL) programs” which is rooted in the communicative focus of language use and the age-appropriateness of the CEFR. He argues, for example, that the levels C1 and C2 are not applicable to children due to the cognitive maturity and experience required in tasks that elicit such language (see also Hasselgreen, Kaledaité, Maldonado-Martin & Pizorn 2011).

Little (2007, 652) also doubts whether the full potential of the CEFR has been exploited in bringing “curricula, pedagogy, and assessment into fruitful interaction with one another”. He discusses the use of the ELP as one example of the triadic combination and calls for an assessment culture that “accords with the CEFR’s action-oriented approach and explicitly accommodates the self-assessment fundamental to effective ELP use”. The CEFR through the ELP is thus perceived as one enabler of assessable, goal-oriented language pedagogy. The problem is, however, that the ELP model templates currently available are intended for EFL rather than CLIL use, although experiments have been made to develop an age-appropriate, CLIL-compatible language portfolio for young learners (Wewer, forthcoming).

SUMMARY

Models of language proficiency or communicative language ability have evolved from the descriptions of individual language skills to an understanding according to which language proficiency is seen as a complex, multicomponent issue, the main emphasis of which is on the communicative language use in a social context. The models serve as a theoretical reference for frameworks such as the CEFR which can be used as tools for designing actual classroom teaching and language assessment. The adopted models and frameworks represent a shared understanding of what constitutes language use in various situations and therefore help users pay heed to the complicated nature of language and view it from diverse angles rather than remain at a surface level. In the European context, regardless of its obvious failure to commit to younger language learners and subject-specific language, the CEFR nonetheless provides a mutual springboard for educational implementations that ensure uniform basis for CLIL operation modes throughout Europe.

3.2 Language proficiency in CLIL

As was stressed in the previous chapter, CLIL instruction carries two objectives: the learning of content and language. The overarching objective of implementing CLIL in the first place is to improve students' general language command, as the NCC (2004, 270) and the CLIL expert team of CLIL Compendium⁷ articulate. The fundamental leading thought behind CLIL is that the learners acquire higher (academic) functional language proficiency than they probably would by merely attending standard EFL instruction. In order to gain access to knowledge through disciplinary language, develop subject-specific literacies and prepare for further studies in the foreign language, students need to be familiarised with the conventions of academic language. This calls for heightened language awareness for both CLIL teachers and students because EFL-type casual language does not suffice in CLIL study in the long term.

Language in the CLIL classroom context has different roles as demonstrated in Table 7: language related to different subjects, classroom discourse and language development (Llinares, Morton & Whittaker 2012, 15). The language proficiency in CLIL is thus manifested in a more complex way than general, BICS-type language proficiency. These roles are applicable to both L1 and L2 since, as indicated in section 2.2, proficiency in L1 supports the L2 acquisition and vice versa.

TABLE 7. The roles of language in CLIL instruction (Llinares, Morton & Whittaker 2012, 15)

SUBJECT LITERACIES		CLASSROOM INTERACTION		LANGUAGE DEVELOPMENT
Genre Register	ASSESSMENT	Instructional and regulative registers (focus)	ASSESSMENT	Expressing ideational meanings (key concepts and understandings)
		Communication systems (approach)		Expressing interpersonal meanings (social relationships, attitudes)
		Interaction patterns and scaffolding (action)		Expressing textual meanings (moving from more spoken to written forms of language)

Assessment is a significant part of this framework; it lends support to the various roles of language in CLIL. The role of “formative and dynamic” (ibid.) assessment is to provide evidence of progress which, in turn, supplies material for reflection and self-assessment to ensure further improvement and continuing progress in the diverse roles of language in CLIL. I will use this framework to investigate the language proficiency in CLIL through relevant research and literature. First, I will look at academic language in CLIL which is constituted of, for instance, various subject literacies (3.2.1). I will then examine CLIL students' language behaviour (discourse) in classroom interaction (3.2.2) and finally, I will touch upon what research reveals about the development of language proficiency in CLIL (3.2.3).

⁷ See the project website www.clilcompendium.com for CLIL dimensions and foci.

3.2.1 Academic language

The crucial factor in the integration of content and language is that the language used in CLIL study is appropriate for the purpose. The consensus view seems to be that *academic language* (CALP) is an essential and integral part of general academic proficiency and content study, as the model of Krashen and Brown (2007) depicts (Figure 11). Krashen and Brown maintain that *academic proficiency* is constructed of two components: “1) academic language, characterized by complex syntax, academic vocabulary, and a complex discourse style [...] and 2) academic content, the content of subjects such as algebra, history, literature etc.”. Both of these components are supported by strategies enhancing their acquisition.

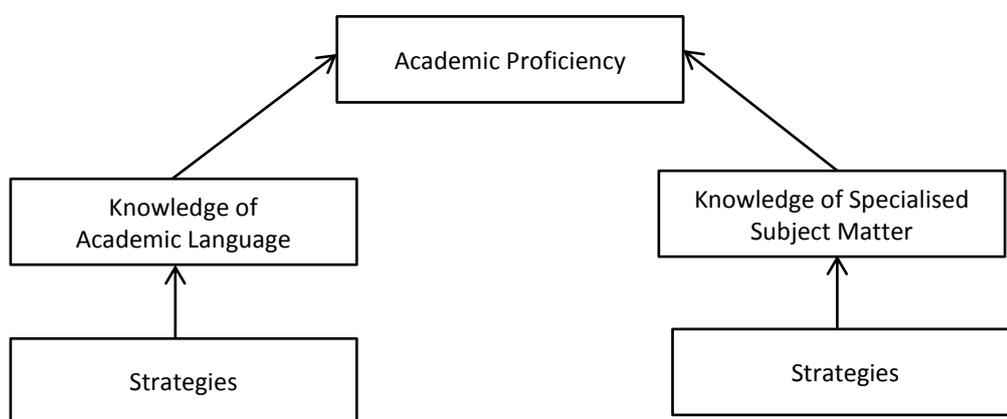


FIGURE 11. A view of academic proficiency (Krashen & Brown 2007, 1)

This model is particularly interesting in connecting both the language and content aspects to general academic proficiency. Krashen and Brown (2007) base the model on two hypotheses:

- 1) **the Comprehension Hypothesis** according to which literacy development and language acquisition occurs rather through grasping messages than conscious, intended learning (Cf. Krashen’s Input Hypothesis on p. 34); it suggests reading as exemplary means of academic language development;
- 2) **the Problem-Solving Hypothesis** claiming that subject content is ideally not learnt through rigorous study but problem-solving which deepens the existing knowledge. The illustrated strategies comprehend reading and writing strategies, and teaching of strategies is worth considering because they assist learners to gain autonomy in learning (ibid.).

An older model, Chamot and O’Malley’s (1987) Cognitive Academic Language Learning Approach (CALLA), which was proposed to alleviate the transfer of students with limited English language proficiency into American mainstream education and to accelerate their academic achievement, also includes literacy and learning skills (i.e. cognitive strategies) in academic competence, which according to Chamot (2007, 317), is “far more than merely becoming proficient in English”. CALLA combines 1) procedural knowledge (language as a tool: *how*,

understanding and generating language), 2) declarative knowledge (content: *what*, e.g. facts and rules) based on Anderson's cognitive theory which stresses the fact that "the interplay between declarative and procedural knowledge leads to the refinement of language ability" and 3) learning strategies nurturing autonomy development in students (Chamot & O'Malley 1987, 232).

Subject content is thus encountered through academic language. Schleppegrell, Achugar and Oteiza (2004, 68) state that "[t]o achieve advanced literacy and disciplinary knowledge, students need to be able to understand how language construes meanings and how concepts of school language are realized in language. In other words, disciplinary knowledge is not taught in isolation from language." For these reasons, Schleppegrell (2006, 51) argues that language focus needs to be linked with subject teaching and teachers should engage students in language analyses resulting in *language-based content teaching*, because the language of schooling features dense information, abstraction and technicality, multiple semiotic systems, expectations for conventional structure and appropriate voice.

Features of academic English

Academic English is one variety of English (Scarcella 2003, 2), and in order to operate satisfactorily in educational settings, students need to master the registers (types of language used in particular situations) and genres (different text types) of English. What is satisfactory in given situations at each level should be defined in the curriculum – mediation of academic English should, naturally, be proportioned to match the age, prior knowledge and aptitude of learners. The students need scaffolding to develop academic literacy, and, as Scarcella (2003, 10) argues, conventions of academic English should be taught, because there are "regular features of academic English that are well defined and teachable".

Scarcella (2003) proposes a framework for academic English consisting of three components: 1) linguistic (the phonological, lexical, grammatical, sociolinguistic and discourse components), 2) cognitive (metalinguistic abilities, higher order thinking, background knowledge, and strategies), and 3) sociocultural/psychological components (norms, values, beliefs, attitudes, motivations, interests, behaviours, practices and habits). The framework thus extends and combines the models of language proficiency presented in this chapter in a different way. The comparative Table 8 describes the features of linguistic components of academic English in comparison with ordinary, everyday English with a few examples (i.e. CALP versus BICS).

It becomes apparent that academic English is more concise, dense and precise in expression of information. Academic language is, as Snow and Uccelli (2009, 118-120) argue, an "inventory of features" such as interpersonal stance, information load, organisation of information, lexical choices, reasoning strategies and disciplinary knowledge. They also posit that the concept of academic language is relative; "language can be *more or less* academic" (ibid., 115) – there are no absolute polarities which is why familiarisation with appropriate registers and genres is crucial for successful study.

TABLE 8. Comparison of linguistic features in ordinary and academic English (slightly reduced from Scarcella 2003, 12)

Linguistic components of ORDINARY ENGLISH	Linguistic components of ACADEMIC ENGLISH
The Phonological Component	
<p>knowledge of everyday English sounds and the ways sounds are combined, stress and intonation, graphemes and spelling Examples: <i>ship</i> – <i>sheep</i> /ʃ/ - /i:/ <i>sheet</i> – <i>cheat</i> /ʃh/ - /ch/</p>	<p>knowledge of the phonological features of academic English, including stress, intonation and sound patterns Examples: demógraphy, demográphic, genéric</p>
The Lexical component	
<p>knowledge of the forms and meanings of words occurring in everyday situations; knowledge of the ways words are formed with prefixes, roots, suffixes, the parts of speech of words, and the grammatical constraints governing words Example: find out</p>	<p>knowledge of the forms and meanings of words that are used across academic disciplines (as well as in everyday situations outside academic settings); knowledge of the ways academic words are formed with prefixes, roots and suffixes, the parts of speech of academic words and the grammatical constraints governing academic words Example: investigate</p>
The Grammatical Component	
<p>knowledge of morphemes entailing semantic, syntactic, relational, phonological and distributional properties; knowledge of simple rules of punctuation</p>	<p>knowledge that enables learners to make sense out of and use the grammatical features (morphological and syntactic) associated with argumentative composition, procedural description, analysis, definition, procedural description and analysis; knowledge of the grammatical co-occurrence restrictions governing words; knowledge of grammatical metaphor, knowledge of more complex rules of punctuation</p>
The Sociolinguistic Component	
<p>knowledge that enables learners to understand the extent to which sentences are produced and understood appropriately; knowledge of frequently occurring functions and genres</p>	<p>knowledge of an increased number of language functions; the functions include the general ones of ordinary English such as apologizing, complaining and making requests as well as ones that are common to all academic fields: knowledge of an increased number of genres, including expository and argumentative text.</p>
The Discourse Component	
<p>knowledge of the basic discourse devices used, for instance, to introduce topics and keep the talk going and for beginning and ending informal types of writing, such as letters and lists</p>	<p>knowledge of the discourse features used in specific academic genres including such devices as transitions and other organisational signals that, in reading, aid in gaining perspectives on what is read, in seeing relationships and in following logical lines of thought; in writing, these discourse features help learners develop their theses and provide smooth transitions between ideas</p>

Development of academic language

According to Cummins (1982, 6), the development of conversational, casual language proficiency (BICS) takes approximately two years for English language learners, and it takes five to seven years to achieve context-reduced academic language proficiency CALP. This estimation was concluded from studies involving English language learners of immigrant background attending language programmes in the United States as well as successful immersion programmes (ibid.),

and it has been reinforced by several other studies. For example, a study by Shohamy and colleagues reported in Cummins and Man (2007, 801) discovered that Russian and Ethiopian immigrant students in Israel “require at least 9 years to catch up to their peers in academic Hebrew”.

It needs to be noted that these studies were conducted on immigrants in an environment in which the target language was the dominant, prevailing language of the society: English in the United States, French in Canada and Hebrew in Israel. One has to keep in mind that the participants of those studies were substantially immersed in the TL environment and the linguistic exposure was therefore more extensive than mere school study. Furthermore, the teachers were either bilinguals or English-speaking natives. The extent of language exposure both within and outside school was significantly higher than it is in the European CLIL context, although English can be often perceived as a lingua franca in the Nordic countries.

Quick results in CLIL contexts with fairly low exposure to the TL are unrealistic, especially concerning academic English. The CLIL conditions are not even remotely similar – even within Europe they are often totally opposite in terms of TL exposure, for instance. Generally speaking, the circumstances are different in Northern Europe in comparison with Southern Europe, which is why it is reasonable to assume that the development of academic language proficiency is likely to take longer in CLIL contexts where the learners are typically speakers of the majority language, the extramural and school-internal exposure to the target language (TL) is lower, and the teachers are not bilinguals or native speakers of the TL.

Based on this purely theoretical hypothesis, I assume that it will take considerably longer than 5–7 years for Finnish CLIL students to acquire academic language. If the teachers are not aware of the characteristics of academic language and do not promote its practice, it may take even longer. It appears that in different CLIL-providing countries the obtained linguistic results vary. Environments with early extramural exposure to English combined with early start of CLIL instruction with high linguistic exposure, national recognition of CLIL policies and the quality of CLIL teacher resources - especially in terms of teacher language proficiency - have a significant effect on language accumulation (Sylvén 2013). Acquisition of academic language in CLIL contexts is thus an interesting topic requiring further investigations.

Academic language and CLIL

CLIL represents a movement away from grammar-oriented language study towards a “genre-based approach” to “all language study” (Lorenzo, Casal & Moore 2010, 435). Language in CLIL and other content and language integrating classrooms is mainly, but not only, used for conveying academic subject content; academic classroom language must therefore be included in CLIL discussion in order to enhance students’ disciplinary learning (Dalton-Puffer 2007b). Research has shown that language has an impact on subject development (e.g. Vukovic & Lesaux 2013, see also Snow & Uccelli 2009).

Various school subjects represent contextualised academic language and demand different types of *subject-specific literacy*. Therefore, academic performance also entails genre mastery in addition to mastering various linguistic features or components in academic language (Snow and Uccelli 2009, 118). Different school subjects adhere to different subject-specific language conventions: history, for example, is expository, often uses past tenses and explains causal relationships, while the language of home economics or physical education is normally instructional making use of the imperative mood (see Llinares, Morton & Whittaker 2012 for an extensive review of genres, grammar and lexis in different CLIL subjects).

In order to recognise subject-specific language and genres, analyses of “the language demands of different content subjects, which include the language of curriculum materials and of classroom participation, need to be analysed so that the students can be taught the actual language functions, structures and subject-specific vocabulary that they will need” (Chamot & O’Malley 1987, 236). Academic language analyses may be useful for CLIL classes, because they ensure proper curricular and syllabus planning and make systematic progression possible, but the relevance of highly academic language in primary schools has to be carefully considered according to the proportion of CLIL exposure and age of pupils.

Subject-literacy and its systematic development seem to be issues that have been brought to the fore in educational circles world-wide (Cf. also NCC draft 2014). The ECML project is an example of such mapping and a first step towards academic language analyses in the European context. The European Centre for Modern Languages (ECML) has initiated a two-year (2012–2013) project of creating language descriptors indicating the academic language competences needed in schooling linked together with the CEFR proficiency scales and subject content objectives (ECML 2013). These descriptors are intended to be useful especially for teachers of migrant and minor language students aged 11/12 and 15/16. Because they define the core subject-matter language requirements in mathematics and social sciences thus integrating academic language and content objectives, the project grids promise guidelines for integrated CLIL assessment. The outcomes will be disseminated through the ECML web page, conference presentations and various publications. Such work should preferably be extended to lower and higher levels of education in order to ensure continuous, curricular progression in academic language from preschool to university levels.

In the U.S., such work was started in the eCALLMS (2013) project, the purpose of which is to raise teachers’ and teacher trainees’ awareness of linguistically responsive instruction that improves multilingual learners’ acquisition of language, subject-specific literacy and content knowledge. This in turn should be reflected in students’ improved learning outcomes. The focus is on identifying vocabulary and genres typical for various topics in mathematics and science and designing materials and classroom activities that enhance learning of especially ESL/EFL speakers. The project eCALLMs will expand in satellite projects of the University of Turku. A Finnish project funded by the Academy of Finland also attempts to find a more solid way to integrate content and language (ConCLIL 2013).

Enhancing development of academic language in CLIL

In a CLIL environment, or any environment combining language and content learning, it is important to demonstrate and understand how language encounters content and especially, as Long (1996) argues from the basis of his Interaction Hypothesis, how language forms are used to encode meaning. In order to promote academic language acquisition, teachers need to ensure sufficient language input, access to authentic materials and texts as well as opportunities to practise the structures, expressions and subject-specific vocabulary. For example, when studying history, the knowledge structure ‘temporal sequence’ is encountered frequently. The logical is to draw pupils’ attention to adverbs of time (e.g. first, then, next, finally) and causal connections (causes and consequences) that help to organise and produce texts, narratives and discussions (see Schleppegrell, Achugar & Oteiza 2004 for a pragmatic example).

Cummins and Man (2007, 807) describe the facets of academic English development through content-based second language instruction as a response to the unsatisfactory situation in CLIL classrooms combining Chinese and English in the following way:

To develop proficiency in academic English, students need systematic scaffolding and instruction to deal with longer texts, structurally more complex sentences, more subject-specific new vocabulary, less visual material, and more creative and higher-order thinking skills. Furthermore, students need greater exposure to readings of different types, such as narrative texts to provide a comforting linear structure for reading fluency, expository texts to provide useful repeated exposure to key vocabulary, and argumentative texts for developing reasoning and justification. Extensive reading and writing is essential for the development of academic English, which students need to acquire for academic success and higher education. Students need to be engaged in knowledge construction in both oral and written form, be supported to understand rhetorical patterns in the language and basic linguistic cues such as prefixes, suffixes and root words, and become familiar with a variety of subject-specific examples.

Extensive reading is, as reported by Cummins and Man (2007), recommendable due to encounters with less frequent loan words of Greek, Latin and French origin and overall improvement of linguistic skills (Cf. Krashen & Brown 2007 model on p. 59). In addition to this, they endorse explicit instruction in comprehension strategies and paraphrasing word meanings as well as drawing students’ attention to the intersection of linguistic features and meanings as means to improve skills of academic language (Cf. Schmidt’s Noticing Hypothesis on p. 41 and Focus on form on p. 37).

Lorenzo (2007) proposes a model for CLIL syllabus construction to make the interface between academic language and content more salient (Table 9). His graphic organiser deconstructs a syllabus into constituents of macrotopic (the overall theme, e.g. Asia), microtopics (e.g. monsoons, Asian countries), knowledge structure (e.g. synthesising), discourse model (speech events), tasks (e.g. preparing a poster on basis of group discussion) and language (lexis, functions and structures). In so doing, the need for various academic language properties, text types and corresponding speech events representing knowledge structures (such as classification, principles, evaluation, description, temporal sequence or decision-making) become

marked, alleviating the choice of needed language and enabling systematic language instruction intertwined with content tuition and, naturally, more targeted assessment.

TABLE 9. The interface of academic content and language (partly adopted from Lorenzo 2007, 269)

Macrotopic	Microtopics	Knowledge Structure	Discourse Model (Speech Events)	Tasks	Language
the overall theme	sub topics within the theme	e.g. according to the Bloom's taxonomy	e.g. comparing, stating	e.g. group discussion	lexis functions structures

Lorenzo (2013) has continued this work by producing a multilingual genre-map integrating both language and content across the curriculum at secondary level. All this indicates that the significance of subject and genre-based academic language becomes an issue to be recognised in CLIL. The NCC draft (2014) for the NCC 2016 reform underlines language awareness in the future Finnish basic education, and explicitly states that every teacher should teach the linguistic and textual conventions of his/her subject (NCC draft 2014, 22). Furthermore, the draft establishes that instruction should advance from BICS to CALP. This general principle applies to bilingual instruction as well and it is highly likely to remain in the final NCC 2016.

Research concerning academic language use in CLIL classroom environments suggests that academic-type language is seldom used or addressed. A study of 7th grade peer discussions during group work in history lessons by Nikula (2012) found very little explicit references to history, but nonetheless, the learners were able to negotiate historical meaning sufficiently. Dalton-Puffer's (2007b) study investigating the incidence of academic language functions of defining and hypothesising in Austrian CLIL contexts (grades 6–13) shows that academic language functions were addressed neither frequently nor systematically.

Furthermore, a study conducted in upper elementary classrooms in the U.S. examining academic teacher talk during content instruction concluded that academic, subject-specific language was used in minute quantities and the used terms were often opaque (Ernst-Slavit & Mason 2011). This means that the ESL learners "had limited opportunities to hear the specialized language of the content areas" (ibid., 430). It is thus important to provide practice in academic language functions from the very beginning in order to initially build up academic language proficiency and refine the mastery of functions during the subsequent years (see e.g. Haag, Heppt, Stanat, Kuhl & Pant 2013).

The majority of the knowledge today is created, disseminated and acquired in English. Mastery of academic English acquired in CLIL is likely to be valuable in extramural contexts where "English for knowledge acquisition" is a prerequisite of successful information retrieval (Dalton-Puffer 2007a, 294). Oral skills only are not sufficient for such an activity, as Dalton-Puffer (2007a, 295) puts forth; also "scanning and skimming texts, listening for gist, note-taking, processing and condensing new information and transferring such information into L1 (or L2), presenting, evaluating, expressing opinions, challenging – thinking". These skills one can acquire through systematic introduction to subject-specific academic language, its genres and registers.

3.2.2 Classroom interaction

Classroom interaction in the framework of roles of languages in CLIL illustrated in Table 7 is divided into three types of language according to focus (language of learning), approach (language for learning) and action (corresponds loosely to language through learning) in analogy with the CLIL Language Triptych (see e.g. Coyle, Hood & Marsh 2010). Classroom interaction is of relevance to this study, because assessment of language proficiency needs to be based on evidence which is partly drawn from classroom interaction. In this sub-section, I will narrow the focus onto research in CLIL classroom discourse.

There is no denying that the linguistic environment in CLIL (focus on content, focus on form) differs from ordinary language classrooms (focus on forms). Dalton-Puffer (2007a, 294) mentions two aspects of CLIL classrooms as language learning environments which she calls “CLIL Bonuses” that unburden the load of second language acquisition. First, because the goal of CLIL is also to develop content knowledge skills, it is easier to talk about content one is familiar with. That is, language is used for a real purpose to convey meaning and messages. Second, the classroom is also familiar as a linguistic surrounding, so the learners are already accustomed to typical classroom language through their L1 or the conventional language of education. They know what kind of utterances are used in given situations and are familiar with the structure of a typical didactic or pedagogic conversation which helps them to translate the knowledge from L1 to L2.

Classroom interaction has apparently not been studied in primary CLIL; the studies known to me are from lower secondary level and above. Based on the research available, it appears that the actual language discourse in CLIL classrooms is very teacher-centred regardless of the ideal of shared negotiation of meaning as “the key element of CLIL classrooms” (Bonnet 2012, 74). The classroom discourse seems to be based on *Triadic dialogue* (Dalton-Puffer 2007b) which refers to the classic *IRF discourse pattern* (teacher initiation–student response–teacher feedback) which Dalton-Puffer (2007a, 72) calls the “prototype of instructional talk”. IRF as “the prime strategy” of CLIL teachers is highlighted by the fact that no instances of teacher lectures were found by Dalton-Puffer, and the repair and correction turns following the IRF pattern were found to be frequent and integral parts of CLIL classroom interaction and negotiation of meaning (Dalton-Puffer 2007a, 90–91).

The types of the teacher questions influence the type of elicited student language. Dalton-Puffer (2007a, 93) found in her study that the majority of the questions asked by teachers were in an instructional register, and questions to which the answers were unknown outnumbered the ones with a known answer, open questions (undefined response) outnumbered the closed ones, but the questions of facts (89%) were clearly preferred over explanations, reasons and opinions (11%). Swain (1988, 81) in turn would rather see “the typical question/answer sequence found in much content teaching” replaced with tasks and activities that enable functional language use, and she would “provide learners with the motivation to use language accurately, coherently and appropriately by writing for, or speaking to, real audiences”.

Teachers' linguistic competence is likely to have a positive effect on the openness of the discourse in the CLIL classroom – and vice versa (Dalton-Puffer 2007a, 125). Takala (1992, 145) claims that flawless and fluent speech production is not the most important skill for the CLIL teacher since communication gaps and misunderstandings also take place in L1, and it might be easier for a non-native speaker to grasp the TL message of another non-native speaker. Pihko (2010, 59-61), however, reports a contrary finding in her study of the Finnish lower secondary learners' perceptions and experiences of CLIL. Her study revealed that 15% of students were disappointed by the poor language skills, especially the Finnish accent, that some of their CLIL teachers displayed. As mentioned in section 2.3, the requirements of linguistic competence for Finnish CLIL teachers are extremely high, but unfortunately they are rarely realised (for more on CLIL teacher competencies, see for example the European Framework for CLIL Teacher Education⁸; Järvinen 2012b; Nikula & Järvinen 2013).

The background education and orientation of the CLIL teacher (language teacher vs. content teacher) also influences on the amount of the TL used in the classroom, as Lorenzo, Casal and Moore (2010) concluded: language teachers tend to use more TL than content teachers who displayed more L1–L2 code-switching, whereas language assistants, mostly native speakers, pertained to conversational language closest to full immersion circumstances. Language teachers appeared to focus more on grammar at sentence level, while content teachers elaborated the textual level. The Andalusian CLIL model combining three different types of expertise (language teacher, content teacher and language assistant) appears to provide the language learners with a rich linguistic environment which is echoed in excellent learning results (*ibid.*). Enrichment of the linguistic CLIL environment with conversational groups led by a native speaker also seems to support conventional CLIL instruction (Rahman 2012).

The issue of initial CLIL teacher training and in-service training has constantly been raised in various connections and CLIL scholars have voiced deep concerns on insufficient teacher education policies and small numbers of qualified teachers (e.g. García 2009a; Dalton-Puffer 2011, Nikula & Smit 2010; Nikula & Järvinen 2013; Sylvén 2013). Even though the CLIL teacher would not be competent as a language teacher, s/he is ideally aware of the basic rules concerning bilingual classroom discourse. For example, Busch (2011, 547) remarks that those bilingual programmes that incorporate two languages “simultaneously in class in an unstructured way” tend to reinforce the dominant language because learners soon understand to follow the instruction given in their language of preference only. Consequently, it is important to structure the linguistic output, avoid immediate code-switching or translating and maintain the selected language – a principle which leads to separate language “sessions” incorporating the typical features of the language of instruction.

⁸ The document is downloadable in <http://www.ecml.at/Resources/ECMLPublications/tabid/277/language/en-GB/Default.aspx>.

English language seems to be more used in CLIL than in EFL secondary classrooms regardless of the type of talk (organisational or disciplinary), but in CLIL, the L1 was used for clarifications in meaning negotiation and to overcome communicative gaps (Nikula 2005). Evidence of paraphrasing was also found in two students' deliberate synonym seeking, e.g. people versus population (*ibid.*). Dalton-Puffer (2007a, 103) also noticed in the Austrian context that the questions asked by CLIL students in L1 pertained to language (vocabulary and spelling), whereas "real content questions" were uttered in the target language English.

Regarding the pragmatic features of English language use – orientation towards others and language use in identity and role construction – Nikula (2005) detected a notable difference between EFL classes and CLIL. The language in EFL was more impersonal and distant ("s/he-they-there-then") in imaginative circumstances treating students as apprentices in respect to English, whereas CLIL language was completely opposite: immediate and including personal investment ("I/we-you-here-now") which became visible in the student activity and initiatives taking place in authentic circumstances where students were perceived as experienced users of the language.

Informality seems to be an aspect on CLIL classroom language according to the study of Dalton-Puffer and Nikula (2006). The overall findings indicate, for example, that directness in addressing interlocutors is typical for CLIL classrooms mirroring the pragmatic and functional approach to language acquisition. However, regardless of the aspiration to establish a linguistically naturalistic and contextually meaningful environment for language learning, CLIL classrooms are still curriculum-regulated, institutional settings, where the produced language may be, according to Dalton-Puffer and Nikula (*ibid.*, 263), "far removed, pragmatically, from language used in other settings". By this notion they wish to point out that the choices teachers make regarding materials, activities, register and language are important in either ensuring or damaging students' opportunities for maximally rich language learning environment, for, as they underline, "CLIL models which propagate using materials in the foreign language but running the classroom in the L1 are not desirable" (*ibid.*, 264).

3.2.3 Language development

Language development in CLIL is probably the most discussed topic in the CLIL literature and the most studied aspect in CLIL research because it ultimately proclaims the efficacy and affordances of the method. This is not surprising, because better language development is the core aim of CLIL. Language development is also what assessment in CLIL is interested in. There are ample studies – both influential and less convincingly executed – in this field of CLIL study (see e.g. Pérez-Cañado 2012 for a critical research review). I have chosen to approach this issue from the questions raised by García (2009a, 211) that explicitly reflect CLIL concerns. Out of the questions that pertain to mother tongue, additional language, subject learning and socio-psychological issues in CLIL, I will answer the two first ones here by providing evidence from CLIL research.

The issue of L1: Will the first language develop normally despite a significant amount of instruction time being conducted in another language?

CLIL is labelled as a bilingual language programme, and therefore the instruction involves the use of the language of schooling which is – although not nearly always – the mother tongue of the learners. The role of the language of schooling (L1) in CLIL is to guarantee understanding and to facilitate interaction (Dakowska 2013), but depending on the level of education (primary, secondary, tertiary) and the amount of TL exposure in the CLIL programme, the ratio and importance of L1 varies. As was noted in the section pertaining to theoretical CLIL underpinnings (see Common Underlying Proficiency and Conceptual Reservoir in 2.3), concepts learned in one language readily transfer to the other. There are plenty of studies of cross-linguistic influence and transfer from previous language knowledge to the L2 (see e.g. Ortega 2009), but the role of L1 has not been widely addressed in CLIL research.

Ricci Garotti (2007) discusses the role of L1 in CLIL mirrored through two opposite views on L1: the first is “as much L2 as possible, as much L1 as necessary” and the second “L1 as a supporter to L2”. She presents the distinction between passive and active bilingualism reflecting the uses of L1 and L2 in CLIL lessons (Table 10).

TABLE 10. Grading of L1 and L2 linguistic objectives in passive and active bilingualism (Ricci Garotti 2007)

PASSIVE BILINGUALISM	
Aims	Linguistic use in production
Understanding of material mostly with non-verbal language	L1 prevailing in all activities
Understanding of complete verbal material	L1 in homework
Understanding of explanations and instructions	L1 in homework
Understanding of simple authentic texts	L1 prevailing, L2 in short answers
Understanding of text analysis	L1 in group work, L2 in answers or in work at school
Understanding of authentic texts of a certain length and complexity	L1 in group work, L2 in answers or in work at school
Understanding of specific authentic texts	L1 in group work, L2 in answers or in work at school, L2 in material for homework
ACTIVE BILINGUALISM	
Aims	Linguistic use in production
Answer to instructions and questions	L1 in homework, L2 only in answers
Elaboration of reasons to answers	L1 in homework, L2 in answers and in improvisations
Elaboration of simple texts	L1 in group work, L2 in plenary activities
Participation in dialogue exchanges that go beyond the mere question-answer exchange	L1-L2 in group work, L2 in plenary activities
Participation in disciplinary discussion and analysis	L2 in group work, L2 in plenary activities

Passive bilingualism entails more use of L1 in instruction whereas *active bilingualism* enhances active use of L2 rather than mere comprehension skills. The grading of linguistic objectives illustrated in Table 10 permits, according to Ricci Garotti (2007, 140), the individualisation of learning goals, and thereby also the activities or tasks which include “hints about which language to use, or how much L1 should be tolerated”. The table is also helpful in determining the decisiveness of either of the two roles of L1 in bilingual content instruction and defining the conventions of bilingual instruction as well as objectives.

The Finnish National Core Curriculum for Basic Education (NCC 2004, 271–273) is concerned of transfer from L2 to L1 and lists potential areas in which “special care must be taken so that the pupils receive adequate instruction in the school’s language of instruction”. The extent of L1 usage in CLIL may vary according to the model of provision, level of schooling and, naturally, the language skills of the learners. The effect of bilingual instruction on the acquisition of mother tongue skills (or in case of students with immigrant background, skills of the language of schooling) has also been of concern to researchers, but the obtained results are anything but alarming.

In fact, the effect of CLIL on mother tongue skills is particularly positive according to the longitudinal study of Finnish primary CLIL pupils in comparison to pupils of mainstream instruction by Merisuo-Storm (2013). The results show that from the 2nd grade upwards the differences in favour of CLIL were significant in each category under scrutiny: reading and writing skills, orthography, creative writing, writing an abstract and textual apprehension. The attitudes of CLIL pupils were also more positive than in the control group – especially among boys. Furthermore, Merisuo-Storm’s study reveals higher language awareness in CLIL pupils’ creative essays. Parallel results have been obtained elsewhere (Gebauer, Zaunbauer & Möller 2012).

CLIL programmes are occasionally referred to as selective and elitist due to admission procedures. Linguistically talented pupils’ opportunities to prosper in CLIL are certainly better, although several studies (e.g. Merisuo-Storm 2013, Seikkula-Leino 2007) substantiate that even weaker pupils may benefit from bilingual instruction. Seikkula-Leino (2007) states that pupils, regardless of their achievement level, succeeded in their study of the Finnish as the mother tongue in a similar vein in CLIL as they did in the mainstream teaching.

The Issue of L2: Will the second language really develop better if a significant amount of instruction time is conducted in it?

It is very difficult to compare the efficacy of CLIL programmes because the amount of time invested in instruction in the TL varies immensely. In that sense, the modifier ‘significant amount’ in García’s question is relevant: exposure to CLIL instruction can be differentiated at three levels: low, medium and high. One has to keep in mind that Finnish CLIL students also receive formal English instruction along with their CLIL lessons which increases the language exposure. It is a well acknowledged fact that frequency and practice effects language acquisition positively; accordingly, the more extensive and intensive the linguistic exposure, the better results obtained (see section 2.1 and Ruiz de Zarobe 2008).

In bilingual settings in general, there is powerful evidence supporting the effectiveness of additive and strong forms of bilingual education such as early total immersion (see e.g. May 2008 for a research review). Since CLIL principally differs from immersion – for instance, in its depth, linguistic surroundings and the language skills of its teachers – general conclusions or inferences regarding CLIL cannot be drawn from the research results obtained from immersion settings. Research acknowledging the unique context of CLIL has been conducted since the 1990s – the

first Finnish CLIL study was carried out by Järvinen (1999). The number of specific CLIL studies multiplied in the 2000s and they have mainly concerned language competence, subject matter competence and teacher research (Bonnet 2012, 66–67).

Language development is one of the most studied areas in CLIL research and the results in general have been very positive (see e.g. Dalton-Puffer 2008 for a research review), as can be noted in Table 11 which pinpoints the areas that are positively affected by CLIL instruction and areas where any specific gains have not been observed or are not clear. In the light of today's knowledge, certain language aspects appear to benefit of CLIL instruction more than others.

TABLE 11. Language competencies favourably affected or unaffected by CLIL (slightly modified from Dalton-Puffer 2008, 5 in the light of recent research)

Favourably affected	Unaffected or Indefinite
Receptive skills	Syntax
Vocabulary	Writing
Morphology	Informal/non-technical language
Creativity, risk-taking, fluency, quantity	Pragmatics
Emotive/affective outcomes	
Pronunciation	

When studying language development in CLIL, the research layouts tend to be contrastive: CLIL students are compared to mainstream learners studying English as a foreign language (EFL). Numerous studies report enhanced overall language performance and competence in CLIL (e.g. Järvinen 1999; Lasagabaster 2008; Lorenzo, Casal and Moore 2010; Serra 2007 and Zydatið 2012). Gains in speech production were found by Ruiz de Zarobe (2008), while Gallardo del Puerto, Gómez Lacabex and García Lecumberri (2009) detected less severe foreign accent and more intelligible pronunciation. There are CLIL investigations showing CLIL students' favourable language development in, for example, more complex grammar and sentence structures (Lorenzo & Moore 2010), academic language proficiency and coping with more cognitively demanding language test items that required grasping more subtle meanings and sophisticated grammar (Várkuti 2010), academic discourse proficiency (Zydatiss 2012), the use of achievement strategies in communication (Rahman 2012) as well as the extent and growth of vocabulary (Lo & Murphy 2010; Ruiz de Zarobe 2010; Sylvén 2010; Xanthou 2011).

The excellent research results beneficial to foreign language instruction programmes such as CLIL and immersion need, however, to be interpreted with some caution, as Dalton-Puffer, Nikula and Smit (2010b) and Merisuo-Storm (2013) highlight. In many cases, the CLIL pupils are selected according to their success in language aptitude tests; teachers are highly motivated in their undertaking to integrate the language in traditional subject teaching; and the stakeholders, mostly parents, are very keen supporters of bilingual instruction. This, as well as student motivation and the sense of being 'special', may well have an influence on the very positive research results. Rumlich (2013) uses the term "creaming effect" to describe the phenomenon where CLIL students are selected from a heterogeneous group of students and describes the accented study in English in CLIL subjects as the "preparation effect". Considering both these

effects, it is not surprising that comparative studies result in reporting flattering language learning outcomes in CLIL.

Bonnet (2012) presents some critical considerations regarding the “powerful atmosphere of optimism and almost limitless belief in the potential of CLIL” created by the political support of the European Union and the lack of substantial, evidence-based research (see also Pérez-Cañado 2012). Each CLIL context, as Lasagabaster (2008, 38) notes, “has its own peculiarities” pertaining to, for example, linguistic surroundings and country characteristics, number of languages, composition of learners and their language profiles as well as teachers’ language competences and knowledge of CLIL methodologies. The outcomes of individual CLIL programmes should be scrutinised keeping those specificities in mind to make more accurate inferences of the ‘(dis)advantages’ of CLIL programmes as a whole because the results are not always generalisable, as the findings from the Swedish context show (Sylvén 2013).

SUMMARY

The language needed for studying content is subject-specific and represents various genres of academic English CALP which is different from ordinary, everyday English BICS. The internalisation of academic language takes several years in totally immersive circumstances, and even longer in low TL exposure conditions. The conventional IRF pattern seems to persist in CLIL environments. Although the language use of CLIL learners in comparison to EFL students seems to be more diversified and extensive, there is still room for improvement in the discourse skills of CLIL students. It is suggested that teachers have a central role as language models and facilitators in that process. The research results yielded from various fields of CLIL language development are impressive, although concerns have been raised about the trustworthiness of the results. The research suggests that CLIL instruction not only enhances second language acquisition, especially perceptive skills, fluency and vocabulary but also the development of L1.

4 SECOND LANGUAGE ASSESSMENT

Assessment is, along with objective setting and actual instruction, an inseparable part of any education – including CLIL, as demonstrated in Figure 10. The overall aim of *language assessment* is to elicit a sample of language performance which makes it possible to infer the potential language use of a person in other contexts, to see which implications these inferences have on instruction and whether the objectives have been reached or need to be changed. The three basic elements of education are thus in dialogue with each other which is why none of them can be ignored.

Just as the understanding of the constructs forming the overall second language proficiency have, as portrayed in the preceding chapter, changed through decades, so have the views on what constructs in the L2 should be assessed (see Stoyhoff 2012). Bachman (2007, 43–45, 70) differentiates three chronologically overlapping and co-existing dialectics (approaches to second language assessment) from the 1960s to the 2000s that have influenced the ways second language has been and is assessed: 1) ability-focussed, 2) task-focussed and 3) interaction-focussed dialectics. Bachman (*ibid.*, 70–71) emphasises the inclusiveness of these approaches to language assessment as follows:

[T]hese different approaches have important implications and present challenging questions for both empirical research in language testing and for practical test design, development, and use. These theoretical issues also provide valuable insights into how we can enrich the ways in which we conceptualize what we assess and how we go about assessing it. For research, they imply the need for a much broader, more catholic methodological approach, involving both so-called quantitative and qualitative perspectives and methodologies. For practice, they imply that exclusive focus on any one of these approaches (ability, task, interaction), to exclusion of the others, will lead weaknesses in the assessment itself, or to limitations on the uses for which the assessment is appropriate. This means that we need to address all three in the design, development, and use of language assessments.

Bachman thus considers the three main assessment approaches as supplementary rather than exclusive. For that reason, I will not strive to give a historical overview on the developments in second language assessment, but I rather approach the issue comprehensively with an overall focus on current, up-to-date approaches to assessment that, in many cases, combine elements of each of the three dialectics.

The terms assessment, evaluation, testing and measurement are often used in overlapping senses or interchangeably as synonyms or near synonyms referring to the very same process of gathering information and providing feedback on something. In this study, I define assessment as follows by combining elements from Satterly (1983), Bachman (1990) and Lynch (2001):

Assessment is either the systematic and well-grounded process of information gathering or the product which describes the extent and/or quality of second language acquisition, its degree of correspondence with the objectives of language acquisition and its relationship with the CLIL environment for the purposes of making decisions or judgements about the language proficiency of individuals or giving feedback in order to enhance learning.

I perceive assessment as a holistic and participatory process which also includes feedback, surveying and documentation. Assessment is both the act of profiling and the end result, an account of a person's language proficiency consisting of versatile quantitative/qualitative and external/internal descriptors such as self-, teacher and peer assessments, psychometric and numerical test results, project works, portfolios, learning diaries and so forth. The assessments project the objectives that have been set for the language acquisition and are primarily intended for decision making on how to further enhance that person's learning but also for final assessment.

Thus, having defined assessment in this research context, I will now move deductively from more general, principal assessment issues towards assessment in CLIL, a specific but neglected sector in the field of CLIL research. In the first section, I will examine the assessment principles in Finnish basic education in general (4.1) and classroom assessment in particular (4.1.1), because those principles also guide assessment in CLIL. Future education and assessments are shortly looked at in 4.1.2. Alternative assessments represent formative assessments which seem to be a global trend in resistance of high stakes accountability assessment policies that prevail in many countries. I will address these alternative assessments in section 4.2. Out of many alternative assessment approaches I have selected three that are exceptionally suitable for CLIL contexts and this study: collaborative testing (4.2.1), task-based performance assessment (4.2.2) and technology-based testing with a focus on computer simulations (4.3.2) due to the computer simulation experiments conducted in this research. Section 4.4 is dedicated to the core content of this research report, assessment in CLIL.

4.1 Assessment in Finnish basic education

In Finland, it is ordained by law that assessment has guiding and encouraging functions, it should promote pupils' competences in self-assessment, and it should be versatile (Basic Education Act 1988/628, 22§). This legal norm directs the National Core Curriculum for Basic Education (NCC 2004), and it is notably seen in the final draft for the renewed NCC 2016 (NCC draft 2014). These principles differentiate the Finnish assessment culture from the assessment cultures of many other countries. This section will give an overview of assessment in Finnish basic education.

Continuous and final assessment

In the classroom context, three distinctive forms of assessment have traditionally been differentiated: diagnostic, formative and summative. Each is situated in various points in the continuum of learning, and they have different purposes. *Diagnostic assessment* prior to study provides a prognosis of learning. *Formative assessment* takes place during learning to inform the stakeholders of, for instance, the efficacy of teaching and the learning progress. *Summative assessment*, at the end of study, concludes the level of skills or competences in relation to pre-set objectives. Assessment may also have other purposes. Newton (2007, 161-162) has

identified at least 18 purposes (or rather uses) for assessment. Examples of such uses are grading, reporting, certification, placement, selection (entrance, readiness), decisions of teachers and educational programmes, accountability and national comparison (see Bachman 1990; Broadfoot & Black 2004; Cohen 1994; Newton 2007).

The current Finnish National Core Curriculum for Basic Education (NCC 2004) has not, however, adopted any of the traditional terms. Instead, it acknowledges two kinds of pupil assessment: *continuous assessment during the studies* and *final assessment* taking place at the end of the basic education. No other forms of assessment are mentioned. As a result, I will concentrate on these two forms of assessment in this section, the main emphasis being on assessment during the studies which I will refer to as *continuous assessment*. Continuous assessment has, according to the NCC (2004, 260), the following tasks:

The tasks of assessment during the course of studies are to guide and encourage studying and to depict how well the pupil has met the objectives established for growth and learning. It is the task of assessment to help the pupil to form a realistic image of his or her learning and development, and thus support the pupil's personal growth, too.

The NCC also presupposes that continuous assessment is helpful in creating awareness of one's own skills, actions and thinking as well as understanding what has been learned (ibid.). The general guidelines for assessment issued by the NCC also apply to assessment in CLIL.

When juxtaposed, the similarities between continuous assessment and formative assessment become protrusive. The task of formative assessment, called *assessment for learning* by Birenbaum and colleagues (2006), or *assessment as learning* by Rea-Dickins (2008), is to give teachers, pupils and their parents feedback on how successful the teaching and studying has been, where progress has been made and which are the developmental issues. The intention of formative assessment is to scaffold future learning, adjust teaching and to give support or additional instruction when necessary. There is ample research evidence that formative assessment is a convenient means to enhance student achievement (Black & William 1998) because it involves and activates pupils. It is transparent to its stakeholders and one part of the learning process.

The NCC (2004, 260) is in concordance with these principles, but it also highlights that "the pupil's progress, work skills, and behaviour are assessed in relation to the curriculum's objectives and descriptions of good performance". The description of good performance is available in the curriculum for each subject. Following from this, continuous assessment is criterion-based. The ultimate purpose of formative, continuous assessment is to promote pupils' learning, but, as discussed later in this section, such assessments can be used for summative purposes as well. Formative assessment represents a 'here, now and soon' approach to learning in comparison to summative assessment which is rather a retrospective 'there and then' statement of something that has passed and can principally no longer be influenced.

Instead of using the term summative assessment or *assessment of learning* (Birenbaum & al. 2006, 62), the NCC (2004) has adopted the term final assessment. The task of final

assessment, which has to be equal and comparable in national level, is “to define how well, at the conclusion of his or her studies the pupil has achieved the objectives of the basic education syllabus in the different subjects” (NCC 2004, 264). In this sense, final assessment is parallel to summative assessment. It is worth noticing that the NCC (ibid.) states of final assessment (boldfacing mine):

*In each core subject, the final grade is to be based on the performance in the final phase of basic education - that is, **in the eighth and ninth grades**. Criteria for final assessment have been prepared for all core subjects. The pupil’s performance is assessed with those criteria, on the basis of diverse evidence.*

In other words: final assessment is also criterion-referenced, but unlike one would expect, final assessment occurs at the very final stage of basic education only, not at the end of each school year. This indicates that **in the course of lower grades 1–7, the task of assessment is always formative**. What is problematic with this division between continuous and final assessments is that in the NCC (2004), there is no specific designation for such summative or final assessment that occurs at the end of the school year or study unit during grades 1–7. This problem has been addressed in the final draft for NCC 2016 which assigns continuous assessment a summative role, when necessary. It retains the two basic forms of assessment as well as their designations and purposes, but specifies that continuous assessment *also includes a description of the level of skills and competences in form of reports at certain times* (NCC draft 2014, 40). The education provider (municipality or individual school) decides a uniform outline for the school year report.

Finnish assessment culture

The current NCC (2004) uses the term ‘*pupil assessment*’ of assessment. This term will most likely be replaced by ‘*assessment of learning*’ (NCC draft 2014, 39) in the renewed NCC 2016. This concept underlines that the object of assessment is not the learner as a person but learning, which in turn emphasises the more active role of the learner as a participant in the assessment process. As Boud (2000) aptly points out, summative assessment “gives the message that assessment is not an act of the learner, but an act performed on the learner”. The NCC draft (2014, 39) lists the central features of the Finnish assessment culture as follows:

- encouraging atmosphere that promotes learning effort
- interactive, conversational conventions enhancing pupils’ participation
- support given to pupils in understanding their own learning process
- making progress visible during the whole learning process
- fairness and ethicality of assessment
- versatility of assessment
- capitalising assessment information in designing instruction and school work

These features are characteristic of formative assessment, which confirms that the Finnish assessment culture in basic education is primarily supportive of formative assessment for learning instead of summative assessment of learning.

In the assessment literature, summative assessment can be divided into *school internal and external assessment*. The former refers to using teacher-made tests for record keeping and passing the results to learners themselves and their parents, while the latter entails granting certifications and qualifications as well as accountability controls (Harlen 2005, 208). Policy-driven, large-scale assessments are often associated with external summative assessment and as a result, much of the current assessment debate in the world deals with shifting ground from summative to formative or school internal assessment practices. Following this definition, the Finnish basic education system endorses formative assessment and internal summative assessment which underpins autonomy in assessment.

It is noteworthy that the Finnish education system does not endorse assessment for accountability purposes. Schools or individual teachers in Finland are not accountable to any internal or external body of supervision; they are liable to follow the guidelines and criteria sketched in the national and local curricula which are normative documents. Moreover, fairly strict admission procedures and the requirement of a master's degree for teachers at any level of basic education guarantee the quality of education. Contrary to many countries such as the United States, Great Britain or Australia, to name a few examples, there are no national large-scale high-stakes assessment programmes regulated by law or controlled by the state for language or any other subject in Finland.

Instead of control, the Finnish system places trust in the highly-educated teaching staff and the stakeholders. The National Board of Education (2013) states, regarding the Finnish assessment policy⁹ at the basic education level:

The student assessment and evaluation of education and learning outcomes are encouraging and supportive by nature. The aim is to produce information that supports both schools and students to develop. National testing, school ranking lists and inspection systems do not exist.

The approach to assessment and use of tests categorises educational systems into two varieties: decentralised and centralised, which are biased in several features (Shohamy 2001, 29). In decentralised systems, equal opportunities and prospects for advancing in studies are available for everyone, and for this reason, a substantial number of students advance to the next levels; no national large-scale, high-stakes assessments at the end of school are administered; and teachers carry out classroom assessments for formative purposes (ibid.). On account of the principles mentioned above, Finland represents a *decentralised educational system*.

Finland has, however, attended external programme evaluations, such as the OECD-governed international PISA programme, but their function is rather developmental than

⁹ See more on the Finnish principles of assessment in the basic education in http://www.oph.fi/download/146428_Finnish_Education_in_a_Nutshell.pdf, p. 16

restrictive. Although the autonomy of educational institutions and teachers has traditionally been strong, implications of very subtle shifting towards accountability assessment or institutional evaluation can also be detected in Finland in the spirit of enthusiasm for several subsequent excels at PISA studies (Väljörvi 2012). One example of such a phenomenon is the national ranking listing of upper secondary schools composed by the Finnish media since the turn of the millennium.

Finnish assessment convention at school and classroom level

Formative, continuous assessment for learning is the current official norm in Finnish basic education concerning grades 1–7, which is why there is no reason to make a specific distinction between formative and summative assessments at that level. This is, however, a matter of definition. In practice, school year reports clearly represent some type of summative assessment in a similar manner as tests concluding a study unit. At the end of the school year, the learners are to be issued a school year report which must be based on the criteria written in the local curriculum sketched following the guidelines given by the NCC (2004, 264). Working skills (process) are assessed in each subject assessment (*ibid.*, 262). Both the current NCC (2004, 260) and the NCC draft (2014, 39) accentuate the importance of feedback and acclimatising pupils to self-assessment.

Typical for the Finnish context is that assessments, originally collected for formative purpose during the school year, can be recycled for summative purposes at the end of school year. In the NCC (2004), school internal summative assessment is considered as one form of formative assessment. The numerical or verbal assessment of a subject concerning the whole of the school year is thus grounded on diverse data instead of one final test. Therefore, continuous assessment drives dual purposes and deploys different criteria which is resourceful, cost-effective and saves both time and effort. I have modified the graphic presentation by Harlen (2005, 220) to demonstrate this principle (Figure 12). The figure could represent any subject, but the number of units is normally slightly higher.

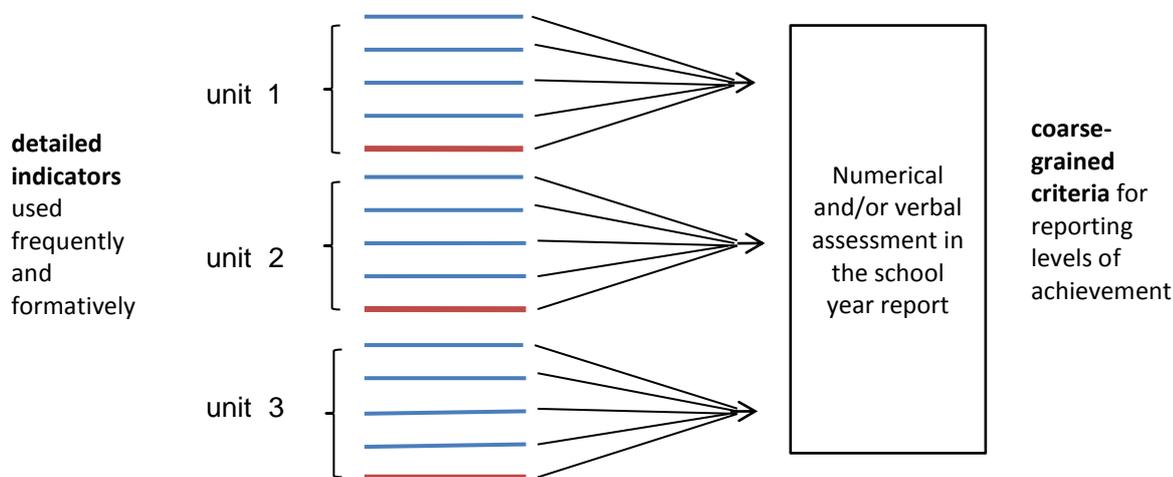


FIGURE 12. Finnish assessment convention at classroom and school level: grades 1–7

The view in the current NCC (2004) and the NCC draft (2014) for the NCC 2016 appears to be that one school year (grade) represents a syllabus within the curriculum for basic education. The syllabi are divided into school subjects (e.g. mathematics), which in turn are comprised of study units (e.g. multiplication, geometry). Each study unit normally contains several methods of formative assessments (e.g. teacher observation or formative tests) represented as thinner blue lines in Figure 12, and the unit is concluded with a larger, summative test represented as bolder red lines.

Educators in Finland have the freedom to apply any assessment method, format, content of the test, composition of scoring or system of feedback they consider appropriate. The teachers may have their own criteria in line with those stated in the local curriculum for both of these assessment forms. As a result, the validity and reliability of testing procedures may alter enormously because assessments, as Boud (2000, 153) emphasises, also convey the values of teachers and the society as well as beliefs about what is important (Llosa 2011). Regardless of that, the benefits of such an assessment convention are seen more valuable than external control. The existence of the curriculum and teachers' conforming to it as a normative document should guarantee that assessment is valid, appropriate and adequate.

4.1.1 Principles of classroom assessment

Formative assessment is often considered synonymous to *classroom(-based) assessment (CBA)* (Shepard 2000). CBA incorporates three critical dimensions of assessment: evidence, interpretation and use, exemplified in Table 13 by Hill and McNamara (2011), originally used as a research framework. The framework with relevant questions pertaining to the scope and dimensions of CBA is helpful in understanding the different aspects of CBA.

TABLE 12. Scope and dimensions of classroom-based assessment (Hill & McNamara 2011, 398)

Scope	Dimensions	
What do teachers do?	<i>Evidence</i>	What activities or behaviours are assessed? Is it planned/incidental, explicit/embedded?
	<i>Interpretation</i>	Does it target individuals, groups, the whole class?
	<i>Use</i>	Is reflection sustained or fleeting? How is assessment used?
What do they look for?	<i>Interpretation</i>	What criteria do they apply?
What theory or 'standards' do they use?	<i>Interpretation</i>	What are the values guiding assessment?
Do learners share the same understandings?	<i>Evidence</i>	What are learners' beliefs about how assessment is conducted, interpreted and used?
	<i>Interpretation</i>	
	<i>Use</i>	

The framework, however, pays attention only to teachers as assessors, while in reality learners may also act as assessors. As seen in Table 12, CBA may involve the class as a whole or target at individuals, it may be embedded in the teaching activities during instruction (dual use) or be

explicit as in tests and assignments. The duration may also alter (sustained/fleeting) and the uses of CBA are subject to variation.

The purpose of CBA is, in addition to inform the teacher of the appropriateness and success of the classroom activities and a range of other purposes, to gather evidence of the performances of individuals in that class as well as the class as a whole (Rea-Dickins 2007). That is why part of the classroom assessment is informal and remains unrecorded, whereas some of it is executed consciously and will be recorded (Rea-Dickins 2001). CBA is continuous, instantaneous and cyclical, recurring in cycles of assessment–decision–instruction, and the mode of assessment may be either explicit or implicit, which denotes that neither the teacher nor the learners are aware that assessment is taking place (Bachman & Palmer 2010, 18–19).

Implicit assessment is used for formative decisions to facilitate learning and organise the course of the lesson while *explicit assessment* is distinct from actual teaching and used for formative and/or summative purposes (Bachman & Palmer 2010). McKay (2006, 141) uses the terms on-the-run assessment for instantaneous, formative assessments and planned assessment for explicit assessment, while Hill and McNamara (2011, 403) talk about formal assessment which is planned and evident (e.g. tests and assignments), planned assessment opportunity (teaching activities also used for assessment) and incidental assessment opportunity (unstructured observation).

Classroom assessment entails a variety of strategies: e.g. incidental observation, planned observation, conferences, portfolios, contract work and projects, self- and peer-assessment and classroom tests (McKay 2006, 152–168, see also Gottlieb 2006 and Ioannou-Georgiou & Pavlou 2003). McKay (2006, 140) considers that administering tests prepared by others is not, however, classroom assessment because the test designers are not immediately concerned with the learners and the classroom context. Other implicit, instruction-embedded assessment methods are, according to Rea-Dickins (2001, 434), “teacher questioning and probing, small-group-interaction between learner and teacher, effective collaboration amongst learners themselves – whilst being observed by their teacher – as well as more formal feedback in terms of comments on learners’ oral and written work”. Rea-Dickins (ibid.) also points out that such CBA strategies potentially develop learner awareness in linguistic content and learning objectives in the given lesson and “stimulate reflection on what is being taught”.

The available evidence shows that the assessment practices of teachers vary considerably and there are a number of factors that affect the way teachers administer assessments (e.g. Stoyhoff 2012). Such factors include 1) the requirements of the education system (in Finland, the curricula), 2) parental and student expectations which are, among other things, under investigation in this study and 3) teacher expertise (McKay 2006, 142–145).

Feedback

Assessment is not completed, as Ioannou-Georgiou and Pavlou (2003, 14) stress, in the collection of assessment evidence, but “offering feedback is an integral part of the assessment

process” (see e.g. Tillema & al. 2011 for a more analytic description of the whole CBA cycle). Feedback, as pointed out by Boud and Molloy (2013) is a “slippery term” that may have different connotations to different people. This study adopts the definition by Boud and Molloy (2013, 6) with some minor additions and changes. Work in the definition may also refer to performance and skills (product and process), and standards are equivalent to objectives.

Feedback is a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work itself, in order to generate improved work.

Feedback is perceived as one of the most central methods of formative assessment (Havnes, Smith, Dysthe & Ludvigsen 2012), thus conveying that assessment information to the stakeholders is equally important. Harlen (2005, 215) explicates that “using [formative] assessment to help learning means that the students [...] have information about where they are in their learning, what steps they need to take and how to take them” (see also Hattie & Timperley 2007, 87). This denotes that praising or scores are not sufficient – students need profound, elaborate feedback, guidance and counselling in order to claim ownership in their studying and be aware of the goals and the means to achieve those goals.

The NCC (2004, 260) requires accurate assessment “based on a diversity of evidence” in which “ongoing feedback from the teacher plays an important part”. The NCC (2004, 261) states:

In addition to the school year reports, the pupils and his or her parents or other guardian are to be given assessment feedback adequately and in a diverse manner. Information is to be provided about the pupil's progress and strengths, as well as those areas of learning that need improvement. Assessment feedback can be provided by immediate reports, different types of notices, assessment discussions, or other means.

However, the NCC does not dictate the form of feedback more closely, nor does it influence the form of the school year report which can, until the eighth grade, consist of numerical assessments, verbal statements and accounts or a combination of those two.

Types of feedback can be categorised in different ways. For example, Hattie and Timperley (2007) distinguish four foci for feedback: 1) feedback about the task (e.g. corrective feedback; how well the task is accomplished or performed), 2) processing of the task (e.g. strategies), 3) self-regulation (commitment, control and confidence) and 4) the self as a person (e.g. ‘good job’). From the broad perspective, there are two kinds of feedback: immediate and delayed, occurring after the evidence gathering and interpretation; both may involve record keeping, which is vital for more formal reporting purposes (McKay 2006, 152). Immediate feedback resembles a formative practice of assessment; delayed feedback can be associated with summative practices.

Hill and McNamara (2011) distinguish two basic types of immediate feedback they identified in their CBA study: person-referenced and task-referenced feedback of which the latter can be divided into confirmatory, explanatory and corrective feedback (Table 13). Person-referenced feedback is targeted at the learner’s ego; it is often comparative in nature and can take “the form of reward or punishment, approval or disapproval” (ibid., 406). Confirmatory feedback is often related to single correct answer situations; the explanatory foregrounds success

in performance; and corrective feedback draws the learner's attention to the gap between expected performance and actual performance (ibid., 406–407).

TABLE 13. Feedback types (Hill and McNamara 2011, 408)

Feedback type	Description
a) <i>Person-referenced</i>	Non-specific; affective or conative; comparison with peers
b) <i>Task-referenced</i>	
Confirmatory	Ticks; repetition of correct response; 'moving on' without comment
Explanatory	Highlights successful aspects of performance
Corrective	Identifies gaps between performance and expectation

Studies in SLA show that corrective feedback particularly correlates with higher learning outcomes in linguistic features (e.g. Havranek 2002; see also Hattie & Timperley 2007 on efficacy of different types of feedback). The language teacher thus needs what Edenlobos and Kubanek-German (2004, 259) call “diagnostic competence – defined as the ability to interpret foreign language growth in individual children”. They argue as follows, especially concerning primary foreign language education:

Teachers will be expected to engage in more one-to-one contact with students, rather than addressing the class as a whole. In order to do so teachers must be able to describe and interpret the individual student's emerging foreign language competence.

The preliminary description of a language teacher's diagnostic competence includes, for instance, various types of observations, common knowledge about children as language learners and various testing abilities which are helpful for teachers in, among other things, “assisting students in interpreting feedback” (Edenlobos & Kubanek-German 2004, 278–280) Research shows that students appreciate discussions of their own learning with their teachers, but complain about receiving too little feedback, while teachers state giving feedback, but that students seem not to grasp it (Havnes & al. 2012). The process of feedback exchange requires bidirectional, reciprocal communication between students and teachers in order to meet the needs of both so that both parties are activated (Black and Jones 2006) – this implies that also the teacher can be a recipient of feedback.

Rea-Dickins' (2007, 517) portrayal of good formative assessment practice is an appropriate way to conclude this sub-section.

Good teaching – where teachers respond to learners' language learning and needs, with different types of feedback of an appropriate kind, of learner involvement through collaborative learning activities and self- and peer-assessment, with ample opportunities to language practice – implies good formative assessment practice.

This description highlights the multimodality of assessments, participatory and social practices as well as the importance of linguistic output in second language acquisition in education and assessment.

4.1.2 The 21st century education and assessment

The world has changed and will continue changing at an extremely rapid pace which exerts pressure on the school institution to sustain that pace in order to educate citizens that are able to function meaningfully in future society. This requires skills and knowledge that the present day school system for various reasons is unable to provide. This undoubtedly requires adjustment from the educational field. It is no longer sufficient to only cultivate knowledgeable individuals, but students also need to develop their social skills, problem solving and higher order thinking as well as critical thinking skills in order to cope in the 21st century satisfactorily.

The current draft of the foundations for the new Finnish basic education curriculum (NCC draft 2014) lists several aspects that form the new school culture; among them are language awareness and plurilingualism. These principles underpin celebration of cultural diversity and underline that every single adult and teacher within the school community acts as a linguistic model. Tuition of subject-specific language is highlighted. It also names several subthemes as future comprehensive skills (NCC draft 2014, 14–18):

- thinking and learning to learn
- cultural competences, interaction and expression
- self-provision and everyday skills
- multiliteracy
- competences in information and communications technology
- skills needed in the work life and entrepreneurship
- participation, societal contribution and development of sustainable future

These themes are compatible with other listings of 21st century skills (see e.g. Binkley & al. 2012, 18–19), although the terminology used for multiliteracy appears to vary.

New literacies are particularly widely acknowledged as an issue to be addressed in future education. Warschauer (2003, 111) uses the generic term electronic literacies to refer to several forms of new literacies such as “computer literacy, information literacy, multimedia literacy and computer-mediated communication literacy”. These literacies should be increasingly addressed in school contexts and used in connection with assessment as early as possible in anticipation of the future educational changes which have already occurred in real life (Cf. 3.2.1). Norrena (2013) argues that in order for this to happen, teachers should first activate these future skills in themselves and then convey them to the learners. However, other possible future prospects have also been examined.

The future barometer of Finnish education in 2030 by Linturi and Rubin (2011, 19) discusses “likely futuribles”, i.e. several challenges, trends and prospects that shape the world of teaching and learning in the next two decades. In other words, the barometer envisions the operational models, skills, practices and schemes needed in and for the 21st century education. The future theses produced through a Delfoi method were appraised by experts who gave their arguments and comments based on which the theses were categorised into three main groups:

disputed issues, further dialogue needed and consensual issues. Among the interesting outcomes were that learning environments are, according to the majority of the panellists, likely and desirable to change (classroom expansion outside the school building as well as virtual knowledge webs and reality as parts of learning environments); social media will also have an important role in education; one future scenario paints the picture of basic education in English being available for all, while another emphasises authentic learning (Linturi & Rubin 2011).

The thesis “Assessment and feedback will be primarily targeted at pupils’ meta- and deep skills of learning and occurs mostly within dialogue between pupil and teacher”, also including multimodal assessment, participation of parents and description of larger entities of competences rather than fragmental knowledge and skills, received divided opinions. Many Finnish experts, however, according to the future barometer (Linturi & Rubin 2011, 32), are hesitant to engage in one assessment system; others are concerned of the arduousness and costs of a new system; some are content with the current excellent Finnish accomplishments in PISA studies and see no reason to deviate; and a few are open to assessment innovations.

Reconceptualising assessment is not a surprising action in the light of previous visions because the assessment practices shape the adopted learning approaches (Dochy 2001, 16); when trivia are required, students memorise such information instead of concentrating on constructing larger knowledge bases. In order to meet the requirements of the modern world of today and the future society, we must find and construct a new epistemological essence of education and its constituents, including assessment, for it is an integral and inseparable part of education. These future perspectives together with critical language testing as a counter movement against accountability policies (see e.g. Shohamy 2001) posit new approaches to assessment which are collectively called ‘alternative assessment’.

SUMMARY

Finnish basic education acknowledges two purposes of assessment: continuous, formative assessment during the studies and final assessment, which refers to the summative assessment at the very end of basic education. Assessment at lower levels is always continuous, and it seeks to enhance learning instead of merely discovering the state of art in achievement. The evidence gathered in formative assessments throughout the school year is typically used for school year reports. Contrary to many other countries, there are no assessments for accountability purposes in Finland which therefore represents a decentralised assessment system. The National Core Curriculum gives the guidelines for assessment which is expected to be constructive and based on diversified evidence, but educational establishments and individual teachers are left with considerable latitude in deciding suitable assessment and feedback practices in their own contexts. Therefore, classroom-based assessment is a relevant approach to assessment. Research shows that formative assessment and feedback has a significant role in learner achievement. Future education needs reconceptualisation due to massive changes in the society which requires new skills; these future educational demands also effect assessment.

4.2 Alternative assessment

Critical language testing in particular, but also the prospects of 21st century education, has initiated demands for assessment reform and new qualitative approaches called *alternative assessment* or a *new assessment culture* (Lynch 2001; Dochy 2001; Shohamy 2001). The definition of alternative assessment depends on, as Fox (2008, 97) articulates, what it is an alternative to. Most often it is contrasted with traditional norm-referenced testing which denotes psychometric, discrete-point testing producing numerical scores and rankings (e.g. Fox 2008, 97; Lynch & Shaw 2005, 263; Shepard 2000, 4), and accountability practices. The shift from assessment of learning (large-scale single-trait accountability measurement) towards assessment for learning thus reflects learning culture which is overtaking the exam culture (Hamp-Lyons 2007).

Similar to the term CLIL, alternative assessment is also perceived as an umbrella term for various approaches (see Fox 2008, 97–98 for a discussion of diverse perceptions) which have emerged along the new assessment paradigm. Alternative assessment is the most frequent term used in the current assessment literature (Suomela-Salmi 2010, 209), although Brown and Hudson (1998, 657) would prefer the term *alternatives in assessment* since the methods are not totally new – the novelty is mere a question of perspective.

In general, alternative assessment means the embedding of policies other than accountability into assessment and shifting from summative assessment of learning to formative assessment for learning, which as such is not relevant for the Finnish context. Nonetheless, alternative assessment, in its pursuit of visualising and authenticating the quality of learning rather than quantifying it, may have a positive impact on the Finnish education as well, especially in the form of enriched teacher-driven assessment methodology. Considering the Finnish context, alternative assessment can be seen as a continuation or an extension of normal formative assessment practices providing various perspectives. Therefore, in order to diversify assessment methods, other means than traditional paper-and-pencil tests are required.

Lynch and Shaw (2005, 265) rightfully note, however, that “the true soul of assessment lies not in the components or tools, but in the perspective or set of assumptions motivating their use”. In respect of language proficiency, alternative assessment “takes the view that language ability and use can best be understood as realms of social life that do not exist independently of our attempts to know them” (Lynch 2001, 362). This means that language proficiency cannot be tested in isolation (as is often the case in paper-and-pencil tests), but new, communicative and activating testing methods are needed – also in CLIL. Therefore, this study sets out to experiment with computer simulations in CLIL contexts. The following characterisations, synthesised from Dochy (2001, 16–18) and Brown and Hudson (1998, 654–655), are associated with alternative assessment and overlap with purposes of formative assessment.

Alternative assessment

- is embedded in instruction and enhances learning
- is multimodal
- activates learners to perform, create, produce or do something
- is keen on both product and process
- is fairer and less threatening to learners
- uses real-world contexts, e.g. simulations
- uses meaningful tasks and activities that involve multiple skills, especially higher order thinking skills and problem-solving skills
- allows assisting tools and often finds time irrelevant
- is multiculturally sensitive
- documents and provides information of learners' progress, strengths and weaknesses
- adopts a holistic view: reports multiple traits rather than single attributes
- ensures human judgement either by instructors or learners themselves instead of machines
- encourages transparency and openness in disclosure of criteria and standards
- is better suited for new learning environments
- requires new roles of teachers

Various realisations of alternative assessment are, for instance, oral presentations, debates, exhibitions, collections of written products, constructions and models, experiments and inventories of student work and behaviour (Darling-Hammond 1994, 5–6). So are composition tests, cloze and cloze elide tests, c-tests, simulated oral proficiency interviews, portfolios, role play tests, group tests, performance assessments, task-based tests, diaries, conferences, self-assessments, video and audiotapes, learning logs, checklists, journals, teacher observation and peer assessments (Brown & Hudson 1998, 657). In addition to these, Fox (2008, 97) also lists simulations, diaries and inquiry-based learning projects as types of alternative assessments. As becomes evident, the variety of these methods is enormous and as a consequence, alternative assessment has also attracted criticism.

In particular, concerns of validity and reliability of the procedures have been voiced. For example, Rea-Dickins and Gardner (2000, 236) raise the validity issue in teacher assessment they discovered in their study of classroom assessment at primary level. They identified a number of potential sources of error related to the inferences teachers make on basis of pupils' language samples. Among these parameters were diverse assessment contexts: degree of task preparation, differences in difficulty level, content or interlocutors in the assessment activity and the limits placed on the assistance the pupil receives during testing. These factors may affect students' language performance and impinge on the "opportunities to demonstrate their language ability, as they may affect both the type and amount of language elicited and subsequently included as a part of their language development profile" either as a false negative (proficiency

gauged too low) or false positive (too high) inference (ibid., 236–239). Consequently, it is of utmost importance that teachers are aware of reliability (consistency) threats and attempt to create homogeneous assessment contexts to every test taker regardless of the technique of alternative assessment.

Lynch (2001) highlights that the structural design, uses and interpretations made are critical for valid and ethical assessment, not the use of alternative assessments per se in replacement of traditional assessments. Furthermore, the debate revolving around the pitfalls of alternative assessments is keen on economic matters, implementation, unintended consequences, insufficient assessment expertise and fears of increasing control and surveillance (Fox 2008, 104). For instance, performance assessments are, according to Brown and Hudson (1998, 662) relatively difficult, expensive and time-consuming to administer. They also discuss the potential reliability problems caused by “rater inconsistencies, limited number of observations, subjectivity in the scoring process, and so on”. They continue by viewing test security and pinpointing a number of factors undermining validity: “(a) inadequate content coverage; (b) lack of construct generalizability; (c) the sensitivity of performance assessments to test method, task type, and scoring criteria; (d) construct underrepresentation (i.e., the problem from generalising from a few observations to the whole spectrum of real-life performances); and (e) construct-irrelevant variance (i.e., performance-characteristics that have nothing to do with the students’ real abilities)” (ibid., 662–663).

All assessment, including alternatives, are value-laden and shaped by the society in which it occurs, thus defining what is worth learning and knowing (Fox 2008, 98). Although Lynch (2001, 360) claims that alternative assessments do not seek evidence to make decisions, Fox (2008, 100) regards that fairness is in jeopardy because “they may also be used for traditional purposes, namely, to sort, sanction, and control, and to define to which (and whose) knowledge counts, and which (and whose) does not” (ibid., 100). This is an especially relevant notion in the multicultural society we are living in, but as the description of the Finnish assessment context (see 4.1 and Figure 12) demonstrates, formative alternative classroom assessments can be used for summative purposes given that the criteria for assessment are available and the common ground for assessment has been defined.

With the above background information on the characteristics, benefits and pitfalls of alternative assessments in general, I now turn to the actual modes of alternative assessments. Some of them can be considered as genuinely new, some may evoke new ideas and perspectives to assessment. Numerous prefix labels have been incorporated into alternative assessment: progressive, authentic, restorative, multimodal, continuous, integrative, sustainable, collaborative and dynamic, for instance. Although all of the following, selected modes of alternative assessment share a few features, most notably authenticity and learner-centeredness, they nonetheless also exhibit their own distinctive features.

The modes of alternative assessment I have chosen under closer scrutiny are collaborative testing, task-based performance assessment and technology-based language assessment.

However, because the field of technology-based assessment is extremely wide and constantly evolving, I will specifically concentrate on computer simulations. Task-based performance assessment will be treated in slightly more detail because it has strong implications for both collaborative testing and computer simulations. I will give some theoretical background for each mode of assessment, a brief introduction based on literature or research and examples of their practical uses. The benefits and shortages of the modes will also be addressed.

The core criterion for being included in this sub-section is that these modes of assessment have, in my opinion, relevance for assessment in CLIL. Furthermore, they provide insights in constituting assessment theory for CLIL, and they can be applied either as such or in a modified version. If not transferred to the classroom practice, they may provide some insights and perspectives worth considering in CLIL contexts.

4.2.1 Collaborative testing

Collaborative testing, following the guidelines of collaborative learning and learning in social contexts, refers to assessment situations where two or more learners are working together in teams to produce test item responses through discussions and co-operation. It is thus grounded in the Vygotskian sociocultural theory (Vygotski 1978), and I perceive it as a potential alternative in CLIL because “CLIL is about learning by construction, rather than learning by instruction” (Wolff as cited in García 2009a, 213). Collaborative testing is also occasionally referred to as a group quiz or a small group assessment or work. According to Kapitanoff (2009), research has generously shown the benefits of collaborative testing situations in the form of better scores; her own study on college students combining both individual and pair test sections showed that the mechanisms leading to enhanced performances were related to three cognitive processes: better recollection of information, enhanced ability to organise and construct information and fruitful discussions. Furthermore, self-reported anxiety was reduced.

The connections between test performance, test anxiety and quality of discourse were examined by Pandey and Kapitanoff (2011) in a collaborative exam of college students. They found out that students most likely to suffer from test anxiety, according to initial anxiety scores, benefitted more from the collaborative test setting than others. Furthermore, a positive connection between high interaction scores and better performance was discovered; that is, the more negotiation and conversation, the better the academic achievement. The principles of collaborative testing can be applied in a range of school subjects, as the following two examples show.

Berry and Newman (2002) challenged the traditional view of testing in which knowledge is presented by the teacher in the classroom and repeated and reproduced again in tests (assessment-driven instruction) by designing a collaborative set of mathematics assessments eliciting transferable problem-solving and mathematical-thinking skills at college level. The mathematical modelling course was designed to include plenty of collaborative learning and the

assessment was composed of active participation in class discussions, an oral presentation, a team test, an individual test and a poster. This approach was embraced by students, many of whom considered collaborative testing more 'real-world'-like (authentic) and relaxed; furthermore, collaborative testing enabled students to use their metacognitions and subject-specific language – teamwork gave students the opportunity to 'speak mathematics' (Berry & Newman 2002, 646).

In language education, group discussion is a fairly common method of assessment, while pair or group writing has received less attention. The study of Wigglesworth and Storch (2009) conducted at university level looked at the differences in fluency, accuracy and complexity of argumentative essays produced by EFL students in pairs and individually. Their findings show that collaborative writing, "pooling of linguistic knowledge", allows for more accurate texts – a feature which the researchers assume to be transferable to other collaborative test settings as well (Wigglesworth & Storch 2009, 460).

Jensen, Johnson and Johnson (2002, 165) suggest that collaborative testing may be best suited for assignments and situations that meet the following prerequisites:

- A list of learning objectives is premeditated.
- The initial list is short to ensure successful onset, but will be expanded and the level of difficulty increased during the course of study.
- The significance of the grade is not too strong.
- The focus of learning is on terms and procedures.
- Group assignments are accompanied with prior or subsequent individual tests to ensure every group member is learning.
- The grading is based on predefined criteria (criterion-referenced assessment) rather than norms (norm-referenced assessment).

The benefits of collaborative testing are many, as the previous paragraphs demonstrate. Among these benefits are enjoyment (Pandey & Kapitanoff 2011) and preparation for demands of future professional life which often requires team work skills (Berry & Newman 2002; Wigglesworth & Storch 2009). Ambivalent factors in collaborative testing, according to Wigglesworth and Storch (2009, 446) are: 1) unequal participation in group work, 2) marking dilemmas (individual or group score), and 3) provision of reciprocal, group-internal feedback.

As to the marking dilemma, Jensen, Johnson and Johnson (2002, 165) obtained an interesting research result which indicates that positive interdependence (collective marking, i.e. same grade for all group members) rather than no interdependence (individual grade for each member) within a group quiz results in "highly effective learning experience". This inference was made because "students were interested in each other's learning, paid close attention to each other's progress, and engaged in considerable promotive interaction to ensure that all group members mastered the assigned information and procedures" (ibid.). In other words, by supporting each other the students created a common responsibility for the group's achievement,

for the grade was chosen randomly – one random group member represented the group as a whole.

Collaborative tests serve as “a learning tool” or “an occasion for learning” in that they invite negotiation of linguistic forms (Ewald 2005, 580) as well as content knowledge and strategies. Good interpersonal skills, as Kapitanoff (2009, 66, 69) remarks, appear to contribute to the positive experiences on collaborative testing which is why it is very important that teachers provide models and tools for various conventions of group work in the classroom. Introverted and shy pupils may even suffer from the social settings and feel overpowered. Although all research examples given here were from higher education where the language proficiency of students is normally substantially high, there is no reason not to apply collaborative assessment methods in primary classrooms. When language and test responses are co-constructed in social interaction, possible deficits in language proficiency and/or subject knowledge are more likely to become reciprocally overcome.

4.2.2 Task-based performance assessment

Assessing language ability or language competence is not an easy undertaking because, as the iceberg model of Cummins (see p. 32) demonstrates, only part of the common underlying proficiency is visible or verifiable. Where language ability and competence are complicated, hidden systems, language performance or language use can be observed and substantiated. In chapter 3 it was established that language proficiency is embodied and demonstrated in language performance. Tasks, in turn, can be incorporated in diverse educational settings to elicit samples of language proficiency for assessment as in the computer language simulations, which are basically comprised of sequenced tasks, experimented with in this study. As a result, *task-based performance assessment* (TBPA), referring to assessment based on tasks designed to elicit language performance, is investigated in more detail.

TBPA has its origins in performance-based assessment (PBA) created for occupational assessment of immigrant L2 speakers in the 1960s (McNamara 1996, 1) and task-based assessment (TBA) which is informed by task-based teaching (see e.g. Willis & Willis 2007). Both performance assessment types incorporate tasks as a test feature. Language performance tests reflect a strong emphasis on communicative language use (e.g. Brown, Hudson, Norris and Bonk 2002, 6); as a result, they often pertain to the assessment of oral skills. Task-based performance tests reflect the view of a language as a social construct because “language skills are assessed in an act of communication”, especially in speaking and writing, and “elicited in the context of simulations of real-world tasks in realistic contexts” (McNamara 2000: 6). Assessment of communicative language use is, as Skehan (1998, 290; 2001, 167) points out, difficult to conceptualise because it needs to be translated from theoretical models to actual performances. To overcome this shortcoming, he proposes a more pragmatic model for oral test performance portrayed in Figure 13.

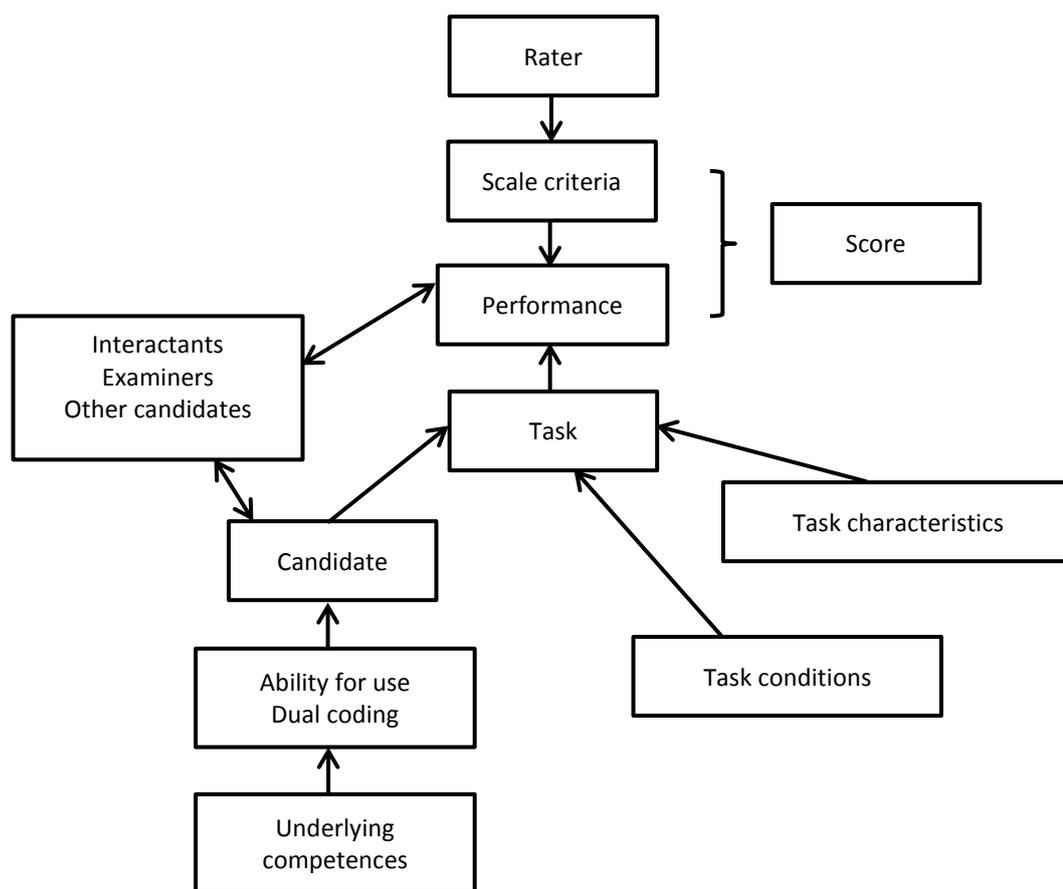


FIGURE 13. A model of oral test performance (Skehan 2001, 168)

The programmatic Skehan model indicates that the intricate interplay between model constituents has an impact on test performance resulting in the actual proficiency not being the sole factor influencing the score (see also Messick 1994; Bachman & Palmer's (2010, 38) model of reciprocal language use). The test designers and raters need to be aware of these interconnected factors. The Skehan model makes the interactive circumstances and the underlying abilities of the test taker more salient and incorporates task conditions and rating issues.

Grounding the TBPA on existing models of competence-oriented language proficiency is important along with considering the ability to actually use those competences. Ability for use, according to Skehan (1998, 171) “draws upon dual-coding capacities and organizes the way processing is adapted to performance conditions”, and it is more than strategic competence. Dual coding refers to two language processing systems available to the interlocutor: exemplar-based (drawing lexical items from memory) and rule-based (organising lexical items in syntactic forms); the former is activated, for example, when time is scarce but contextual scaffolding is given, the latter in opposite cases (Skehan 1998, 88–91).

There is plenty of evidence that the actual performance of an individual operates on the basis of three performance areas which are distinct, appear to compete with each other for the interlocutor's attentional resources during the performance and “represent different stages of the learning process” (Skehan 2003, 393). These performance areas, which emerge in order, are 1) complexity (usage of more advanced language), 2) accuracy (derived from growing control and

concerns to avoid errors) and 3) fluency (speech production without unnecessary pausing and interruptions at a normal rate in real-time), collectively referred to as CAF (Skehan 1998, 172–173; 2003, 393; 2009, 510). Skehan (2009, 528) concludes based on a research review that there is a need for a fourth, supplementary dimension of lexis or lexical performance. The task performance thus activates a number of linguistic sub-areas.

Tasks in TBPA

Agreement on the definition of a task has been difficult to reach among researchers, as Ellis (2003, 2–9) sketches. The consensus view seems to be the following (*ibid.*, 9–10):

- A task is a work plan that defines the student activity and indicates the needed content.
- The focus of a task is primarily on meaning and it “seeks to engage learners in using language pragmatically rather than displaying language [...] and to develop or assess second language proficiency through communication which requires a motivating information gap”.
- Real-world, authentic language use is elicited in the task.
- A task may involve one or any combination of four basic language skills of speaking, listening, reading and writing.
- Cognitive processes influencing but not determining the language are activated and employed in the task.
- The communicative outcome can be explicitly defined.

Tasks may vary in terms of design variables which are, according to Ellis (2012, 200–202), for instance:

- focus (unfocussed – focussed on a specific feature of language),
- mode (input-providing – output-prompting),
- gap type (information – opinion – reasoning gap),
- openness (closed – open regarding the number of outcomes),
- complexity (language of here-and-now – there-and-then or single – dual activity) and
- familiarity (known – unknown topic).

There are also other variables such as number of participants, preparative instruction of linguistic forms, receptive and productive activeness of the test taker(s), allocation of roles according to first language or other learner attributes such as personality, opportunity for strategic and online planning, repetition of the task and post-task reporting (Ellis 2012, 200–202). A more elaborate TBPA task construction model, the Triadic Componential Framework by Robinson and Gilabert (2007), will not be discussed here, but might prove useful for those interested in task design. Also, Bachman and Palmer’s (1996, 47–57; 2010, 66–82) framework of test conditions and test design allows for purposeful and efficient analysis, comparison and sequencing of tasks within

performance tests which often are essays, simulations of various forms such as role plays and oral interviews or discussions.

Pierce (2002, 2) outlines a number of profits made from implementing well-constructed assessment tasks in comparison to traditional assessment methods. For instance, TBPA 1) provides comprehensible input; 2) employs meaningful and naturalistic context-embedded tasks through hands-on or collaborative activities; 3) shows the gaps and gains in students' knowledge through diversified assessment tasks; 4) supports both cognitive and linguistic needs; 5) meets individual needs flexibly; 6) uses criterion-referenced assessment for judging students' work; 7) provides feedback for further improvement by pinpointing weaknesses and scaffolds self-esteem by indicating strengths; as well as 8) generates descriptive data that can be used for improvement of instruction. Furthermore, she mentions the visible, shared criteria and self-assessment which both contribute to student motivation as the key assets of TBPA.

Three different types of TBPA can be differentiated: products, performances and process-oriented assessments (Pierce 2002). Examples of these are listed in Table 14. They vary in terms of focus and implementations. Products are concrete items produced by the learner, whereas both performances and processes provide a channel to examine the skills and knowledge, ways of thinking or affective factors that influence learning and its outcome.

TABLE 14. Types of task-based performance assessments with examples (based on Pierce 2002, 2)

Types of TBPA	Characteristics	Examples
Products in TBPA	works produced by students providing concrete examples of their application of knowledge and skills	writing samples, projects, art or photo exhibits and portfolios
Performances in TBPA	students demonstrate their skills and knowledge under direct teacher observation	oral reports, skits, role plays, demonstrations and debates
Process-oriented TBPA	provide insights into student thinking, motivation, reasoning, reflection and learning strategies	think-alouds, self-assessment checklists or surveys, learning logs and individual or pair conferences

It is fair to state that TBPA is chiefly evidence-based assessment. The learners produce something visible or audible for the raters to verify. A more thorough investigation of linguistic features can follow language samples in several manners. Typical methods for assessing complexity are amount of subordination and lexical richness i.e. type-token ratio; accuracy is measured by number or percentage of error free clauses and target-like use of a specific grammatical feature; fluency is examined by measuring syllables per minute, mean length of pauses and number of repetitions (Ellis 2012, 207). Such investigations are likely to be applied to individual task performances which are affected by task characteristics.

The substantial body of task-based research allows for some generalisations regarding the influence of task characteristics on the CAF performance areas (complexity, accuracy and fluency) of a received language sample. I have collected these generalisations in Table 15 from Skehan (2003; 2009) representing both what he calls earlier and newer research. This is valuable information for assessment task designers navigating the expected outcome of the assessment.

For instance, when greater fluency is desired, the tasks should be structured, contain concrete, familiar information possibly pertaining to personal issues, and the test taker should deliver a monologue and have pre-task planning time.

Regarding the assessment of the interactive task performances in dyads or groups (collaborative testing), in addition to discourse analysis, three principle approaches based on interactionist-cognitive and sociocultural theories of SLA are used to “examine orderliness, structure and sequential patterns of the interactions that arise in the performance of different tasks” (Ellis 2012, 204). These approaches are: 1) negotiation of meaning, 2) language-related episodes and 3) focus-on-form episodes (ibid.).

TABLE 15. The influence of task characteristics on performance areas (Skehan 2003; 2009)

Task characteristics	Influence on performance areas
structured tasks	clearly greater fluency, tendency towards greater accuracy
concrete or familiar information	greater fluency and greater accuracy, markedly greater complexity of language
personal information exchange tasks	accuracy and fluency are raised, but not complexity
narrative tasks	higher complexity, but lower accuracy and fluency
outcomes requiring justifications	markedly greater complexity of language
interactive tasks	advanced accuracy and complexity
monologue tasks	higher fluency
tasks requiring information manipulation	higher complexity
pre-task planning time	greater complexity and fluency
post-task (public) presentation or transcription	raised accuracy

Furthermore, Ellis points out that a primary task focus on meaning does not exclude a secondary focus on form (see 2.2). This is an interesting point from the CLIL perspective, signifying that conveying – and especially assessing – messages or content can include foci on both meaning and form.

4.2.3 Technology-based assessment: computer simulations

Technology-based language assessment has its roots in *computer-assisted language learning* (CALL) which in turn is an adjustment of SLA to the technological changes in the modern world (see Chapelle 2001 or Kern 2006 for an earlier review and Stockwell 2012 for more recent developments). CALL is defined by Stockwell (2012, ii) as “an approach to teaching and learning languages that uses computers and other technologies to present, reinforce, and assess material to be learned, or to create environments where teachers and learners can interact with one another and the outside world”. This view is adopted in this study.

There are influential voices in the assessment field calling for a more rapid and drastic change in the modernisation of assessment. For instance, Birenbaum and a number of colleagues (2006, 61) claim:

Whilst modern societies have dramatically changed with the advent of technological changes and the development of information technology systems, most schools still rely on teaching according to an out of date information transmission model. Current assessment practices fail to

address the needs of today's learners and the modern, complex and globalised societies that they are a part of. Teachers need to be supported in changing their current practices in order to assess learners in ways that reflect the future needs that will be placed upon them.

In the light of modern education outlined in the previous section, this demand is justified and acknowledged in this study which attempts to broaden the range of assessment alternatives in CLIL by experimenting with technology in assessment. One area within CALL is *computer-assisted language testing* (CALT), a merited example of which is the international Test of English as a Foreign Language (TOEFL iBT®)¹⁰ which can nowadays be performed online in most countries. Another, project-based innovation is DIALANG, a diagnostic, solely computer-based test resting on the CEFR scales (see Alderson 2005).

Chapelle and Douglas (2006, 39), in examining the advantages and limitations of CALT in comparison to test method characteristics, conclude that “[c]omputer-assisted language tests are different from other types of language tests in the method characteristics of input and response, the interaction between them, and assessment”. Table 16 presents advantages and limitations of CALT in various test characteristics in which the differences to other test types are most marked. This binary presentation demonstrates that the use of technology is not an end in itself but needs to be contemplated thoroughly in terms of possible advantages to be gained through CALT.

TABLE 16. Test method characteristics and CALT advantages and limitations (Chapelle & Douglas 2006, 23)

Test method characteristics	CALT advantages	CALT limitations
Physical and temporal circumstances Location, time, personnel	CALTs can be taken at many convenient locations, at convenient times, and largely without human intervention.	Security is an issue in high-stakes tests; equipment not standardized nor universally available; IT expertise required for establishment, maintenance.
Rubric/Instructions Procedures for responding	Test tasks are presented in a consistent manner for all test takers and instructions and input are presented automatically and uniformly, making for enhanced fairness.	Different levels of instructions, voluntary help screens, different languages of instructions can detract from uniformity
Input and expected response Features of the context: setting, participants, tone Format: visual/audio/video	Multimedia capabilities allow for variety of input and response types, enhancing contextualisation and authenticity.	Input and response types are limited by available technology.
Interaction between the Input and Response Reactivity: reciprocal	Computers can adapt input in response to test takers' responses and actions, allowing for computer-adaptive tests and rapid feedback.	Interactiveness is more controlled than certain other formats; computer's ability to sample fairly may be limited; computer-adaptive tests are expensive to develop.
Characteristics of assessment Construct definition Criteria of correctness Scoring procedures	Natural language processing (NLP) technology allows for automated scoring of complex responses, affecting the construct definition, scoring criteria, and procedures.	NLP technology is new, expensive, and limited, thus creating potential problems for construct definition and validity.

¹⁰ See <http://www.ets.org/toefl/ibt/about> for more information.

Among the most obvious advantages of CALT are, for instance, flexibility in terms of time and location of performance, fairness of testing in terms of equal treatment for all test takers, multimodality, rapid feedback and automatic recognition of correct answers when scoring. Among the disadvantages are security issues, need for technical support, limitations in technological capacity and ability to interpret nuances, controlled test interaction and expensive production.

Computer simulations are one form of CALT. There seems to be plenty of discussion and research reports on various simulations but very few offer a sound definition of a computer simulation, which is probably due to the vast variety of them used for diverse purposes. At its simplest, a simulation refers to the replication and mimicking of the real, surrounding world, while a computer simulation is a simulation executed by means of a computer. Computer simulations take place in virtual, digitalised environments, and the simulation experience is composed of the executor's sensory perceptions, cognitive processes and imagination (Wewer 2013b). Laurillard (1992, 164) remarks that the essential character of a computer simulation is that "they will enable students to experience a model of the world more directly, and thereby formulate a better conceptual understanding of it".

In literature, simulations are often handled together with gaming and role play, and they are not, by any means, new to educational settings. For instance, the peer reviewed academic journal *Simulation and Gaming* dates back to the 1960s, and Russell and Shepherd (2010, 994) claim that educational computer simulations have been used for assessment since the 1970s – although with slow feedback, which could take weeks. Contemporary computer simulations can provide instant feedback, take place in virtual worlds, make use of the Internet and they may involve avatars, virtual identities as well as either artificial or direct, authentic social interaction.

The rise of computer simulations in language classrooms coincide with communicative language teaching and its desire to bring authentic and meaningful language to the classroom (Miller & Hegelheimer 2006, 313). Contextualised, educational SLA computer simulations "with realistic activities and tasks" for all four basic language skills were proposed for the enhancement of communicative competence as early as in the 1990s (Lee 1993, 221). A language computer simulation incorporates the use of language, and it aims at providing, as Crookall (2002, 273) denotes, "a relatively realistic language-using environment for learners to practice their new language and also for them to be able to make mistakes without a teacher clobbering them".

Computer simulations can be seen as a result of a long development in language testing which, in the latter half of the 20th century, also pursued the inclusion of social and functional aspects in the testing process in addition to assessment of language in more genuine situations: "One inevitable conclusion has been the realization that tests need to find some way to achieve authenticity, to measure the ability to perform in situations not unlike the real world." (Spolsky 2008, 450–451). The value of language simulations, then, "must be the extent to which it produces, in the performance of users, the appropriate language, i.e. language that would be appropriate in the corresponding "real-life' situations" (Jones 1986, 180–181). In relation to assessment, this means that instead of having to make inferences concerning the transferability

of the elicited language samples to the real-world situations, the language produced in the computer simulations is as close to real-world-like situations as possible in artificial contexts. Therefore, the inferences drawn on the basis of computer simulations may – given that the medium and method is familiar to the test takers – be much more reliable than those obtained through more conventional testing methods such as paper-and-pencil tests.

Assessment using computer simulations draws from the principle of experiential learning (see Kolb 1984). Feinstein, Mann and Corsun (2002, 741) describe experiential learning as something that “involves immersing learners in an environment in which they actively participate in acquiring knowledge”, and define computer simulation as “an experiential learning activity that allows learners to visualize situations”. Experiential assessment through simulations in CLIL can be defined as a process that involves immersing learners in a virtual environment in which they actively participate in demonstrating both content knowledge and language proficiency while performing various content and language integrated tasks. Computer simulations thus represent task-based assessment in a digital, virtual form and follow the principles of task-based performance assessment (TBPA) discussed in the previous sub-section (see also Skehan 2003).

Computer simulations also incorporate a range of representational modes such as “image, movement, gesture, music, sound-effect, and voice quality” (Jewitt 2003, 84) which all contribute to meaning making besides language, which has traditionally been perceived as the meaning-conveying mediator in learning. In the digital era, students need to master visual cues (visual literacy) in addition to the more traditional textual literacy. The NCC draft (2014) for the Finnish NCC 2016 reform refers to this in the general objective of multiliteracy. The multimodality, Jewitt (2003) rightfully argues, also requires a new approach to assessment while traditional assessment focuses primarily on linguistic oral and written modes. Consequently, multimodal realisations call for leaping into areas beyond language – it requires multimodal assessment, of which computer simulation is one example.

Assessing assessment, or quality control, is an important step in the validation of any assessment method. Because there seems to be no assessment criteria available for computer simulations, I have used the evaluation references of CALL task appropriateness by Chapelle (2001, 55) to create such assessment references. The original six criteria (language learning potential, learner fit, meaning focus, authenticity, positive impact and practicality) were originally written with a learning focus in mind. I rewrote them from learning mode to assessment mode and gained six criteria for evaluating the appropriateness and affordances of CLIL computer simulations experimented with within the project PROFICOM (2013) and reported in this study. I will return to these criteria to evaluate the computer simulations in this study in the section 8.3 of Discussion.

The six criteria are:

- **language assessment potential:** evidence for the foundation of inferences,
- **test taker fit:** the amount of opportunities for engagement with content-specific language use,
- **meaning (content) and form (language) focus:** the balance between these two and the extent of test taker's attention directed toward language form and meaning,
- **authenticity:** the degree of correspondence between the simulation tasks and possible real-life language use situations,
- **positive impact:** the positive impacts of simulations and the simulation environment to the test takers in the short and long run, and
- **practicality:** the reasonable amount of resources bound to simulation design, production and implementation with its various phases.

SUMMARY

Alternative assessment, the reformulation of assessment perspectives and foundations of assessment, emerged as a result of critical language testing opposing accountability policies in assessment and to better meet the challenges posed by global changes experienced in the 21st century. The reform is based on the societal, economic and technological changes of our society towards actions in digital networks, rapid media and news coverage as well as altering concepts and construction of knowledge. In order to educate capable and active citizen, we need to adapt to that change. There are many diverse approaches to alternative assessment which is intended to promote learning instead of solely stating the level and extent of attainment. The types of alternative assessments chosen for closer inspection in this study (collaborative assessment, task-based performance assessment and technology-based language assessment, especially computer simulations) are all adaptable to CLIL circumstances.

4.3 Assessment in primary CLIL

The current National Core Curriculum for Basic Education (NCC 2004, 273) maintains that assessment in CLIL must give *adequate information* on learners' target language proficiency to all stakeholders (teachers, pupils and parents). This section of the study aims, in the absence of assessment studies in primary CLIL, to answer the question of adequate, appropriate assessment in primary CLIL (or CLIL assessment, as I will refer to it interchangeably) at a theoretical level. I will argue that assessment in primary CLIL is ideally informed by three particular fields of language assessment: *assessment of young language learners* (YLLs),

assessment of *English for academic purposes* (EAP) and *content-based assessment* (CBA). I will address all these areas in this section starting with assessment of YLLs and assessment of EAP after which content-based assessment will be covered. In so doing, I will also simultaneously investigate CLIL assessment research and literature as well as introduce a few European developmental CLIL assessment projects.

Assessment of young language learners

The most decisive factor in primary CLIL assessment is the age of pupils which is why *assessment of young language learners* (YLLs) is highly relevant. Assessment of YLLs has been, particularly from the 1990s onwards, an independent, definite area within language assessment research, and it pertains to approximately 6 to 13-year-old pupils learning a foreign or second language (Hasselgreen 2005; McKay 2006, 1; Rea-Dickins 2000). YLLs exhibit special characteristics as language learners because they are still growing in more senses than just physically; they are developing literacy skills and new identities in addition to being vulnerable to criticism or failure (McKay 2006, 24, 30).

Young children in general are very optimistic about their academic skills potential (Kärkkäinen 2011), but Katz (1997) brings forth the risk of labelling YLLs as achievers, for “[y]oung children are notoriously poor test takers”, likely to make more mistakes the younger they are. A false label on a child may be detrimental to the budding language self-concept. For this reason, alternative, diversified assessment methods are highly recommendable: children can portray their strengths in multiple ways and view their weaknesses from a developmental angle. YLLs should be assessed in a different manner than more mature students; assessment should entail familiar contents (e.g. simple genres), and it should be executed by familiar adults in a “psychologically safe” environment; scaffolding is recommended and immediate feedback valued (McKay 2006, 9–10). Assessment of YLLs is ideally multimodal.

Some assessment tasks are more appropriate than others for YLLs. Brown and Hudson (1998, 658) differentiate three broad categories of language assessment: 1) selected-response assessments referring to true-false, multiple-choice and matching assessments, 2) constructed response assessments which are, for example, fill-in, short-answer and performance assessments and 3) personal-response assessments including conferences, self- and peer assessments and portfolios. The last is most appropriate to start with YLLs, and as pupils’ proficiency level increases, teachers could shift from the second category to the first (Shabaan 2001, 18). On the whole, Shabaan proposes using informal alternative and non-threatening assessment techniques with YLLs, such as nonverbal responses, oral interviews, role plays, written narratives, presentations, student-teacher conferences, self-assessment, dialogue journal, peer and group assessment and student portfolios – but not exclusively, because pupils should be familiarised with all kinds of methods in order to sketch the most multidimensional picture of the child’s skills and knowledge.

Hasselgreen (2005, 338) lists widely accepted characteristics of assessment tasks for YLLs (boldfacing mine):

- Tasks should be appealing to the age group, **interesting and captivating**, preferably with elements of **game and fun**.
- **Many types of assessment** should be used, with the **pupils', parents' and teachers' perspectives involved**.
- Both the tasks and the forms of feedback should be designed so that the pupil's **strengths** (what he or she can do) **are highlighted**.
- The pupil should, at least under some circumstances, be given **support in carrying out the tasks**.
- The **teacher** should be given access to and support in **understanding basic criteria and methods for assessing language proficiency**.
- The assessment activities should be **good learning activities** per se.

A young age does not prevent a pupil using technology-based assessment methods. Indeed, “computer assessment tasks that give immediate responses (with sounds and visual effects) and teachers responding kindly to the child’s efforts, are ideal for young learners” although the provision of immediate feedback can decrease with time and age of pupils (McKay 2006, 9–10). More demanding assessment tasks and methods could be introduced in upper-primary classes.

Assessment of English for academic purposes (EAP)

Another angle from which to consider assessment in primary CLIL is assessment of *English for academic purposes* (EAP). EAP is normally associated with higher and vocational education – especially with aptitude and placement tests in tertiary level (see Clapham 2000). Since CLIL deploys academic language and aims at bilingual academic attainment, it might be reasonable to consider whether EAP is of any relevance for CLIL. Even at the primary level it is useful to keep in mind that the pupils are being prepared for academic study that is gradually becoming linguistically more demanding – practitioners therefore need to be aware of issues relating to acquisition of academic English and the interplay between background knowledge and language knowledge displayed in language use discussed in the following paragraphs.

A specific-purpose language test differs from a general language test in that it is “one in which test content and methods are derived from an analysis of a specific purpose target language use situation so that the test tasks allow for an interaction between test taker’s language ability and specific purpose content knowledge, on the one hand, and the test tasks on the other” (Douglas 2000, 90). This is where an understanding of the differences of BICS and CALP as well as different subject literacies becomes important (see 3.2.1). In addition to acknowledging the different roles of languages in the school context and the subject-specific genres or language functions needed to express the content knowledge, teachers also have to consider how to combine – or separate – the assessment of both language and content.

Douglas (2000, 29) acknowledges the problems involving simultaneous assessment of content (background knowledge) and language knowledge because it is difficult to discern them from one another; it is thus utterly crucial that test developers have “an understanding of how specific purpose background knowledge interacts with language knowledge to produce a communicative performance in specific purpose contexts” (ibid., 33). Clapham’s (2000) research review on university level EAP reading comprehension testing reveals that language proficiency is at least as important as the background knowledge. Additionally, Douglas (2000, 34) lists basic, research-grounded guidelines for EAP test development:

1. Sufficient field specificity contributes to the test takers’ success in their own content area.
2. The more field specific a test is, the stronger the impact of background knowledge: good language proficiency does not compensate when the test content is highly specific and contains complex concepts.
3. It is likely that there is “a language proficiency threshold below which test takers are unable to make effective use of background knowledge, and a higher proficiency threshold above which a lack of relevant background knowledge could be compensated for by test takers making fuller use of their language resources”.
4. The amount of context-embedded information contributes more to the field specificity than content-specific vocabulary.

The inference concerning assessment in CLIL contexts I draw from these principles is that if high language proficiency can compensate for insufficient background knowledge and if high field specificity predominates language proficiency, then it is crucial for the success of learners in the CLIL programme and their academic language development that the conventions of academic language (e.g. interpretation and use of textual clues, cohesive elements) are specifically addressed in instruction along with content-related vocabulary. This implies that the focus in instruction should be balanced between form and meaning in order to use the principles of assessment of EAP filtered through assessment of YLLs in CLIL assessment.

Content-based assessment (CBA)

Content-based (language) assessment (CBA) is the third research area intrinsic to CLIL assessment. It denotes assessment in which content and language are assessed together (McKay 2006, 74). Assessment in CLIL is often portrayed as difficult and problematic (e.g. Hönig 2010, 3–4; Morgan 2006, 59; Mustaparta & Tella 1999, 36; Serragiotto 2007, 271). There are several reasons for this. For example, the decision of whether to assess content and language separately, both together or one through the other has to be made, and an appropriate assessment method chosen (Hofmannová, Novotná & Pípalová 2008; Morgan 2006, 60–61; Poisel 2007; Serragiotto 2007, 271). One should also consider what the target of assessment is: the end product or the process (Serragiotto 2007, 271). Product refers to concrete examples of the

learners' language proficiency whereas process refers to aspects such as working skills, metaknowledge and affective factors (see e.g. Table 14).

There are two major sources of complexities in CLIL that pose a challenge to assessment: the bilingual attainment (two languages in instruction) and the dual focus (content and language). Pure content or mere language tests for CLIL assessment are not appropriate because they have been designed for one-dimensional assessment (Short 1993). Additionally, language knowledge, as traditionally seen in L2 instruction, is strongly associated with lexical knowledge and production of sentences according to morphosyntactic rules out of specific context – the isolation persisted even after the rise of communicative language use which clearly moved towards content emphasis due to task-based instruction (Byrnes 2008, 37). Furthermore, teachers find the 'rivalry' between integrated content and language and the choice of language in testing (L1, L2 or a combination of them) confusing (Poisel 2007; Serragiotto 2007, 272).

The dual focus in CLIL practically leads to two approaches to assessment: either *integrated assessment* where both aspects are assessed simultaneously or *discrete assessment* where these aspects are considered individually (Barbero & Järvinen 2009; Serragiotto 2007). When the content knowledge is expressed through the foreign language, the data gathered for assessment contains both elements, and the investigation of the two sides of the 'CLIL coin' can occur separately. The emphasis placed on either aspect may vary depending on the objectives and context. For example, Bentley (2010, 84) claims that low exposure CLIL programmes are more language-focussed, whereas high exposure versions focus on both or content only. CLIL assessment may have other foci as well, such as cognitive skills, communication skills, learning to learn, practical skills (e.g. carrying out studies or experiments) and attitudes towards language learning (ibid., 84-85). These are examples of process assessment rather than product assessment.

The survey of Serragiotto (2007) in Italian middle and high school CLIL contexts investigated various questions related to assessment: what, who, when, with which methods and why. In Italy, team teaching or co-teaching is a typical way of organising CLIL instruction, as both the expertise of content teachers and language teachers are capitalised in such an arrangement. The study revealed, for instance, that 45% of both teacher types assessed both content and language equally, 45% favoured content assessment at the expense of language and the remaining 10% assessed language more than content. This finding discloses inconsistency in the assessment practices which, as Serragiotto also states, are far from established and systematic. The assessment focus is inevitably affected by the Italian test design which could occur in pairs or either teacher takes the responsibility of the assessment (Serragiotto 2007, 271).

The separate treatment of content and language in assessment, i.e. discrete assessment seems to be the method suggested by the majority of scholars (e.g. García 2009a, 370; Mohan 1986, 122) and developmental assessment projects described below. Alternative assessment methods appear to be favoured in CLIL contexts, for example Short (1993, 633) recommends adopting methods such as "performance-based tests, portfolios, journals, projects, and

observation checklists” where the assessment of content and language knowledge are separated from one another and scrutinised only one at the time mirroring the precise than rather broad objectives which, according to her strong recommendations, are defined prior to instruction and assessment.

Thus, the curricular base is a fundamental prerequisite for assessment. The prerequisite for CBA is the existence of a CLIL curriculum and its baseline objectives for both language and content. Weigle and Jensen (1997, 211) stress the importance of anchoring the proportional assessment of content and language in the requirements of the curriculum. With this rule of thumb, if the proportion of the TL in CLIL is 25%, then 25% of the assessment in CLIL should occur in the TL.

The alignment of separate objectives becomes very clear in Figure 14, adapted from Gottlieb (2006), which demonstrates the bridge from language proficiency to academic achievement: the concurrency and relations of curricular objectives, instructional objectives and assessment in strands of content and language which are bridged together and affecting each other. Figure 14 depicts language proficiency and academic achievement as parallel, equally significant components in learning, both first grounded on general standards (objectives or desirable level of proficiency stated in the curriculum) of which the more specific instructional objectives are drawn and assessed. Content objectives define the academic language needed in achieving the content standards and naturally shape the instruction aiming at English language proficiency and academic achievement.

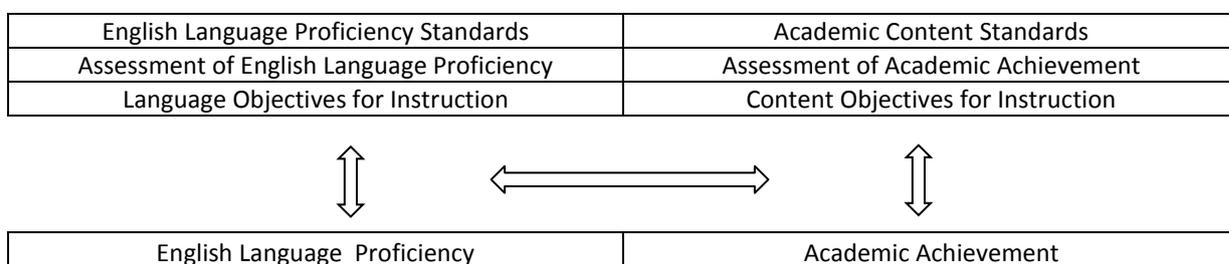


FIGURE 14. The alignment of standards, assessment and instruction of English language learners (Gottlieb 2006, 64)

The separation of content and language is not, however, as simple a procedure as it may appear to be at first glance, because the grading or assessment seems to be vulnerable to external influences. If tested in the target language, the incomplete learner language may affect the scoring of content mastery negatively and vice versa when the testee is able to express him/herself eloquently, as the Austrian CLIL study by Hönic (2010) has discovered. Llosa (2011) also reports findings of such inconsistent scorings in CBI contexts in California, where classroom teachers assessed the language proficiency of English language learners against local language development standards. She noticed that when scoring, teachers paid attention to other than language features such as “students’ personality and classroom behaviour, the teachers’ beliefs about assessment and grading, and external pressure to advance students to the next level” appeared to have an effect on assessment (Llosa 2011, 370). For those reasons, separate,

premeditated criteria for both content and language instead of impressions as reference for assessment are more appropriate.

Criterion-referenced inferencing

The curricular objectives serve as references for more detailed assessment criteria. Implementing *criterion-referenced assessment* (or rather criterion-referenced inferencing, as Griffin 2009, 187 specifies the terminology) in the form of various rubrics or matrixes appears to be the most common solution considered for CLIL assessment. Two European developmental assessment projects in CLIL, which have both taken the quest for finding appropriate assessment activities, methods and tools, suggest rubrics for CLIL assessment. The outcome of the project Assessment and Evaluation in CLIL, AECLIL (2013) is a package of downloadable assessment and evaluation tools ranging from teacher observation sheets to assessment rubrics and an online publication (Quartapelle 2012) which covers practical CLIL issues from primary level to in-service teacher training and gives a range of examples of various assessment grids for diverse purposes.

The primary aim of the other research project, CLIL Learner Assessment, CLILA (2013), is to generate an assessment tool for primary level to measure pupils' content and language command based on four pillars: 1) the Common European Framework of Reference (CEFR), 2) *Lingualevel*¹¹ which is a Swiss collection of assessment tools for English and French as a foreign language from the 5th to 9th grade, 3) the curriculum of the subject in question and 4) the descriptions of disciplinary skills of the given subject (CLILA 2013). It appears that CLILA is also inclined to present a grid matrix which is a two-dimensional synthesis of all documents chosen relevant for the assessment process consisting of axes "*Kompetenzaspekte*" (Aspects of Competence) and "*Themenbereiche*" (Theme Domains). A handbook of achievement enhancement and assessment in primary CLIL by Massler and Stotz (in preparation) will later report the results obtained from the CLILA project.

Other principles in CLIL assessment

Furthermore, other principles and attributes than employing rubrics or matrixes as references have been discussed in the literature. For instance, Johnstone (2000) uses the term "*embeddedness in a flow of events*" to describe the holistic nature of simultaneous language and content learning in the primary classroom, and expresses concerns related to assessment tasks that do not display similar embeddedness:

In the case of primary school children an assessment task is unlikely to be valid unless it represents a type of activity with which they have some familiarity; however, in addition, if they are asked to make 'a cold start' in an assessment task, when they are accustomed every day to being 'warmed up' for it cognitively as well as linguistically, then questions must arise about the validity of the process.

¹¹ See www.lingualevel.ch/ for more information.

Put in other words, there is little if no sense in attempting to test language proficiency in, for instance, paper-and-pencil tests if such activities are not typical for everyday classroom activities. Conversely, if the CLIL language is used in everyday activities such as group work or presentations, then the appropriate method is to accommodate assessment into that as well or administrate assessment while such activities occur.

Most CLIL assessment experiments appear more often than not to be various forms of formative, alternative assessment methods. Poisel (2007) submits and describes such a formative CLIL assessment mode consisting of student-centred study, portfolio work, peer tutoring and teacher facilitating. Task-based assessment (see previous section), leaning on criterion-based assessment, is yet another option for CLIL assessment. Barbero (2007, 296–297) gives an example of a task-based assessment grid consisting of a content range (the task achievement) and language ranges further divided into sub-ranges of vocabulary and structures, accuracy, fluency and interaction and coherence. The task is assessed on a 0–5 scale through all these descriptors.

Technology-based CBA is examined in the Finnish project Profiling Learning Progression in CLIL Environments through Computer Simulations, PROFICOM (2013) looking into the appropriateness of computer simulations in primary CLIL assessment (see 5.4.2). This project employs CEFR scales as a reference, as is typical in the European CLIL contexts. (see e.g. Díaz Cobo 2009). In addition to the CEFR, also other assessment systems have been experimented with in CLIL environments: the Middle Years Programme (MYP) by the International Baccalaureate Organisation, the European Language Portfolio (ELP) based on the CEFR in various country-specific versions and some national language curricula (Morgan 2006). Although admitting the many language-related arrangements of MYP are congruent with CLIL, Morgan (2006, 63–64) is more inclined to prefer the ELP as a more appropriate assessment tool for CLIL learners especially in the Austrian context partly due to the flexibility of the educational system, partly because “the breaking down of language skills into subskills (‘I can...’) allowed for the range and depth of CLIL pupils’ skills to be accommodated”.

Accommodation in assessment contexts means differentiation and scaffolded assessment that takes individual learner needs into consideration; in other words “modifying the assessment and giving [pupils] support strategies to produce answers either orally or in writing” (Bentley 2010, 95). For example, García (2009a, 371) proposes allowing translanguaging (intentional and accepted change of languages) in tests. In praxis, questions may be in the TL whereas pupils can decide which language they use for answering.

In addition, theory-based assessment initiatives have also been taken in CBA. Mohan, Leung and Slater’s (2010) *Integrated Assessment of Language and Content* (IALC) leans on a Systemic Functional Linguistic framework based on Halliday’s concepts of language as a meaning potential (see p. 53): “If language is the primary evidence for learning, then assessment is primarily assessment of text or discourse, and how wording constructs meaning in text” (ibid., 221). Assessment is thus modelled as “a language process” with functional assessments of genre

and register (see 3.2.1) as the starting point of IALC in written texts as well as in classroom discourse. IALC may be adaptable for upper primary classes, but conducting linguistic analyses requires highly linguistically competent CLIL class teachers. The IALC may be better applicable to secondary or tertiary CLIL education where the students' language potential is higher and they possess more sophisticated discourse skills.

In order to keep track of the variety of methods used and to ensure that the role of language and the relationship between content and language are carefully premeditated and defined prior to actual investigation of assessment data, it might be appropriate to organise an assessment scheme or inventory. Short (1993) proposes a framework for integrated content and language assessment consisting of two axes: *what* (the assessed features, e.g. problem-solving skills, content area skills, concept comprehension, language, communication skills, individual and group work as well as attitude) and *how* (the methods used, e.g. checklists, teacher observation, student self-evaluation, oral and written reports and portfolios). These two axes form a grid matrix serving as a compilation of individual documents contributing to the assessment inventory, and it is useful in showing the distribution of alternative assessment activities warranting that the process of data gathering is as favourable as possible to different kinds of learners in terms of learning styles, cognitive styles and individual preferences. Such a practice is transparent to all parties involved and, according to Short (1993, 652), "balance control and responsibility for assessment outcomes between teachers and students".

SUMMARY

In this study, it is suggested that language assessment in primary CLIL classrooms is theoretically based on three pillars: assessment of young language learners, assessment of English for academic purposes and content-based assessment. It seems that the separate, discrete assessment of content and language and use of criterion-referenced inferencing in rating are favoured in CLIL literature rather than the integrated assessment of content and language. CLIL assessment research is extremely rare. It appears that, so far, there are the studies of Hönig (2010) and Serragiotto (2007) explicitly concerning assessment issues, both at upper levels of education. Further empirical research is thus valuable, and formulating a more solid theory basis for CLIL assessment than presented in this study is an important future task. Assessment in CLIL is grounded in objectives from which assessment criteria are drawn. Various assessment grids or matrixes appear to be recommended for CLIL contexts. Multimodal and alternative assessments are applicable for young language learners in CLIL, and in Europe, the CEFR is a common basic tool to rely on.

5 RESEARCH METHODS

The previous chapters examined the research topic from the theoretical viewpoint. In this empirical part of the research report, I will elucidate the research methods, present the obtained results and place them into perspective with prior research as well as discuss their validity and significance. The implications of the findings for CLIL will be touched upon in the final section concluding the study. This chapter concentrates on explaining how the empirical research was conducted. I will begin with a presentation of the empirical research framework (5.1). Then I continue to the research questions (5.2), after which the description of the participants (5.3) and the three research phases (5.4) will follow. Section 5.4 is divided into two sub-sections expounding on the CLIL assessment survey (phase 1) and the two computer simulation experimentations (phases 2 and 3). In the final section 5.5, I will illuminate the data analysis methods.

5.1 Research framework

The core areas of study are current practices and computer simulations in CLIL assessment. The empirical research has four aims. First, it aims to distinguish the practices and methods used by CLIL class teachers for assessment of pupils' English language proficiency and its development as well as the extent of those practices. The second purpose of the research is to pinpoint challenges and development areas in CLIL assessment. Furthermore, this study is interested in investigating the adequacy of CLIL assessment and the feedback on CLIL pupils' English language proficiency as perceived by pupils themselves and their parents. The fourth aim is to examine how computer simulations are perceived as an alternative assessment method in primary CLIL.

The empirical research framework of this study in Figure 15 visually outlines the choices I have made to carry out this study. This study is a synthetic-heuristic, mixed methods research with a qualitative emphasis and developmental intention. The *synthetic-heuristic approach* denotes that the purpose of the study is to discover and describe patterns or aspects in a phenomenon (in this case, assessment of YLL's English proficiency in CLIL) based on the gathered data with no or few predefined specific preconceptions of the phenomenon under scrutiny (Seliger & Shohamy 1989, 31–32, 55–58). It is thus very holistic in nature which is why the degree of control over the studied phenomenon is fairly low (*ibid.*).

Mixed methods indicate that the two traditional research practice categories of *quantitative* (numerical, psychometric data of 'how many or how much, to what extent' etc.) and *qualitative* (descriptive data of 'how, what kind of, why' etc.) research are mixed or merged. Since this study aims at exploring, describing, interpreting and understanding the state of art of assessment in bilingual CLIL education as well as seeking various possibilities for such

assessment methods both theoretically and empirically from the perspectives of all parties involved in assessment (teachers, students and parents), the mixed research approach is predominantly qualitative.

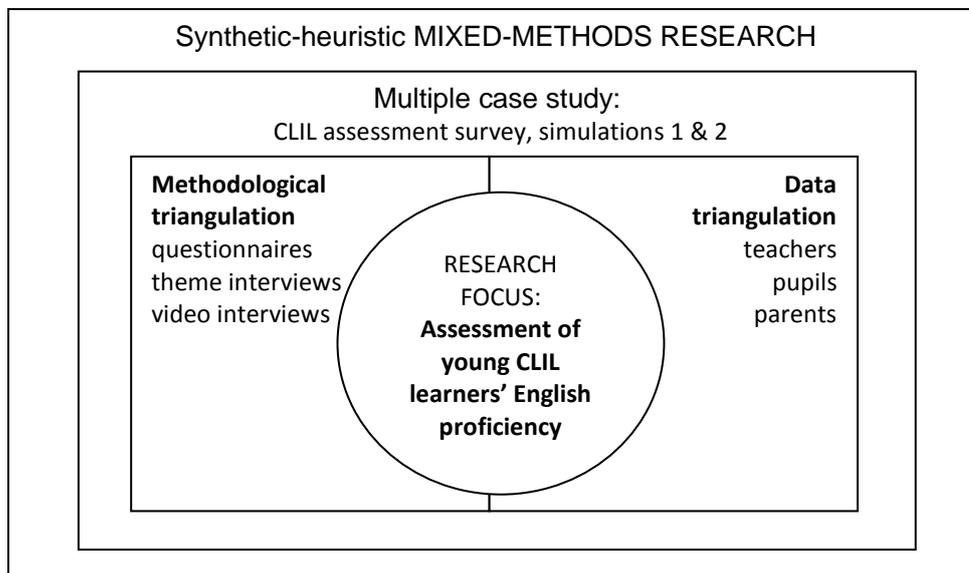


FIGURE 15. Empirical research framework

Duff (2007, 974–976) defines qualitative research as “a cluster or continuum of approaches that generally seek contextualized, naturalistic, holistic understandings and interpretations of phenomena that occur in particular types of contexts” whereas quantitative research is more interested in causal relationships between diverse variants and their strength. However, since quantitative analysis methods are also used, this study represents a mixed methods (MM) study rather than a purely qualitative approach. Teddlie and Tashakkori (2003, 14–17) argue that there appears to be three areas that advocate for the use of mixed rather than single approach designs:

- MM research “can answer research questions that the other methodologies cannot.” While quantitative research is typically confirmatory and related to theory verification, qualitative research is exploratory and connected with theory generation. Mixed methods research thus combines the advantages of both approaches and is hence more versatile.
- *MM research “provides better (stronger) inferences.”* This postulation is based on the idea that the use of mixed methods compensates for weaknesses and reinforces the strengths the single approaches would intrinsically demonstrate. The choice of mixed methods thus contributes to the quality of conclusions made on basis of the data.
- *“Mixed methods provide the opportunity for presenting a greater diversity of divergent views.”* This may refer even to completely opposite results drawn from the qualitative and quantitative data, but since they “reflect different voices and perspectives”, such deviation and “diversity of opinion is welcome”.

Onwuegbuzie and Johnson (2006, 49) point out that MM research is able to address practical questions, examine a variety of perspectives and “if well documented, practitioners can obtain some sense of what might be useful in their local situations”. This statement is particularly valid in respect this study, one purpose of which is “to investigate and share information about current assessment practices (McKay 2006, 65; see p. 13).

In order to present a diversity of views, to approach the research focus from different angles and enrich the findings, a *triangulation* approach was adopted in this study. Triangulation refers to finding patterns from different sources (Seliger & Shohamy 1989, 123) and indicates a multimethodological approach to data collection and analysis, sources of data or team research, and the use of triangulation aims at creating a deeper, multifaceted understanding of the phenomenon under scrutiny (Rothbauer 2008, 893). Theories, sites of study and interpretations are also options for triangulation (Duff 2007, 976). The study at hand incorporates triangulation in data collection and sources of data. The validity threats of triangulation are discussed in 8.4.

In using triangulation, it is possible that some sources of data and data collection methods are more sensitive to the studied phenomenon than others. Different data sets and methods may also introduce such aspects of the phenomenon that were not anticipated. For example, in this study, the role of language in CLIL as perceived by the teachers was one of such phenomena. Particularly theme interviews as data elicitation methods are likely to expose aspects and features of CLIL assessment in more depth than questionnaires because the situation is interactive and not tied to a rigid set of questions. It is also possible that adults, teachers and parents, are more adept at disclosing and analysing their perceptions and thoughts about the studied phenomenon than children, but regardless of that, every stakeholder group adds its own viewpoints.

An MM research paradigm as well as triangulation are recommended and desirable as attributes of CLIL research because this operation mode produces more diversified account of the studied phenomenon (Pérez-Cañado 2012, 332) and they increase the conclusiveness of findings (Bonnet 2012, 66). This is also achieved by taking the product (outcome of CLIL education), process (the methods used in CLIL) and participant perspectives into account (ibid.). The participant perspective is inherently included in micro-approach research which, from the inside of the phenomenon, focuses on the participants (teachers, students, parents) “trying to find out how they act under the conditions of CLIL, what happens to them, and how CLIL influences their states of mind, that is their cognition, knowledge, emotions, beliefs, opinions and attitudes” (Dalton-Puffer & Smit 2007, 12). In this research, the product is second language proficiency, whereas the process is the act of assessment. Because this is a descriptive study, I have given the participants a strong voice in expressing their thoughts, preferences, opinions and visions.

A *case study* as a research method is used to provide extensive, in-depth description of a contemporary, educational phenomenon in its real-life context “especially when the boundaries between phenomenon and context are not clearly evident”, but the context is pertinent to the study (Yin 2009, 18). This research aims to describe and to contribute to the further development

of current practices in language assessment in CLIL, so the research method is a developmental and applied multiple-case study which entails three cases which I call phases: the assessment survey, and simulations 1 & 2. There are also traits of pedagogical action study in this research due to the participatory role of mine especially in the simulation experiments, attempt to initiate change in the studied practices and to generate actionable knowledge (e.g. Somekh 2008). The case study method allows both quantitative and qualitative approaches.

5.2 Research questions

Research questions advance deductively from the general towards the specific. The first entity of research questions pertains to CLIL class teachers, the second to CLIL pupils and their parents, while the last deals with computer simulations from the viewpoint of pupils and their parents.

The first main research question aims at creating an overall conception of language assessment in primary CLIL. The first sub-question is interested in current assessment methods but does not differentiate between the two predominant assessment purposes, formative and summative assessment, due to the fact that during basic education, grades 1–7, the current official national assessment scheme acknowledges no other form of assessment than formative, continuous assessment (see 4.1). Furthermore, this study is, as devoted to its synthetic-holistic nature, interested in any possible indication of language assessment in CLIL. The sub-question 1.2 queries the extent of language assessment and the following two sub-questions the challenges in assessment as well as future visions for developing future practices.

1 How is assessment of English language proficiency in CLIL organised according to class teachers?

- 1.1 What kind of assessment practices do CLIL class teachers employ for assessment of English language proficiency in CLIL?
- 1.2 To what extent do CLIL class teachers assess language and provide feedback to pupils and their parents?
- 1.3 What kinds of challenges are related to CLIL assessment?
- 1.4 How should CLIL assessment be developed?

The second main set of questions is related to the adequacy of assessment in CLIL as experienced by the pupils and parents. Adequacy is the quality of assessments in CLIL determined by the NCC (2004), but there is no definition for adequacy of assessment. Therefore, the stakeholders in assessment are the expert informants in this matter. The sub-questions elucidate the frequency and sufficiency of the received assessment information in CLIL contexts as well as the preferred methods and practices to receive that information. No division between formative and summative practices was made in respect of this set of questions either since any instance of perceived language assessment in CLIL or ways of receiving feedback were considered important to the comprehensive nature of the study.

2 How adequate is assessment of English language proficiency in CLIL according to pupils and parents?

- 2.1 Do pupils and their parents receive information on pupils' English language proficiency and its development in CLIL contexts frequently and sufficiently enough?
- 2.2 Which practices do CLIL pupils and their parents prefer in receiving information on pupils' English language proficiency and its development?

The third question is in connection with the appropriateness of computer simulations for innovative, alternative language assessment in primary CLIL. This question is important to ask to face the challenge of creating more modern, technology-based and future-oriented assessments (see 4.1.2. and 4.2.). Computer simulations incorporate many features appraised as valuable in future education (see e.g. NCC draft 2014) and they familiarise learners with computer-assisted language testing which is now already a megatrend in the world, let alone in Finland (e.g. electric baccalaureate). In addition to describing current practices and stakeholders' opinions on them, I wished to contribute to the CLIL assessment discussion by investigating two alternatives in the field of assessment: LangPerform computer simulations and language portfolios. Since the portfolio experiment is still ongoing, I will report the computer simulation experiment in this publication, and the portfolio report will be published in Wewer (forthcoming).

This set of questions primarily addresses the participants' own perceptions of the issues and advantages of using computer simulations. It significantly deepens the data obtained from questions 1 and 2 as well as provides a new, technology-based and alternative perspective on assessment in CLIL. The actual linguistic analysis of pupils' simulation performances was not manageable within this context and is a matter of further research. The sub-questions illuminate specific aspects of pupils' English language proficiency revealed by the simulations as well as the user experiences of pupils and their parents.

3 What are the key issues and advantages in using computer simulations as an alternative assessment method in CLIL as perceived by pupils and their parents?

- 3.1 What kind of information do computer simulations yield on pupils' English language proficiency?
- 3.2 What kinds of thoughts and experiences are produced by the simulation experiments?

The question formatting is deliberately open because no comparable studies at primary level exist, and according to the heuristic objective of this study, the phenomenon is approached with discoveries in mind.

5.3 Participants

The research group consisted of primary CLIL class teachers as well as CLIL pupils and their parents or guardians – the stakeholders that, according to the current Finnish National Core Curriculum for basic education (NCC 2004), assessment in CLIL should provide information. The three groups were included in order to achieve data source triangulation and a deeper understanding of the topic.

CLIL teachers

The CLIL class teachers were only involved in the first research phase, the CLIL assessment survey. The 42 participating teachers were recruited through two channels: participation in this research (most CLIL class teachers in the two research schools participated) and CLIL Network web site¹² which lists many, but not all, as it turned out, CLIL-providing schools. According to the background information given by the teachers themselves, their educational and vocational backgrounds varied substantially, and some of them fell into more than one of the predefined categories.

All of these participants were qualified as teachers (master's degree), but one fifth (9/42) had no language studies at all. Most (18/42) had a class teacher qualification and basic studies in English (25 credits), and 10 of them fulfilled the CLIL teacher language proficiency prerequisites ordained by the Ministry of Education (2005). Six teachers possessed a double qualification as class teacher and English language teacher. Two teachers had obtained a CLIL teacher qualification by accomplishing teacher training in an Anglophone country and two were English subject teachers. Altogether, 15 teachers indicated having other training or applicability for becoming a CLIL teacher, e.g. in-service training or the Juliet programme¹³ offered by the University of Jyväskylä.

Many participating teachers had an extensive experience in teaching CLIL: 74% of them more than six years and 26% 15 years or more experience. It is therefore reasonable to assume that the participants can provide a relevant insight into the procedures of CLIL assessment. The novices and practitioners with a few years' experience can contribute significantly to the amelioration of CLIL assessment especially by pointing out weaknesses and deficits they have detected as well as stating what kind of assessment tools would be applicable and needed.

Different age groups require different assessment methods, which necessitates the investigation of grade distribution among participants. By far the most representative group of CLIL teachers was that of beginning instruction, i.e. grades 1–2 (pupils aged 7–8). Almost half (46%) of the informants belonged to that group, whereas division between grades 3–4, pupils aged 9–10 (28%) and 5–6, pupils aged 11–12 (26%) was more even.

¹² <http://clil-network.uta.fi/index.php?id=2&L=1>

¹³ <https://www.jyu.fi/edu/laitokset/okl/opiskelu/sivuaineet/juliet/en/intro>

CLIL pupils and their parents

The CLIL pupils and their parents were involved in each research phase, but the participants were only partially the same. They were chosen from two schools located in South-Western Finland; a town school (TS) and city school (CS). The school profiles were quite contrary as shown in Table 17 which describes the state of art during the empirical research phases in years 2012–2013. The most marked distinctions are the location, school type, pupils' ethnography, years of CLIL provision, curricular specifications, the extent of EFL instruction for CLIL learners as well as experience and linguistic training of CLIL teachers. The number of participating CLIL pupils and parents varied according to the research phase (Table 18). Pupils ranged from 3rd to 6th graders. They all had been admitted to CLIL instruction through an entrance test relating to general language skills and other factors that may influence studying.

TABLE 17. Contrastive table of participant school features

Feature	School 1 (TS)	School 2 (CS)
School type	municipal primary school	teacher training school, comprehensive school
Location	town residential area in South-Western Finland	city suburb in South-Western Finland
Number of primary pupils	altogether ca 400, whereof six classes of CLIL pupils (ca 135)	altogether ca 345, whereof six classes of CLIL pupils (ca 115)
Ethnography of pupils	predominantly native Finnish-speaking	ca 50% native Finnish-speaking, ca 50% with multicultural immigrant background
CLIL admission test	yes	yes
CLIL provision	grades 1–6	grades 1–9 (+ International Baccalaureate Programme)
Years of CLIL provision	ca 10	ca 20
CLIL curriculum	lists vocabulary and topics	specifies themes, language functions and structures
Extent of CLIL instruction	not specified	ca 25%
Desirable language level objective	not specified	not specified
Extent of formal EFL instruction	two lessons (2 x 45 min) per week	three lessons (3 x 45 min) per week
Qualification of CLIL teachers	some experienced, none with formal language qualifications, few with language training, most with in-service training	mostly experienced, some with formal language qualifications, most with language training, all with in-service training
CLIL assessment and reports	none (no common practice)*	none (no common practice)*
General assessment scheme	development discussion, middle term report and school year report	development discussion based on self-assessment form, school year report

* Both schools provide, however, a school year report stating in one additional sentence that the pupil has attended English class instruction. The success of CLIL study or language proficiency is not assessed nor reported specifically.

Regarding the reasons for applying for CLIL, 88% of parents (n=98) chose achievement of better English proficiency. The second most frequent (46%) motivator was the child's own wish and the

most third frequent (45%) contingent success in work life. Language acquisition was thus seen as a highly motivating factor in opting for CLIL.

Research ethics and integrity

A very firm ethical research code was followed especially due to the participation of minors. Gray (2009, 73) lists four categories of ethical principles which I took into account as described underneath.

Any harm to participants should be avoided. All research phases were based on voluntariness; the participants could therefore decide themselves whether or not take part in various research activities. The research activities were such in nature that no physical, mental or emotional harm was intentionally caused. Apart from experiences of individual test anxiety and annoyance in situations of technical problems, the participants reported no discomfort.

Informed consent of participants should be obtained. The research plan was discussed with the head teachers of the participant schools and their approval was obtained prior to actual implementations. Every empirical research phase involving children was preceded by sending an information letter with the written consent formula to pupils' homes for their parents to read and sign. Since this research contained several informed consent letters for multiple groups of parents, an example of the survey and simulation letters are attached in this publication (Appendices 1 & 2). The letters are in Finnish and identification references have been removed.

The privacy of participants should be respected. Every questionnaire in each research phase was anonymous. Names and contact details were asked and submitted only in cases where participant was willing to be interviewed. Teachers' and parents' quotes detached from interviews or questionnaires are not labelled in any way; those of the pupils' are presented with generic identification labels of gender and grade. The identity of participants is thus carefully preserved. Personal or other corresponding data is and will be confidentially and appropriately stored.

Use of deception should be avoided. The research model required no deception of any kind; the research process was totally transparent and based on sincere scientific intentions and practices.

5.4 Research phases

The data were collected multi-methodologically in three autonomous research cases which I will call phases: 1) the CLIL assessment survey, 2) the experimentation with computer simulation 1 and 3) the experimentation with computer simulation 2. The data were triangulated. The research settings are slightly different in each phase as described in Table 18, featuring the main characteristics of each phase. A specific description of them is given in subsequent sub-sections.

5.4.1 CLIL assessment survey

The purpose of the assessment survey was to answer the research questions 1 and 2 (see 5.2). The overall objective of the CLIL assessment survey was to form an overview of assessment in CLIL in Finland because no prior studies in this field exist. The survey was conducted in spring 2012 in two different stages (questionnaires and interviews) to explore which assessment practices teachers use to assess and monitor CLIL pupils' English language proficiency in CLIL subjects, and what the identified problem areas and visions are in respect of future development. Furthermore, pupils and their parents were questioned over whether they were pleased with the prevailing practices, and their wishes regarding assessment were revealed. The purpose of this was to find out whether or not the assessment practices were adequate as necessitated by the NCC (2004). The questionnaire pertained to more general assessment issues, while the role of the subsequent, theme interviews was to elicit more detailed information on those assessment practices.

TABLE 18. Main features of the research phases

Research phase	CLIL Assessment Survey	Simulation 1	Simulation 2
Time	spring 2012	autumn 2012	winter 2013
Aim	to investigate current assessment practices in CLIL classrooms	to experiment with a computer simulation and study its appropriateness in CLIL assessment	to experiment with a computer simulation study its appropriateness in CLIL assessment
Participants	CLIL teachers nationwide (n=42), CLIL pupils (n=109) and their parents (n=99) in three classes (grades 3–5) of two research schools	CLIL learners (5 th and 6 th graders, n=74) in the research schools and their parents (n=26)	CLIL learners (4 th and 5 th graders, n=72) in the research schools and their parents (n=13)
Methods	1) Webropol internet questionnaires for CLIL teachers teaching in CLIL providing schools listed in CLIL Network web site and 2) paper questionnaires for teachers, pupils and parents in the two research schools (TS & CS) 3) theme interviews of volunteers in each participant group	1) paper questionnaires for participating pupils and their parents 2) video interviews of volunteered pupils	paper questionnaires for participating pupils and their parents
Obtained data	250 questionnaires and almost 19 hours of transcribed audio-recordings	146 pupil and 39 parent questionnaires 50 video interviews, documentation of experiments in CS (video footage and photos)	
Data analysis	quantitative: percentages and numerical frequencies qualitative: thematic content analysis	quantitative: percentages and numerical frequencies qualitative: thematic content analysis	

TS - town school CS - city school

Participants in the questionnaire stage were CLIL teachers (n=42), the pupils of three CLIL classes in the town school (n=58) as well as three classes in the city school (n=51) and the parents or guardians of the given pupils (n=99). Volunteered teachers (n=10), pupils (n=20) and parents (n=7) were invited for an elaborative theme interview. The interviews will be addressed in more detail after the questionnaires on page 118.

5.4.1.1 Questionnaires

The Finnish-language CLIL assessment survey was adapted for each group. All three questionnaires were semi-structured in that they consisted of both close-ended (quantitative) and open-ended (qualitative) questions. To ensure a high response rate, I designed them to be filled in quickly and easily by ticking a choice, yet not disregarding the possibility to write further, additional comments to those questions. Each questionnaire was peer reviewed. Additionally, to be certain of the age-appropriateness of the pupils' questionnaire, it was linguistically edited especially for primary-aged children and pre-tested with a class of CLIL 2nd graders prior to the actual deployment. *Each questionnaire emphasised that the questions pertained to English language in other subjects than English in regular EFL instruction.*

Teachers' questionnaire

The CLIL teachers were approached through a questionnaire (Appendix 3) in order to gain first-hand information of the current, at that time, language proficiency assessment practices. The questionnaire enquired, for instance, about background information (e.g. education and number of CLIL years), stance towards assessment in CLIL, used assessment methods, frequency and ways of informing pupils and their parents of the assessments as well as possible ideas for how to improve and develop assessment in CLIL. The questionnaire did not make any explicit difference between formative and summative assessment, because the normative National Core Curriculum for Basic Education (NCC 2004) does not specifically differentiate between the two forms of assessment during grades 1–7 but recognises assessment during studies. Two sets of identical questionnaires were released, the sole difference being that the first set was in paper for the teachers in the research schools, and the second was an electronic web questionnaire necessary to reach CLIL class teachers nationwide.

I sent, via e-mail, an introduction letter containing a link to the Webropol questionnaire to the head teachers of 25 schools providing English-language CLIL tuition at primary level listed in the national CLIL Network web page (see p. 112) in March 2012. I asked the head teachers to forward the e-mail to the CLIL teachers of the school. Within the following month, 36 teachers filled in and returned the internet questionnaire. Assuming that each of the contacted schools had six CLIL classes, the potential maximum number of replies was 150 in which case the response rate was around 24%. Because the questionnaires were identical, the answers were combined into one corpus by entering the paper questionnaires manually into the Webropol data bank.

Combined with the paper questionnaires, 42 teachers altogether formed the CLIL teacher sample.

Pupils' questionnaire

I was always present while carrying out the pupils' questionnaire (Appendix 4) in six classes (3rd, 4th and 5th grade classes in two schools in an identical manner. I always started with a mutual introduction in which I explained the purpose of the survey and stressed common principles of replying, such as answering according to one's genuine opinion and making sure that none of the questions is missed. Additionally, pupils were encouraged to ask for clarifications in case they did not grasp something. I also emphasised that the questionnaire does not pertain merely to their own class teacher, but to all CLIL teachers they have encountered as well as all CLIL classes or lessons they have attended.

The questionnaire touched upon issues such as the perceived importance of receiving feedback of coping in English in CLIL subjects, the actual feedback received and desired assessment methods or ways of receiving feedback. Pupils who had finished the questionnaire were allowed to draw while others were still continuing. Altogether, 109 pupils, 63 girls and 46 boys, from both schools, all with parental consent, filled in the questionnaire. Table 19 displays the exact numbers per grade and school.

TABLE 19. Quantity of pupils answering the CLIL assessment questionnaire

PUPILS	3 rd Grade	4 th Grade	5 th Grade	Total / School
TS	18	23	17	58
CS	19	15	17	51
Pupils/ Grade	37	38	34	109

TS - town school CS – city school

Parents' questionnaire

The dispatching of parents' questionnaires (Appendix 5) occurred simultaneously with the pupils' questionnaire filling. Every single pupil, regardless of their own participation in the survey, received an envelope with a parents' questionnaire inside. In case of an absent pupil, the class teacher was requested to forward the envelope when the given pupil was again present. The pupils were asked to act as couriers and bring the sealed envelope to their parents and return the envelope back to school with the filled questionnaire inside. In the classroom, there was a box where the questionnaire envelopes were to be left. The parents' questionnaire concerned, for example, the sufficiency of information received from school regarding the child's management in English in CLIL subjects, development of English language proficiency, means of receiving assessment information on CLIL language issues and preferred ways of getting information. After three weeks' time, the questionnaires were collected for analysis. The number of parents

returning the questionnaire was reasonably high and very satisfactory, as can be seen in Table 20, the response rate of the CS being somewhat higher than the TS.

TABLE 20. Statistics of parents' questionnaires in the CLIL assessment survey

DISPATCHED PARENTS' QUESTIONNAIRES					Returned questionnaires	Percentage of returned questionnaires
	3 rd grade	4 th grade	5 th grade	Total		
TS	21	23	21	65	51	78%
CS	19	18	17	54	48	89%
Total	40	41	38	119	99	83%

TS - town school CS - city school

It is not uncommon that siblings in a family attend CLIL education, and therefore two or even more questionnaires may have ended up with the same parents. It is therefore possible that parents who received more than one questionnaire filled in and returned only one. It is also possible that not all pupils remembered to forward the research envelopes or return them back to school. I asked the class teachers to remind parents of the questionnaires through Wilma¹⁴ messages to increase the response rate.

5.4.1.2 Interviews

Questionnaires were not the only investigation method in this research phase. In order to obtain a deeper and more profound understanding of the topic, I considered elaborative theme interviews appropriate. A themed (aka semi-structured) interview focuses on certain topics and themes which are similar to all interviewees, although variation in order, depth and even formulation of questions may occur (Hirsjärvi & Hurme 2010, 45–46). Therefore, a call for voluntary interviewees was included in each questionnaire version. A surprisingly high number of volunteers enrolled for the interviews, which were conducted from late spring to early summer 2012. The purpose of the theme interviews was to gather participants' knowledge, experiences, opinions, thoughts and ideas of CLIL assessment in general and assessment of English language proficiency in particular.

Each participant group was interviewed according to its own body of themes (Appendix 6), and all interviews were audio-recorded using a digital voice recorder. The teachers and parents were interviewed at a location of their choice, most often at their homes or in the school building after working hours. The pupils were always interviewed at school during school days. I interviewed all of the volunteered participants personally. It was not, however, possible or rational to interview every volunteer due to the large number of them. The town school was especially

¹⁴ Wilma is a widely used educational application. It is an electric environment where, among many other things, teachers can send messages and enter data (e.g. notions and feedback) on pupils which parents can read and react to in their own data terminal equipment.

enthusiastic in volunteering; altogether 21 pupils wished to be interviewed, but only four pupils per grade were invited (n=12); half were selected based on their questionnaires and the other half by drawing lots. The pupil sample of TS is thus partly selective, partly random. Eight volunteered pupils from the city school participated in an interview, which makes a total of 20 interviewed CLIL pupils and approximately five hours of recorded material.

During this process, I paid attention to some basic principles of interviewing children. For instance, I ensured that the pupils knew why and for which purpose they were being interviewed. Additionally, as pointed out by Hirsjärvi and Hurme (2010, 131), one should not rely too ingenuously on children's knowledge of concepts; nor should one express emotions or reactions to children's responses or opinions even when inappropriate or irrelevant. In each child interview, I made comprehension checks to verify the pupils' understanding of the substantive term 'assessment'. The concept was defined whenever necessary, but almost without exception every pupil was capable of giving an adequate definition of assessment. The following examples of children's definitions are translated from Finnish by me:

It's like I would draw a circle and then someone would come to assess. That is, s/he would have a look at it and say if it's good or bad. (Girl, 3rd grade)

Assessment, in my opinion, means that if there's a self-assessment, you must estimate your own skills and what you can. And assessment, well, it's assessed what someone can do and what someone knows. (Girl, 4th grade)

Well, it's how well you master English, how well you learn it and if it feels difficult or easy and how it goes in general. (Girl, 5th grade)

All interviewed parents (n=7) were from the TS. Only one parent from the CS offered to participate in the interview, but when contacted, there was no reply. The total time of the recorded parent interviews was roughly five hours and 30 minutes. Twelve teachers volunteered for the interview, but nine of them actually organised time to be interviewed. One teacher was invited to take part in the interview due to her experimental assessment methods which I became aware of from the parents' interviews. The nine interviewed teachers represent six different CLIL schools situated in Western, Southern and Central Finland. The audio-recorded teacher interviews cover more than eight hours of speech. In short, 20 pupils (12 girls and eight boys), seven parents (all mothers) and 10 teachers (nine women, one man) were interviewed, and the combined recorded interview data is almost 19 hours in duration.

Without any delays, I transcribed all the audio files during spring and summer 2012. During the process, a need for clarification and some new questions arose, and I contacted the given interviewee for further information. I had also offered the adult interviewees a chance to read through the transcribed interviews with the intention to allow them check their answers. Only three wanted to read their transcriptions, and two of them wished to change or specify their ideas. This happened via e-mail. The data analysis methods of the interviews as well as the questionnaires are elucidated in section 5.5.

5.4.2 Computer simulations

In addition to approaching the assessment topic from theoretical and stakeholder perspectives, I also wanted to concretely experiment with a few potential CLIL assessment methods, language portfolios and computer simulations, of which only the simulation experiments are included in this report due to the fact that the portfolio experiment is still ongoing. This sub-section portrays the actual empirical experimentation with the computer simulations built on *the LangPerform concept* (Haataja 2010). LangPerform simulations consist of three main parts: 1) a language biography, 2) the actual film-based simulation and 3) a language laboratory – all accessible via the Internet regardless of time and place as long as the user has been given a valid user name and password (see e.g. Haataja & Wewer 2013). I will describe these parts below.

The first part, the **language biography**, is an area where the users can enter data and details of their language background, for example the languages mastered, where and when acquired and used for which purposes. The second part, the **simulation**, is a film-based narrative with various embedded tasks that the test taker encounters and has the possibility to react to in speaking or writing, for example. The performers' reactions, speech and actions are saved in the **language laboratory**, the assessment area. The language laboratory is located in an external server from where these performances can be retrieved for monitoring and assessment either instantly after the performance or later at a suitable time. The users can self-assess the performance, and/or an external rater, with rights to enter the environment, can give feedback on the performance either verbally or by indicating the level of proficiency in various features (e.g. accuracy, fluency) in the CEFR descriptor line continuum.

PROFICOM simulations for CLIL assessment

The simulations experimented with in this study were produced within the project Profiling Learner Progression in CLIL Environments through Computer Simulations, PROFICOM (2013). This three-year project, from 2012 to 2014, funded by the Finnish National Board of Education, is a joint venture of three agents: 1) the University of Turku (Teacher Training School), 2) the University of Tampere (Research and Development Unit for Languages in Education) and 3) the City of Tampere (Basic Education). The project aim is to design, implement and pilot three computer simulations of which the first two, already completed simulations comprise the data of this study and are perceived as two cases with traits of action study due to my active, subjective and participatory role in the project.

The assessment scheme, as depicted in Figure 16, gives an overview of the LangPerform concept and the assessment scheme in PROFICOM simulations. LangPerform simulation was experimented with for the first time at primary level and in CLIL contexts. New to the LangPerform concept was also the assessment scheme involving the test takers' parents. The PROFICOM simulations represent a new type of a communicative language test particularly designed to measure content-based language proficiency in primary CLIL environments.

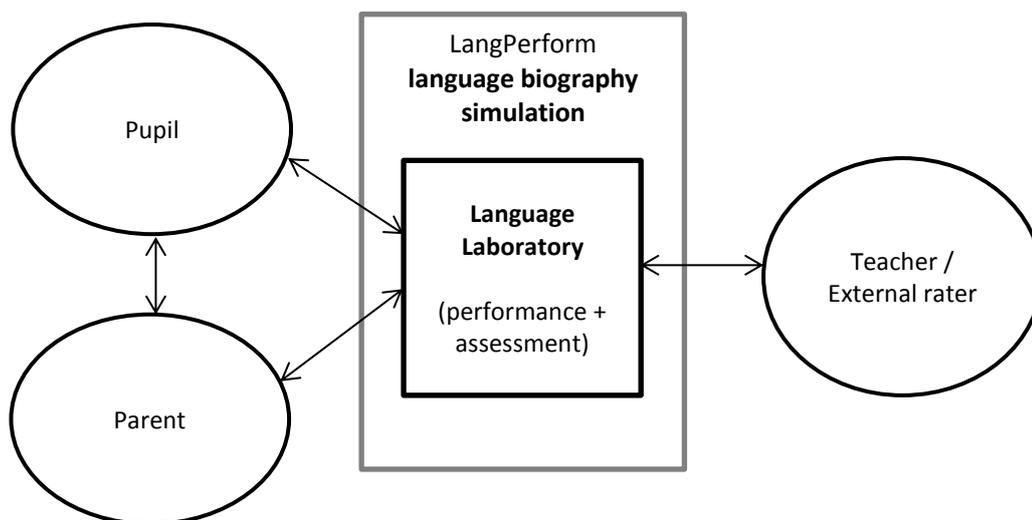


FIGURE 16. Assessment scheme in PROFICOM simulations (based on Wewer 2013b, 106)

These first CLIL simulations can be perceived as prototypes (see Fulcher & Davidson 2007, 76–85) that are created to ensure the feasibility of the simulation prior to the deployment of the final simulation test version. The following two sub-subsections describe the simulation implementation, piloting and the corresponding research methods providing answers to research question 3 in more detail (see 5.2).

5.4.2.1 Implementation

The PROFICOM simulations are film-based, narrative and immersive, and they entail integrated, embedded sections with coded, digital or animated tasks testing CLIL content mastery in various subjects with many possibilities for communicative language use (e.g. Haataja & Wewer 2013; Wewer 2013a; Wewer 2013b). The PROFICOM computer simulations are based on a filmed frame story under the pretence of which children are planted in the TL environment. The first simulation narration is located in Michigan, U.S., where the children are on a school visit and accommodated in a local home. The second storyline rests on a visit in an international school (with no specific indication of its location) received as a lottery prize. The simulation production can be divided into three phases: preproduction, production and postproduction.

In the **preproduction** phase, I was responsible for designing, content (task) production and writing both screenplays. The scripts were, however, negotiated and revised in the simulation team consisting of professional film makers and linguists. Both simulations contain two basic types of premeditated tasks (see 4.2.2 for task-based performance assessment) that had both already been designed in the screenwriting phase and written into the screenplay: 1) *communicative tasks* eliciting performers' everyday language use (e.g. introducing themselves, talking about family, school and hobbies) and 2) *content tasks* that predominantly measure subject-specific academic language proficiency (e.g. explaining how multiplication in columns should be calculated, working in diverse ways with texts containing the genre and vocabulary of environmental sciences). In the second simulation, the sequence of tasks was designed to follow

the cognitive processes in Bloom's taxonomy (see 2.2) from lower order thinking skills (remembering, understanding and applying) to higher order thinking skills (analysing, evaluating and creating). Figure 17 is a screenshot of a content task requiring analysing and some academic language and specific vocabulary. The test taker was to organise the sentences in right order and then read them aloud in the following task. The order could still be altered if needed. The tasks included in the two simulations; their primary purposes, type of tested language proficiency and specific features for assessment are listed in Appendix 11 for the simulation 1 and in Appendix 12 for the simulation 2. For simulation 2, the objectives of Bloom's taxonomy are also included.

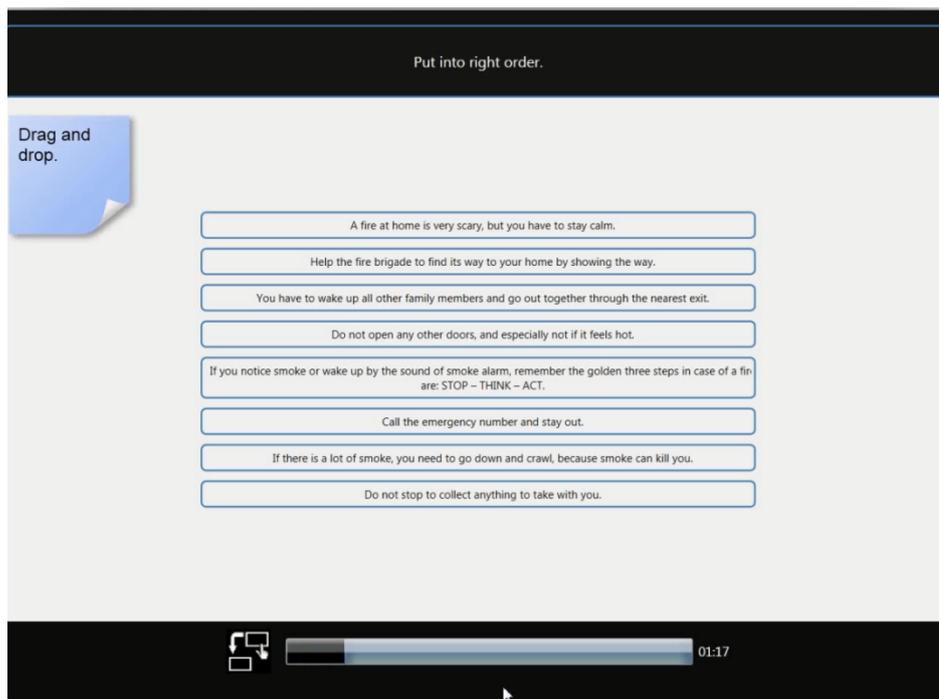


FIGURE 17. An example of a task on the screen: organising text

In the **production** phase, I also acted as a producer in both shootings with various responsibilities, including casting, auditing and preparing both child and adult actors, organising filming locations, procuring props and catering as well as working as a timer. The footages took place in the Turku area, in Finland, and involved a two-membered professional film crew. The actors were mostly native speakers of English or bilinguals with different national backgrounds. After the footage was shot, the films were edited in Tampere and the task integration phase, i.e. **postproduction** began. For the postproduction, I provided visual data and audio files to be added in the tasks. These subject-specific tasks were integrated by a coder based in Germany who also took care of the platform system as a whole.

Intense co-operation and correspondence took place in Turku (I as a screenwriter, producer and process co-supervisor), Tampere (co-supervision, filming and editing) and Germany (coding, system administration). Once the simulation was integrated and captured, it underwent a phase called *alpha testing*. Normally, prototyping consists of alpha and beta testing (Fulcher and Davidson 2007, 80). The former refers to “in-house testing [...] to decide if the design is adequate

and to eliminate any very obvious design faults through expert judgement” and the latter to small-scale sample testing, for example a few people performing a partial test. Both simulations underwent the alpha testing phase, which involved the whole simulation team and included checking and revising, but beta testing was bypassed in favour of field testing called piloting.

5.4.2.2 Piloting

Piloting is a larger testing phase with the actual potential users, in this case primary CLIL pupils. The principle aim of the field testing was to gain knowledge of the affordance of these computer simulations for content-based language assessment and YLLs (see 4.3) at primary level in CLIL contexts for the first time. By affordance I mean the potential of the test to be valid, adequate and practical as an assessment method. The secondary aim was to collect information on how to improve the simulation concept as a whole as well as to specify and fine-tune the simulation functions and tasks.

Both simulations were piloted in the participating schools (town school TS and city school CS) within the same academic school year: the first simulation in autumn 2012 and the second in winter 2013. The participating pupils were not always the same (Table 22); only fifth graders performed both simulations. Altogether, more than 100 children piloted the simulations, of which girls outnumbered boys, at approximately one third of the total number. None of the children participated without parental consent. The introduction letters (one for each simulation, see e.g. Appendix 2 for simulation 2) sent home indicated that the simulations were in a field-testing phase which is why they may not function perfectly. The letter also stated that the performances would not be assessed and nor would they affect assessment at school. The parents were also informed of the possibility for pupils to attend video interviews as well as of the following research questionnaire for both children and themselves.

TABLE 21. Participants in the simulation piloting sessions

	Simulation 1		Simulation 2		Total
	Girls	Boys	Girls	Boys	
TS 4 th graders	0	0	15	6	21
TS 5 th graders	10	11	10	10	41
TS 6 th graders	8	12	0	0	20
CS 4 th graders	0	0	6	8	14
CS 5 th graders	13	5	13	4	35
CS 6 th graders	11	4	0	0	15
Total girls and boys	42	32	44	28	146
Total	74		72		146

TS – town school

CS – city school

The piloting dates and timetables were agreed upon with the relevant CLIL class teachers so that half of the children remained in the classroom while the other half was conducting the pilot, under my guidance, in a computer room or a corresponding space. There were altogether 16 individual piloting sessions. Every group received the same initial instructions on the simulation procedure with the help of a PowerPoint presentation starting with logging into the LangPerform platform.

The language biography was executed together step by step prior to introducing the simulation itself and its screen view (symbols, time bar, etc.), functions (e.g. untimely continuing) and other issues (e.g. test takers' asynchronous advancing in the simulation, provided scaffolding).

In Figure 18, the screenshot of the simulation film is blurred in order to accentuate the symbols and functions. This slide illustrates the time bar function which indicates the elapsing time and how much time is left for the task. In this case, the task is to introduce oneself and the post-it tag scaffolds by giving hints of topics to talk about. The symbol for speaking is highlighted, and in the upper bar, the task instruction is seen in written form.



FIGURE 18. Example of the simulation introduction PowerPoint slide

Each pupil needed a computer, keyboard and a headset to perform the simulation (Figure 19); in the TS the computers were small laptops of various sizes whereas in the CS, there was a computer room with 17" desktop screens.



FIGURE 19. Pupils performing the simulation

The sessions with the CS pupils succeeded technically considerably better than the ones with the TS pupils.

Piloting problems

Prior to both piloting phases, I had tested the simulations several times (double alpha testing) and reported on all problems to the system administrator in Germany. The very thorough technical preparation included pre-testing, purchasing and installing diverse types of headsets to various laptop models, ensuring the laptop battery power charge and so forth. Unfortunately, regardless of all precautions taken, the piloting sessions of TS, which preceded those of CS, were particularly characterised by technical and system failures caused by a number of reasons. One was probably the occasionally poor internet connection which crashed the simulation system in a few instances. Problems with headset software installations in turn resulted in sound difficulties: either the test taker could not hear the simulation sounds through the headset or the microphone was not recording his/her voice. Also, the mini laptop screens proved to be very small for the simulation. The room where the laptops were stored and charged was not appropriate for such performances – pupils were listening to each other rather than concentrating on their own performance. No computer or information and communication technology (ICT) room was available. These technical problems frustrated both the pupils and me, and they were reflected in some of the pupils' questionnaires.

In order to avoid such technical recurrences in the piloting sessions of the simulation 2, the pupils of the TS were transported to the CS, because everything had worked well with the CS computers in the first piloting round and the ICT room of the CS was especially equipped for computer work, although there were no dividing partitions which still remained a problem as there was not enough privacy. Again, regardless of all prearrangements, the piloting was not very successful since practically none of the performances were saved in the language laboratory due to a minor, but crucial, change the coder had made earlier but had subsequently forgotten about.

I had noticed and reported this failure in my pre-tests which I ran several times the previous evening, but I was assured that the simulation was not defective and the piloting would succeed. Unfortunately, it did not, and the majority of performances were not saved. Most pupils, however, gained the simulation experience (a few system crashes occurred this time as well), but due to the performances lost in the saving stage they could not listen to their speech, nor compare their written answers to the default and anticipated ones in the language laboratory. Following from this, the self-assessment phase had to be abandoned. This was a devastating loss research-wise, because the children could not fill in the research questionnaires entirely, nor was there anything for the parents to monitor at home. As a result, I lost virtually the whole set of TS parental questionnaires.

Self-reflection

After finishing the simulation, the pupils whose performance had been successfully saved in the language laboratory were able to listen to their replies, answers and reactions and also read what they had written. The default answers, when applicable, were also visible. The simulation thus gives instant feedback on content mastery and provokes thought regarding language proficiency.

The CEFR-grounded self-assessment system, originally designed for adults or upper-secondary students, was not adapted for primary pupils within the project. The task-related language in the laboratory, however, was Finnish so that pupils and their parents could understand the tasks and content. The film clips and still pictures of tasks were also available for retention.

Video interviews and documenting

Two piloting sessions of the CS pupils were documented by using a video camera and taking photographs. After all of the simulation 1 piloting sessions, in both the TS and CS piloting, the pupils were given an option to participate in a short video interview conducted in the so-called 'diary room' which in both schools was a book storage room. The purpose of these video interviews was to capture and collect the immediate first impressions evoked by the simulation experience. In the TS, I videotaped the interviews myself, but in the CS, three teacher students assisted in the task according to predefined instructions. They also videotaped two of the CS sessions as a whole. The questions posed in the video interviews were:

- How did you feel (having carried out the simulation)?
- What was easy?
- What was difficult?
- What did you notice of your own language proficiency?
- Did you gain something from doing the simulation?
- Were the simulation situations authentic?
- Which would you rather opt for: a paper test or a simulation?
- Why?
- Would you like to try out the same or another simulation again one day?

Altogether, 50 pupils were video interviewed, 24 from the TS and 26 from the CS. These interviews were treated as a complementary data set, and therefore not transcribed. The data, however, is saved and retrievable for further use.

Pupils' questionnaires

When the whole class had completed the simulation, each pupil participating in the piloting filled in the pupils' questionnaire (Appendix 7 for the first piloting round, Appendix 8 for the second) which was slightly different in the second piloting round for the 5th graders who had already piloted once (Appendix 9 contains the deviating questions only). None of the questionnaires were pre-tested, although they were peer reviewed, because the simulation experience is a prerequisite to the questionnaires. I was always present when the piloting class as a whole filled in the questionnaires and encouraged them to ask whenever they needed help or clarifications. For the second piloting questionnaires, I illuminated and defined some concepts (e.g. language structure, grammar) I had estimated to be difficult based on the first simulation data. Some minor wording changes were also made to increase the intelligibility. Both questionnaires dealt with

themes such as understanding instructions, simulation anxiety, language proficiency, appropriateness of the simulation as an assessment method in CLIL and simulation improvement. The two data sets were combined as the questions were mainly similar.

Parents' questionnaires

Parents were involved in the assessment experimentation process (Figure 16) because the NCC (2004) stipulates that parents should be adequately informed of their children's language development in CLIL. The computer simulation scheme allows parents to familiarise themselves with the performance of their children and learn how they cope in English regardless of time and place as long as they have the login information and a valid password. Teacher's language assessment is available in the language laboratory. Additionally, this model allows evidence-based discussions of, for instance, CLIL-type language provision or specific language aspects between the pupil and the parent at home or in the development discussion. It was possible for parents to write comments on the performance in the language laboratory. In the field-testing phase, however, the teacher was not involved as a rater. The performances are still available for later scrutiny and research. Each pupil ($n=74$) participating in the simulation 1 piloting received an envelope containing the parents' questionnaire (Appendix 10) and instructions for how to log into the LangPerform Laboratory using the child's user name and password.

The total amount of data received from parents remained low, 39 filled questionnaires out of a potential 146 (27%). In the simulation 1 piloting round, out of the 74 dispatched envelopes 34 were returned, but three were totally blank and five contained comments such as "The simulation was unsuccessful, the microphones did not work" and "The simulation failed totally, apparently because of malfunction of the equipment" indicating that the simulation performance had not been saved or was otherwise unsuccessful. These questionnaires were eliminated from the data. Only 26 questionnaires (35%) were analysable. In the simulation 2 piloting round, only 13 parental questionnaires were returned, predominantly from the CS piloting sessions because, as explained above, practically none of the TS simulations were saved. These two data sets were, when applicable, combined because they mainly contained similar questions. The second also entailed questions about whether it was the first or second time the parent encountered a simulation and whether or not the simulation was saved.

The low number of returned questionnaires can be explained in several ways. One of the main reasons for the loss of questionnaires was the coding problem which resulted in the simulation performances not being saved in the system. Since there was no performance to examine, the parents could not answer the questionnaire. Another reason may have been the reluctance of some children to let their parents witness their 'self-perceived failure' in the simulation. Becoming acquainted with the simulation surroundings takes time and effort. Additionally, a few parents reported having technical difficulties in the login stage.

5.5 Data analysis

The data corpus of this study consists of various data sets: several questionnaires (CLIL assessment survey for teachers, pupils and parents as well as simulation questionnaires for pupils and parents), audio-recorded theme interviews of CLIL teachers, pupils and parents as well as video interviews of pupils regarding first impressions on simulations. I analysed all of the data personally thus gaining a closer relationship with and understanding of the information. Pupils' questionnaires in all research phases were analysed so that it is possible to draw class-specific (e.g. TS 5th grade, CS 5th grade) and gender-specific (girls, boys) information from the data. This procedure enables the comparison of various aspects when necessary and found relevant, although not guided by the research questions. The heuristic research paradigm is inductive, and therefore open to any patterns or relationships found in the data and thus "more likely to arrive into new insights" (Seliger & Shohamy 1989, 88). Furthermore, when possible and applicable, the pupils' answers were categorised by gender for further and deeper analysis. Then, the class-specific analyses were combined. As to the two pupil simulation data sets (1 and 2), they were merged after the itemised analysis. The same was executed for the parental simulation data. Both quantitative and qualitative data analysis methods were applied.

Quantitative analysis

The numerical data were obtained from the questionnaires by calculating frequencies and reporting them as percentages, thus providing answers to questions such as 'how many', 'how much', 'how often' and 'to what extent'. Commonly in the quantitative analysis in the case of ambiguity in questionnaire answers, the answers were omitted from the data. Such cases were blank questions, multiple answers instead of one, or ticking where ordinal numbers were expected. Due to this practice, the number of included participants varies occasionally.

Qualitative analysis

The majority of the collected data necessitated, however, qualitative data analysis to complement the quantitative data by adding the 'why' or 'how' perspective and depth. A wide range of textual data were collected in this study: open-ended replies and free comments in the questionnaires as well as the transcribed interviews and the video interviews. Qualitative content analysis, sometimes assimilated with thematic analysis (Braun & Clarke 2006, 98), is an intellectual process method used to reduce large bodies of textual data and to categorise that data into "clusters of similar entities, or conceptual categories, to identify consistent patterns and relationships between variables or themes" (Julien 2008, 121). According to Braun and Clarke (2006, 98, italics in the original) the sole difference between content and thematic analysis is that in thematic analysis, "themes tend *not* to be quantified". In this study, thematic and content analysis are considered to be identical and will from here onwards be referred to as thematic content analysis.

Thematic content analysis is an interpretative method independent of any theoretical approach that requires close reading from the researcher and a realisation that non-written texts such as video interviews, as in this study, also comprise textual data and that the “text is open to subjective interpretation, reflects multiple meanings, and is context dependent (e.g., part of a larger discourse)” (Julien 2008, 121). In thematic content analysis, patterns and themes are sought. Braun and Clarke (2006, 87) differentiate six individual phases in conducting thematic content analysis:

- 1) familiarising oneself with the data (reading and re-reading, noting ideas)
- 2) generating initial codes (systematic coding of interesting features)
- 3) searching for themes (collation of codes into themes)
- 4) reviewing themes (checking the relations between extracts and the data sets, data map generation)
- 5) defining and naming themes (refinement of themes)
- 6) producing the report (final analysis, selection of extract examples, checking interaction with other research components)

Although the textual data set was relatively extensive, no computerised coding tools were available. Through intense reading, I marked, coded and categorised themes using a traditional method: colour identifiers and pencil marking. I also searched for extracts in the Word documents by using key-words. Since the data corpus is large and partially rather complex, it is not relevant or even appropriate to present the analysis of every single questionnaire or interview question in this research report – the results are mainly presented thematically as clusters in the following two chapters.

SUMMARY

This study is a mixed methods study with qualitative emphasis pertaining to the year 2012’s current CLIL assessment practices, their adequacy and the appropriateness of computer simulations in CLIL assessment. Therefore, the research was conducted in three phases: 1) the CLIL assessment survey, 2) experimentation of simulation 1 and 3) experimentation of simulation 2. The data were triangulated by sources (questionnaires, theme and video interviews as well as documentation) and participants (teachers, pupils and parents). The amount of data is extensive: 250 questionnaires related to the CLIL assessment survey and 185 questionnaires in connection with simulation experiments. The interview data is comprised of almost 19 hours of transcribed audio-recordings in phase 1 and 50 individual video interviews in phase 2. Phase 2 was partially documented by filming and taking photographs. The data were analysed using both quantitative and qualitative methods: calculating numerical frequencies and percentages as well as executing thematic content analysis.

6 CLIL ASSESSMENT SURVEY

The purpose of the CLIL assessment survey was to provide answers to the first two sets of research questions (see 5.2) pertaining to CLIL teachers' assessment practices and the adequacy of these practices as perceived by primary CLIL pupils and their parents. Furthermore, identified challenges, wishes and visions for improved practices were collected. In other words, the goal was to find out what kind of language assessment methods, if any, are used in Finnish CLIL classrooms, and to determine if the assessment practices are adequate, sufficient and frequent enough to cater for pupils and their parents. Moreover, the goal was to investigate how CLIL assessment could be developed so that it would better serve those concerned.

All three parties of assessment (teachers, pupils and parents) were included in the survey in order to compose the most accurate understanding of the phenomenon. The data were collected using semi-structured questionnaires and elucidated by elaborative theme interviews with volunteers. In the questionnaires, it was clearly stated that the questions concern *content-related English* only, i.e. *English used in other than language subjects*. The most relevant survey results are presented in this chapter following the order of research questions and starting with teachers' assessment practices (6.1) which include methods of providing feedback (6.1.1) and frequency of feedback provision (6.1.1.1). Then, I will move to the challenges of assessment in CLIL (6.2) and adequacy of CLIL assessment (6.3). Section 6.3 presents the views of pupils and their parents, and is divided into three sub-sections: perceived feedback (6.3.1), frequency of feedback (6.3.2) and sufficiency of feedback (6.3.3). The final section 6.4 involves with preferences and future wishes of all stakeholders concerning assessment in CLIL. Each section is concluded with a brief summary of the main results.

Whenever participants' quotes from questionnaires or interviews are given, they are marked by the abbreviations T (teacher), L (learner i.e. pupil), P (parent) or R (researcher) to indicate the interlocutors in that discourse, unless otherwise obvious. The original quotes were, with a few exceptions, in Finnish. Therefore, I have translated the excerpts into English. I have striven to translate as accurately as possible, simultaneously trying to retain the original tone and style. Obvious traits of spoken language such as 'like', general extenders, false starts and repetitions have mostly been ignored and the quotes modified towards norms of written language, because this is not a discourse analysis study – I prioritised conveying the meaning of the quotations. Due to limitations in space, I do not include the original Finnish versions to give more room for the participants to voice their opinions and ideas.

6.1 Assessment practices in CLIL

This section and its sub-section 6.1.1 attempt to shed light on the research question ‘*What kind of assessment practices do CLIL class teachers employ for assessment of English language proficiency in CLIL?*’ and the first part of the question ‘*To what extent do CLIL class teachers assess language and provide feedback to pupils and their parents?*’. The second part of that question is addressed in 6.1.1.

According to the questionnaire data, **assessment of language in CLIL was considered important to varying degrees**. The perceptions of importance did not vary significantly according to the grade level the teachers were instructing – assessment was thus considered important throughout primary education. Almost half of the CLIL class teachers (20/42) considered assessment in CLIL either highly or very important, and 16 of the teachers regarded it rather important.

This perception is in slight contradiction to the finding that **assessing pupils’ English proficiency in CLIL contexts does not seem to be an established practice**, for nine teachers (21%) disclosed that they are not assessing systematically (Figure 19). These nine teachers were predominantly class teachers with no language studies or class teachers with basic studies in English, and they were evenly representative of grades 1–2, 3–4 and 5–6. The reasons behind not practicing CLIL-related assessment may not always lie in the individual teacher but in the institutional practice or in the resistance or ignorance of management, as the following example from a teacher interview exemplifies:

R: Tell me first what kind of thoughts the issue of CLIL assessment awakens in you.

T: Actually during this spring I have become aware of the necessity of it. Firstly, regarding the child perceiving his/her own language proficiency but also regarding feedback for parents stating where we are going in the language development. [This awareness] has awakened in the discussions with parents because, at the moment, we don’t have [assessment in CLIL] in our school and the decision-maker has not taken a positive attitude towards it. Regardless, children, already when little, and their parents, have the right to receive feedback on the language acquisition.

This quote illustrates how decisive the role of the management is, and how individual teachers may conform to the prevailing practices without questioning them. This teacher was not the only one revealing that in his/her school, no language assessment occurs in CLIL contexts. Subtle signs of a customer culture were also detected in this research: the pupils and parents (customers) bring forth an issue, often a source of discomfort, and the teachers or school (service providers) adapt accordingly. A curriculum predefining the content taught in the TL prevents this from happening.

Almost half of the participating teachers were teaching the first or second grade (see 5.3). This may have had some influence on the results concerning the used assessment methodology and outlook on the necessity of assessment. The following quote illustrates how various reasons related to affective factors, differences between children as language learners and building

passive bilingualism (Cf. Table 10) are used to justify ignoring language assessment during the first two grades.

They were first and second graders, little pupils, so I didn't test it [language] in any way. In the primary school where I have mainly worked, to my view, it's about building passive language proficiency. And many are slow in starting to speak and produce. I'd like to think that we create a positive attitude towards the language and learn to use it without measuring that too much. And then, another thing is that pupils are different language learners, so encouraging is very important. It might be awkward to notice that "I haven't learnt at all" and compare a lot.

This teacher also perceived assessment (measuring) as a psychometric action involving comparison, yet acknowledges the importance of feedback. The teachers' perceptions of assessment varied considerably – a discovery made also in other studies, also in relation to CLIL (e.g. Serragiotto 2007).

Some of the teachers of first two grades articulated that the role of assessment in the first years of CLIL is not eminent or noteworthy, but some also insinuated that the role of assessment in those grades could be more substantial. Similar remarks were given in respect to upper grades as well. This signifies that **the issue of language assessment is thus controversial and based on teachers' own perceptions on the necessity and nature of assessment rather than the curriculum**. In the following, I will investigate assessment methods more closely.

Frequently used assessment methods

The **most frequently mentioned assessment methods** by CLIL class teachers (n=42) were **mostly**, in Finnish contexts, **traditional** methods such as teacher observation, tests or test sections in English (i.e. bilingual tests), dialogic interaction and pupils' self-assessment (Figure 20). This result is logically in line with the study of Bovellan (forthcoming) which concludes that the tuition and therefore also learning, is still rather teacher-centred, and thus traditional.

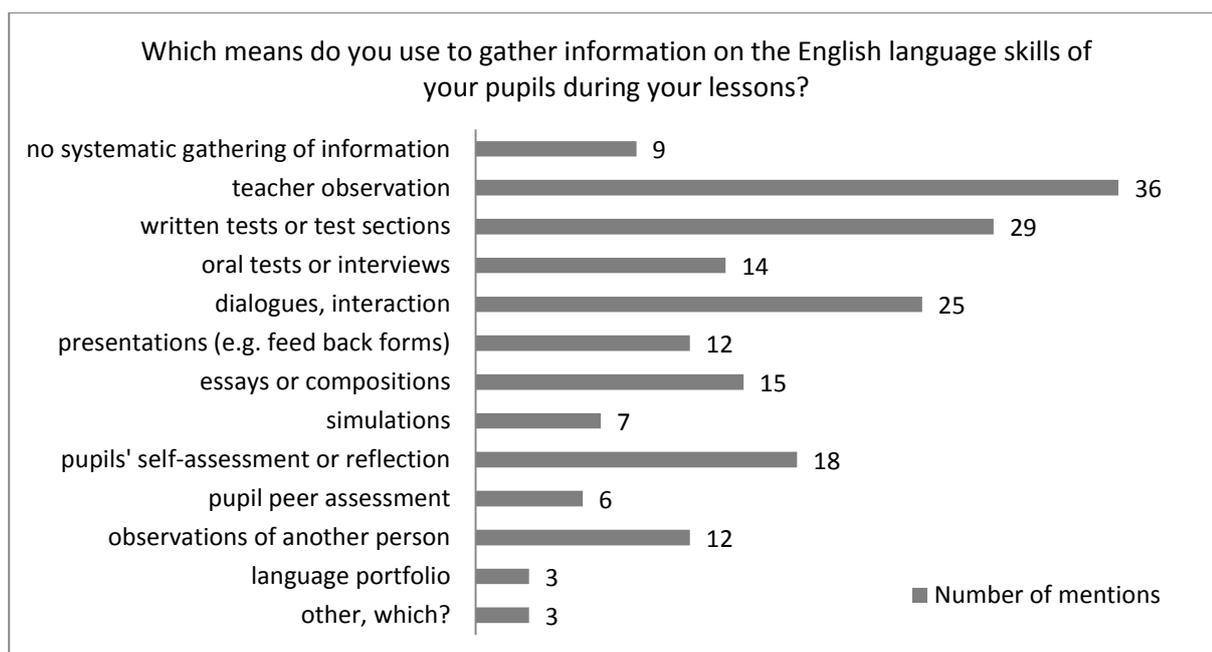


FIGURE 20. Means of collecting assessment information on pupils' language proficiency

Teacher observation

The most popular means of gathering information on pupils' language proficiency and its development in primary CLIL contexts **is teacher observation**, as portrayed in Figure 17. Teacher observation was rarely mentioned as an assessment method by teachers in interviews or in the open questions of the questionnaire; it may be that it is perceived as an inherent and self-evident assessment method. A teacher of grades 1–2 describes her assessment of pupils' language proficiency during the elaborative interview:

R: If I understood correctly, you don't assess pupils' language proficiency during grades 1 and 2 in any way?

T: *Well, I do assess it. That is, I make observations of it and see how active pupils are during the lesson, but I don't do things like giving a paper and then they have to be able to do certain things, write in English and such, and then I would gather the papers and see how it went. I haven't done that. At some point I did it, but now I haven't.*

R: So it is mainly observation?

T: *It is precisely that.*

It seems that teachers make observations “on-the-run” (McKay 2006, 141), hence practising instantaneous, implicit assessment which is part of the cyclical assessment–decision–instruction structure (see 4.1.1). Observation opportunities emerge in the everyday classroom situations as many of the practices the teachers described in the subsequent passages indirectly show, but it remains unclear how these observations are actually used for feedback of different forms. It is likely that teachers keep mental records of the observations and use them as the basis for gut-feeling assessment, i.e. creating an opinion of the children's English proficiency.

I don't see it necessary to assess the language proficiency of individual pupils in the first two grades in [name of the school] – the skills of the group, though, I do.

Assessment is fairly often based on pupil activity and gut feeling.

Pupils' working skills are part of process assessment as is activity shown during lessons. As apparent in the preceding quotes, behavioural aspects and affective factors are open to observation. Content-related language proficiency is likely to be assessed using different methods, as shown in the following passages.

Bilingual or monolingual subject tests

It is not surprising that **the second most frequent method of assessment was paper tests or test fragments in English**. The **test practices seem to vary considerably** according to diverse factors such as the extent of English exposure, model of CLIL provision, teacher personality and materials available. Among the various tests, test practices and assessment principles, several categories were identifiable ranging from test types to test-related practices.

- weekly spelling tests
- content vocabulary tests (e.g. in physics)
- L1 (Finnish) content tests with sections translated into English
- tests and tasks according to language used

In the absence of authentic tests or the ability to produce tests or test items in the TL, teachers tend to translate test tasks. This presents a danger of inauthenticity. Depending on the grade level, the amount of the TL in tests seems to vary substantially. 'The older the pupils, the higher the ratio of English' appears to be a rather common rule. Code-switching is also occasionally allowed sometimes resulting in 'Finglish'. The underlying idea is to encourage using even the most modest language resources available. The following practices and quotes reflect the subjective pedagogical freedom of Finnish teachers to adopt assessment systems they see fit for the given group, context and purpose. They also depict the controversy in CLIL assessment.

- what has been taught in English will usually be tested in English
- possibility to choose the language of answers or switch codes
- extra bonus given of using English language
- willingness to interpret test answers and ignore language errors
- adjusted rating scales and different/same standards

We are trying to unify our standards at grade level. We have done common rating scales. It is not obligatory to use them but quite many do. The tables are for basic tests of different subjects.

We also have a right to compose different tables and assess differently from our own group. Also the tests are different and their standards are different.

- linguistic preparation and selected contents in English

One teacher in particular had a practice I will call supplemented tests, the purpose of which is to supplement any possible gaps in learners' content knowledge. The tests were primarily in English, but at school, code-switching was allowed because the teacher prioritised manifesting content knowledge over linguistic knowledge. At home, the children must produce test answers in English using any source available. The benefit reaped from such a practice is that it enhances pupils' awareness of their own skills in an unthreatening way.

- supplemented tests

We have paper tests because you have to have them. When the test is done, if the children don't check them themselves so that they have their own papers, we have a look of the answers together. But if I check them at home, when returning the papers, we go through every single answer. After that I have usually allowed them to take a blank test paper home which they then have filled in using internet or work books, whatever, together with their parents. Fixations about numbers [grades] have become less, now that they are doing the test once more at home so that they secure their knowledge.

The following points clearly show how content assessment and language assessment are strictly separated from each other. This may indicate two things: either language is tested in EFL contexts or subject-specific language is not assessed at all.

- testing content, ignoring language

The essay questions in the 5th and 6th grade are quite demanding; those are not tiny little things. But then again, the English language is not assessed in them but the content matter.

Subject tests are not language tests. Language tests are separate.

At primary level, the complexity of academic language needed for content study varies between the beginning (grades 1–2) and upper (grades 5–6) levels. Therefore, pupils' academic proficiency needed for understanding subject-specific texts and producing them may not be high enough as indicated in the quotes underneath. The language proficiency + 1 principle is important to keep in mind in order to allow the TL to develop beyond standard EFL towards more academic command.

- less demanding content in English

There is no escaping that the things asked in English rather than in Finnish are easier.

In my opinion, what I ask in English is really easy.

The role of L1 (or the language of schooling) was also perceived as important. In bilingual education, the role of both languages needs to be carefully considered, especially in multicultural environments because both languages should be developed equally.

- opposite language

*We have an English language math book, *Laskutaito in English*, so our math tests are in Finnish. Also because this shows where we go regarding the Finnish language and we also practise in Finnish for the tests.*

Regarding the ratio of English in tests, one interviewed teacher pointed out that the school CLIL curriculum requires the use of English in testing: "In the curriculum of our own school, there is an article stating that when we pursue having around 25–30% in English, so also the tests have around 25% [proportion of English], maybe 30%." Indeed, the curriculum of that school states that during grades 5–6, the pupil uses the target language actively in diverse situations including tests, which are specifically mentioned.

CLIL Group tests

A few of the interviewed parents as well as some pupils mentioned a test type they called "a group test". The **group test is a form of collaborative testing** (see 4.2.1). One parent, in describing group tests, pointed out the spirit of ZPD (p. 28) and current socio-constructivist views on learning:

In my opinion, it has been actually really nice that now, when they have had more essays and larger issues, they have engaged in group tests in which you can benefit from your more capable peers. It is not that "I didn't know" but collective seeking. The value of learning or assessment of skills does not decrease terribly if you can absorb things from your friends and notice that 'wow, that kid's really skilled, I'd like to be as skilled as well'. The juxtaposition of skills is not that pronounced when you do group things like these which I have considered nice.

I contacted the relevant CLIL teacher and invited her to an interview in which she explained her group test concept which I will summarise here. The origin of the CLIL group test was in music, recorder playing tests, and the purpose of group testing was to bring new dimensions to traditional solo playing. The group performance also alleviated pupils' test anxiety, the teacher noticed. She decided to apply the concept in other subjects as well: religion and environmental sciences. The basic idea was "combining, assembling pieces and joining them together as well as introducing English in group work". The first experimental group test activity was related to biblical Easter events and consisted of three main parts: 1) motivation and attuning, 2) group work and 3) the actual test.

Pupils were allowed to form pairs which the teacher then coupled into groups of four: two girls, two boys. The first, motivating and relaxing task was to create a 'sound pilgrimage' – a sound representation of the Easter events which the audience listened to with closed eyes when each group performed their part sequentially. The next task was to create a traditional, written poster-like group presentation with mind maps and visuals. The posters were partially in English. Pupils made notes when other groups were presenting after which the notes were once more checked together so that everyone had the information considered essential by the teacher.

When the group test was announced, the teacher observed that some pupils started sharing responsibilities and content areas according to their strengths, thus creating expertise. The initial idea of the teacher was also to observe how their communication and working skills are enhanced. She also let them seek preferences in learning and working styles and encouraged the development of those preferences and styles further. In the actual test situation, each group adopted individual strategies: some worked everything through together, while some delegated pages and tasks. The tasks ranged from simple connection activities and statements to more demanding open-ended essay responses based on Bible extracts and the production of a mind-map on a given topic (Cf. Blooms taxonomy on p. 29). The language choice was optional: English or Finnish and code-switching were all allowed.

The teacher reported that, on the one hand, pupils had estimated the Bible extract task as especially difficult, but on the other hand, the group support enabled them to manage it: "It wasn't too outreaching; they really had to converse and think it over as a group which was the intention also". The group received a mutual grade for the test which was also communicated to parents. Space-wise the test was demanding, according to the teacher, because the test type entails conversation. The learning environments in schools are not designed for such activities and the room utilisation capacity is normally high. Therefore, it was not possible to disperse the groups around the school.

The next group test was about domesticated (farm) animals, but the implementation was slightly different. Each group chose an animal and made a mind map which was copied to other groups after the teacher checked it. Then, pupils as a group designed test questions of their own animals from which the teacher composed the group test, adjusting the questions when needed. The groups answered their own questions as well as those of others, which prevented the

temptation to design too demanding questions for others to answer. The teacher reported pupils' feedback on the group tests as follows:

Many said that the work load is not that huge when you can share it with others. Collaboration pleases them, although many reported that it didn't always succeed. Working together was fun, they felt. Then, about the development of English language: when reflecting the mind map tasks and their own skills in creating information of Baltic countries in English, they thought it was useful when they could use a dictionary and work together. And learning new things – they had to compromise and ask for my presence and let me say how to proceed. So it wasn't always easy.

Since the initial experimentations, the teacher has further developed the CLIL group test concept by altering the following variables: number of group members (pairs); use of tools (dictionaries), individual test in addition to the group test (multiple tests) and function of the group test (a group test as a "warm-up", e.g. vocabulary brainstorming, preceding the individual test).

Dialogic interaction

Interaction and dialogue were the third most common assessment methods used by CLIL teachers in this sample: six out of ten teachers informed having using various types of interaction for assessment purposes. I will combine the terms and refer to classroom interaction – whether between the teacher and students or students themselves – as dialogic interaction. One interviewed teacher stressed the importance of dialogic interaction as the backbone of his instruction:

Let's say that my whole teaching ideology is based on dialogue and conversation and especially on language use in English and that I'm not a talking head there. But children solve problems among themselves and the working on the whole is very problem-centred. And we do a lot of project work.

This assessment method implies that the TL in spoken form is actually present in the lessons and used by both teachers and students. It seems that teachers who are confident in their own use of the TL in spoken communication are more adept at playing with language and perceive it as an active tool in everyday classroom communication.

It is not that you falter or stop thinking things, but also the children need to get the experience that it's a language in use. It's all the time there. You're thinking about it; you explain to the children how certain things are said. You can suddenly just [...] say something funny. You have to be relaxed enough with the language.

Such language use requires good TL fluency from the teacher. If the language is naturally and constantly present in the CLIL classroom, the pupils will also more quickly reach the point where they start to express themselves in English (Cf. the Silent Period and Language Reservoir in 2.2). In order to provide pupils opportunities to use English meaningfully, teachers need to organise English **language use situations in which they can observe language in use and gather assessment information** in other ways. The following thematisation of such situations was to be formed from the interview data; the teachers described TL use situations in which language assessment takes place.

- teacher-initiated discourse (IRF pattern)

Assessment occurs also so that [...] I always shake hands with pupils when they go home and then I always ask a question [...] such as 'Tell me what you did last evening', and when they come, they desperately think and tell in English what they have done. Or I could give a calculation and they have to solve it in English or they tell what they are going to do today.

- situational use of language

We have agreed on a principle that our meal language is English. We always speak English in the lunch hall.

- soirées and performances

A short fairy tale musical, there are these songs naturally [in English].

The pupils are responsible for the programme, I don't interfere with that. They announce themselves and when they prepare drama plays; their linguistic level comes across terribly well. For example, what kind of lines they have written for themselves – if it is just "yes" and "no". And then the more enthusiastic or gifted ones, their lines are from another planet.

Various performances related to thematic gatherings seem to be common in CLIL contexts, and they provide plenty of opportunities to display communicative and creative language use, as the previous examples show. Drama and presentations can also be harnessed into pedagogical use as in the following examples. **Collaborative work and study methods**, in this data, **appear to be fairly often directly related to content study and subject-specific language use**. It seems that such methods provide ample assessment opportunities in relation to language use.

- pedagogic drama

In relation to the water cycle we have 'the Story of Droplets' in which Droplets discuss. And those [activities] are proportioned to, for example, third graders and allow them to participate and get enthusiastic.

- talks, presentations and interviews

- group work, subject-related projects

Project work fairly often stems from topics of geography. And then in physics-chemistry the pupils have been able to choose a topic to show their [language] skills. The space is a typical topic. Group work and presentations are usually prepared of the topic 'human', not booklets that often, but sessions where we teach each other. And then we use collaborative learning quite a lot: tasks in which we work together and organise learning stations.

With the current group, we organised a so called Science Fair about all contents of the 6th grade: physics, biology and prior stuff. They could choose the most important topics and then they had two weeks' time to organise a Science Day. Then everybody from the first to the 5th grade, English classes, circulated around and the pupils explained those phenomena to them.

The world has become smaller; classrooms are also more open to the world. Teachers in this research sample, in general, appeared to be active in organising contacts with native speakers or people who study English as a foreign language. Such activities provide plenty of opportunities to natural language use. Among the teachers were fairly many whose pupils had chatted or emailed with or wrote letters to their peers in other countries.

- Comenius projects and twin schools
- interaction with foreign visitors (e.g. artists, athletes, teachers, students)

These examples highlight how the TL development in primary CLIL tuition is holistic and comprehensive. Sometimes, but not always, these situations are related to the use of everyday English (BICS). The above quotes and theme areas exemplify well various language use situations in primary CLIL classrooms – however, often without clear assessment intentions. It is possible that such situations are in connection with implicit teacher observation. The teachers were not very explicit about how the dialogic interaction is assessed, to which purposes and how this assessment information is further elaborated and conveyed to the other parties of assessment, pupils and their parents. This requires further research.

It also became obvious that some CLIL teachers teach formal English to their own class which naturally gives many more opportunities to observe the spoken language in general as this short quote implies:

Then, in my opinion, contemporary books have quite much pair work; and then as extra practises we do those pair conversations, group conversations. So if I have a lesson with half a class [...], I have time well enough to walk around, observe and listen to how their utterances are.

Differentiating between the two varieties of English (BICS and CALP) does not seem be inherent. In everyday language use at school, these two may often merge.

Pupils' self-assessment

Self-assessment is an assessment method favoured and encouraged by the NCC (2004, 264) - a principle drawn from the Basic Education Act (1988). However, the method **was not copiously represented** (43%) in this sample. Some CLIL teachers viewed self-assessment as a method that advances the development of pupils' language awareness and makes the language development more explicit to the pupils. The following quote makes this view clear:

I personally am a very devoted developer of especially pupils' self-assessment; for example, by using a language studio and letting pupils to monitor and reflect their own progress. In my opinion, linguistic awareness is the prerequisite for the genuine development of pupil's language proficiency, and I don't think that the contemporary teacher-centred external assessment solely makes this possible. In my view, several possibilities to reflect own language proficiency and progress should be created during the semester and school year.

Some CLIL teachers regarded self-assessment as an assessment fashion that has already seen its height, but in the same token, teachers described various, effortless and simple ways to practise self-assessment: thumb assessment, smileys, marking estimation on a biased line continuum as a reaction to a statement or a question. Self-assessment is already practised in the first two grades.

In some schools, CLIL-related self-assessment is part of the official reporting system and the self-assessment documents are sent home for parents so see and sign. Working skills (assessing process in addition to the product) are also included on some occasions, an example of which is the interview excerpt below.

T: *We are always conducting a self-assessment related to working skills in the middle of the autumn semester, and then in the spring term we do one related to subjects. It also contains working skills to some extent, but then there is a separate attachment for CLIL classes.*

R: *What are pupils asked in relation to CLIL studying?*

T: *Well, actually all kinds of things: their own outlook on things such as in what kind of situations they use the language, how they are as language learners, as children. There are quite many different points.*

It was back then when those self-assessments were tremendously in. These trends vary. So we had various self-assessments of attitudes towards language use, how you can express yourself in spoken and written language and a little bit of what you like and what you don't and how you have succeeded.

Instances of exploiting modern technology in assessment were also in the data. One teacher in particular had experimented with various computer-assisted language assessment methods as in the following pupil self-assessment of pronunciation and reading skills combined with teacher feedback.

A simple example is an experiment I did with my pupils. I asked them to read a certain text in the computer, and the act of reading was recorded by a web camera. After this the given pupil could listen to the performance by him/herself without any disturbance in a room so that social pressure and other complications were minimised. Having listened to the performance, the given pupil and I had a one-to-one conversation, and the pupil could assess his/her own skills and level of language proficiency. Such an exercise was pleasing to the pupils, but comfortable and trusting atmosphere and relationship between teacher and pupil are prerequisites for this. This experiment focussed mainly on language proficiency instead of content, but this could certainly be applicable to self-assessment of content-related language proficiency. But how? 😊

As the teacher mentions, such an assessment situation requires appropriate space and equipment in order to succeed, and the discussion with the teacher afterwards makes this an example of dynamic assessment occurring in the pupils' ZPD (see p. 28). The benign question the teacher expresses at the end of the report is relevant to this context and depicts how **teachers in the field are in need of methodological tools, assessment principles and advice.**

Less frequently used means of assessment methods

Among the less popular means to gather assessment information on CLIL pupils' English language skills (see Figure 17) **were many assessment methods that can be viewed as alternative.** In order of frequency, they were essays or compositions (15 mentions), oral tests or interviews (14), presentations and observations by another person (both 12), simulations (7), pupil peer assessment (6) and a language portfolio (3). Some of these less used methods were already indirectly mentioned in previous paragraphs. I will dedicate separate passages for native teachers and portfolios.

Three comments were obtained under the depiction 'other means of assessment': interestedness which falls into category of affective factors or working skills (process), projects in various subjects in English and a more lengthy explanation which I quote underneath as a whole [Quotation marks in the original].

I don't teach formal English to my class. English proficiency is manifested foremost in connection with content mastery. Especially in each mathematics test, one page is in English, but also in the test comprehension is more emphasised than exact competences. The topic [language proficiency] is discussed in development discussions. I don't gather systematic written information. Essays and compositions, for example as a part of a history project, are done in English. These all come out as side product of "normal" classroom work.

This quote reveals that English is used for content study but it is not assessed in any particular manner. The first sentence implies that language assessment occurs in EFL contexts.

Essays and compositions seem to be utilised primarily in the upper-primary grades (5–6), and those teachers who commented on longer pieces of writing stated that language has a minor role compared to content. Not all teachers pay attention to the correctness of language, but some do. The age of pupils might also play a role in how important the correctness of language is perceived to be. Pupil peer assessment and simulations were hardly experimented with or mentioned by the interviewed teachers. A safe atmosphere and responsibility in giving “fair and to the point” assessments “irrespective of the persona”, as one teacher put it, is the prerequisite of successful peer assessment. The inference is that pupils are not seen as inherent feedback providers.

If someone said something, they remarked that 'you could say it this way as well' and 'are you looking for this word' or 'you are expressing that really well'. Especially the presentation review situations were wonderful.

In the above quote this teacher describes a situation s/he interprets as peer assessment. It could also be interpreted as mutual negotiation of meaning and an example of dialogic interaction which was the third most significant means of assessment indicated by teachers.

Native teachers and external language sources

Observations obtained from an external source, for example native teachers and language assistants, **were significantly present in the data.**

We just had a Turkish Comenius assistant teacher, so related to external assessments, s/he has given a few comments on language fluency and comprehension and so forth.

We have another native teacher teaching one lesson per week. At the moment it goes by the name English Club. [...] It is for the older pupils so that some have it in the autumn, others in the spring so that it is every week, the class always divided half and half. And it is specifically conversational but with different activities. We don't have a curriculum for it yet.

Many schools, according to teachers' accounts, favour employing native speakers with various backgrounds to enrich the pupils' linguistic environment. The teacher accounts revealed many aspects to using native teachers. The native speaker, who is not necessarily a language teacher, often works with pupils separately instead of co-teaching, and the emphasis appears to be on spoken language. **The native speaker is perceived as a proper language model who gives linguistic support to the teacher** as in the following statements.

I am always worried about my language – it regresses terribly [...] so that side of the matter is then taken care of.

R: Is there anything else you would like to add concerning assessment we still have not discussed?

T: *Executing oral assessment, that is one. Textual assessment goes, but oral assessment [is difficult]. Also that the assessment should not pertain to word or concept level only but different levels such as pronunciation and so on. It stays, however, at the teacher opinion level, and there the native teacher is needed. S/he would hear better than I, who am not even a language teacher of my main profession, although I have obtained a qualification for primary level. So that is what I miss.*

These quotes above illuminate how the linguistic expertise of a native teacher is perceived as compensatory for teachers' linguistic deficiencies and is complementary in reinforcing teachers' own impressions of children's language proficiency. Conveying information on pupils' communicative proficiency may not occur on a regular basis or using formal documentation.

A native teacher may also be directly involved in gathering information on pupils' language proficiency as was the case in several schools. This happens by recording pupils' speech.

The native teacher records everybody [talking] – I don't know how much teachers listen to the recordings. It is a system that has been working in this school at least while I have been here. I thought, then, that this would be a nice way to assess, to gather a sample from each pupil during six years. [...] I don't know if the teachers get to listen to what has been recorded. But I do that out of interest and I have files for grades 1, 2, 3 and 4 and just a while ago I listened to them.

How these recordings are actually used for assessment, other than the teacher listening to the samples him/herself, remains relatively unclear. The teacher said that s/he prepares an audio file of those recordings and gives it to the pupils as they are leaving the primary school at the end of the 6th grade. Apparently the recordings are not employed during the primary years as the next quote shows.

For the first time this year children asked what's going to happen with these recordings. For the first time: 'What are they for?' Now they would like to listen themselves. When I was relocating them after school hours, some pupils were here and they were listening: 'hey, that one is talking there'. They got interested: 'I'd like to listen to this as well'. It would be nice that I'd let everyone listen to their own production here in the middle [of primary years] because they are interested in them now. That would be nice.

The quote above also shows how pupils are interested in their own linguistic performance and how gathering audio recordings for no specific purpose can be perceived as a wasted assessment opportunity. Modern technology provides numerous, fairly inexpensive possibilities to gather linguistic data. Another teacher reported plans to buy memory sticks for each older CLIL pupil for storing textual and audio data. In such cases, the memory sticks work as electric language portfolios.

Portfolios

Language portfolio work was the least used assessment method in this sample. Some of the interviewed teachers said that they have experimented with portfolios at some stage of their careers or had plans to do so, but none of them used language portfolios systematically, as revealed in the following interview quote.

R: Have you used portfolios?

T: *Yes, to some extent, but personally, I have abandoned them. I think no one else has used them. It is just quite hard together with all other things. It's difficult to find time for it [...]. We basically collect papers in a folder and then bundle them. But it is not a structured portfolio.*

The portfolios were manifested in different forms, and were viewed as elaborate and time-consuming but adequate on an ideological level. Portfolio work was portrayed by some teachers as an 'assessment fashion' in a similar manner to self-assessments. Seeing alternative assessment methods as passing fashions gives more room for traditional assessment methods, as already seen in this study (Figure 19). **The concept of a portfolio was ambiguous and widely perceived** by the teachers; portfolios were paralleled with a Comenius project or a notebook in which pupils can search for English words alphabetically. Furthermore, finding the appropriate structure for the portfolio was seen to be difficult, as becomes evident in the next quote.

I have tried [to use language portfolios]. I have saved pupil work and such but finding the time is difficult within this instructional frame. It was one of my big ambitions, but at the end I didn't find the time and structure for it. [...] It's good that a child gets feedback on the language mastery over a longer stretch of time. Then s/he can make comments, gather material and so on. You have to sacrifice at least one lesson for portfolio work so that it really is a portfolio lesson. Yep, these are the dreams.

However, a few teachers in the sample had experience in systematic, linguistically enhancing portfolio work.

We'll do a large portfolio, a folder, during two years, in which we'll systematically attach self-reflections of all those certain things in English systematically. The teacher will collect them every now and then and comment on them. [...] It teaches them to assess their own work and the language comes sort of aside, in passing. But it is always totally in English, which is the point.

Showing, sending or presenting the produced language sample contributes to the sense of meaningfulness. The key to successful language documentation is the meaningfulness of the language use leading to authentic assessment, as one teacher pointed out.

6.1.1 Provision of feedback

This sub-section is related to the methods of assessment in so far that, as stated in 4.1.1, feedback is an inseparable part of the assessment process. There is little, if any sense in gathering assessment information and then ignoring the other parties involved in assessment, the pupils and their parents. When teachers are forwarding any kind of information obtained through assessment to stakeholders, they are providing feedback. Figure 21 demonstrates that **not all CLIL teachers** in the sample **give pupils feedback in relation to their language proficiency** in CLIL contexts. The explanation for this may be, as in this comment: "first grade, I do not assess". As was already noted in this section, **it is obvious that some teachers do not relate language assessment to CLIL at all**, or they justify this lack of assessment by the fact that their language

learners are very young. In the following, I will tap into some of the feedback methods, and the frequency of feedback with further examples that will be addressed in the following sub-section.

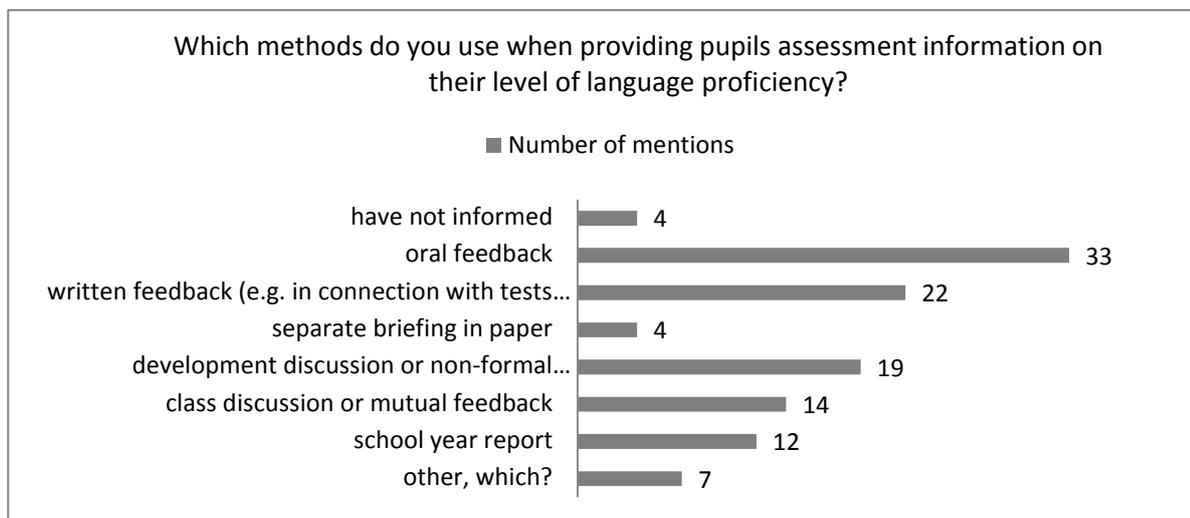


FIGURE 21. Means of providing assessment information to pupils on their level of language proficiency

Oral feedback

Oral feedback was by far the most common (79%) method of giving feedback to pupils on their own language proficiency (Figure 21). One interviewed teacher described his/her feedback practices in the following way.

T: I give feedback pretty much all the time.

R: How?

T: Well, we write a lot and speak a lot. We do all kinds of projects and my aim in all these is that the children themselves identifies how they have progressed and if they have progressed and how the language comes off. And we have recorded their speech, perhaps listened to it then. During the year I always personally give feedback to them. Of course it takes time and there is quite a lot to organise, but in a way it comes quite naturally.

In this practice, among the interviewed CLIL teachers, this teacher was an exception, which indicates that individual feedback that would pertain to issues other than general ones is not very common. The overall quality (person-referenced, task-referenced, explanatory or corrective, see 4.1.1) of the oral feedback remains unclear in the teachers' responses – an interesting issue for further study.

Written feedback

The second most common (52%) method of providing feedback to pupils was in written form in connection with test or self-assessment papers. The questionnaire provided proof of partly parallel methods of imparting assessment information to pupils with those methods used to inform parents on language issues in CLIL. The written feedback pupils get from their teacher is most often shown at home and signed by parents. The methods teachers brought up in relation to

providing feedback to parents were, for example, discussions, notes, Wilma¹⁵ messages (private and collective, also including test ratings), general feedback through the school website and comments on pupils' homework (see also Table 23).

One teacher stated that using the telephone is applicable especially when a pupil has problems in various areas of learning; another said that s/he writes Wilma entries in particular on pupils' lesson activity and its effect on language proficiency. The method of giving feedback and dealing with pedagogical communication is highly a matter of preferences, although nowadays Wilma is the official means of communication between school and home in most Finnish municipalities. The choice of feedback method probably also depends on the wishes of the given parents and the established practices of the given school. The teacher trade union is also influential in giving instructions to avoid any legal matters with parents.

Development discussions

Many primary schools in Finland make it imperative that teachers offer parents the opportunity for a development discussion at least once a year, which partly explains the reasonably high frequency of that method (45%). However, the percentage is low considering the prevalence of the practice. This probably indicates that **development discussions are held, but CLIL language issues are not notably brought up** in them. The underlying idea is that the parents and the child, if present, receive first-hand information on school issues from the teacher and have the possibility to express their concerns or raise school-related issues, as the teacher in the next quote mentions.

The parents are always asking in development discussions; they many times separately ask how the English study goes, for example, in mathematics.

Sometimes the discussion follows a predetermined pattern or is based on an outline or assessment documents. This reflects the pedagogical freedom of Finnish teachers to organise their teaching and work as they see fit. The two extracts below also exemplify how development discussions can be evidence-based which helps to keep the focus on learning.

When they were in the third grade, I had saved all their exams and tests, their own assessments, and they were bundled for the development discussions.

In our school it is accustomed to arrange these discussions with parents so that every teacher does it a minimum of once a year, I do it twice a year. Pupils get a self-assessment form to assess school issues and their language proficiency. But this is my system, not everyone uses such a form.

The problem, some teachers adduced, is that these discussions are often limited in time. On many occasions, these discussions are referred to as 'parents' quarters' which some teachers take literally and some dislike because time-wise much more time than 15 minutes is needed for

¹⁵ Wilma is a widely used electric application environment where teachers can, among many other things, enter data (e.g. notations and feedback) on pupils, which parents are able to read.

a proper, in-depth discussion. Additionally, there are also quite often other topics than language proficiency to discuss, as can be seen in the next extracts.

There has been communication on other things than language learning; on some pupils' part there have been quite many behavioural issues.

And this parents' quarter, i.e. 15 minutes, is in my case always normally 45 minutes. Just because of that [I don't have any assessment forms for the discussion]. But that's actually for the best; then I don't have to explain too much.

On basis of this study, it is reasonable to argue that **the potential of developmental discussions as a feedback method has not yet been fully discovered.**

School year reports

It is noteworthy that only 12 teachers (29%) marked the school year report as a means of giving feedback. This indicates that **it is not a widespread practice to include CLIL studying and language acquisition** as an integral component **in school report cards**. A few of the interviewed teachers reported that an additional document on English language proficiency is enclosed with the report cards in their schools. These schools had in common a relatively high ratio of TL instruction and a more detailed CLIL curriculum with language objectives. It is likely that CLIL-providing institutions with serious linguistic intentions and development trajectories in mind are more inclined to issue school reports or its attachments with CLIL language aspects. The extract from a teacher interview below displays such a practice, applying to every pupil from the 1st to the 6th grade.

At the stage of grades 3 and 4 the teacher writes freely formulated sentences. And the language skills are covered there. And then we give a separate report card on language skills which is more like a checking system: oral skills, written skills which are further divided into sub-skills.

An exceptional issue in relation to school reports was mentioned in the option 'other' means of assessment: one teacher specified the optional 'school year report': "separate language proficiency report card at the end of 2nd, 4th and 6th grade in cooperation with the native teacher and English subject teacher". This note of collaboration in assessment was not common in the data which reinforces the conception of disintegrated assessment in CLIL.

6.1.1.1 Frequency of provided feedback

This part of the report provides answers to the second part of the research question 1.2 "To what extent do CLIL class teachers assess language and provide feedback to pupils and their parents?" In order to avoid giving any answering models and to gain authentic descriptions of frequency rates, teachers were asked to describe the time spans for providing feedback to pupils in their own words. I divided the teachers (n=39) into four interval categories: those who give feedback 1) rarely (38%), e.g. once or twice a year in the form of developmental discussion or otherwise, 2) regularly but not very often (23%), e.g. once a month or when needed, 3) fairly or very often (31%), e.g. weekly or daily, and 4) miscellaneous answers (8%).

The gathered data suggests that teachers do not always mediate the observed or gathered assessment information forward to other stakeholders, pupils and parents. A significant result is that, according to the data, **62% of CLIL teachers give feedback to the learners either rarely or not very often**. Among these teachers were those who did not gather assessment information systematically, many of whom had either no language studies or only basic studies in English. The inference drawn from this is that **linguistic studies are a factor positively influencing how assessment in CLIL is administered and feedback forwarded to the stakeholders**. The feedback method they reported using most was oral feedback. Table 22 gives examples of representative descriptions of the feedback provision frequency.

TABLE 22. Frequency of teacher feedback to pupils on language proficiency or its progress

RARELY (n= 15)	REGULARLY BUT NOT OFTEN (n= 9)	FAIRLY OR VERY OFTEN (n= 12)	MISCELLANEOUS (n= 3)
<i>rarely, because in my opinion, the English subject teacher takes more care of assessment</i>	<i>Several times during the semester: always after a composition, performance or test. Also in connection with reading, listening comprehension is checked.</i>	<i>I give oral feedback on a daily basis. Quarterly, a written assessment form also including a self-assessment section.</i>	<i>It depends on the pupil and his/her level of development. When a weak pupil is involved, I give feedback more often and on a general basis. In the case of an advanced pupil, I give general feedback more rarely and concentrate on the elaboration of language proficiency and giving more detailed feedback.</i>
<i>When thinking of my first-graders, not at all. We're just in the vocabulary input stage, so actual testing is a restricted procedure only. In the 6th grade the tested language proficiency was graded with numbers and also partly by verbal representation, and I conducted the development discussions with pupils.</i>	<i>Mainly when returning tests or during the lessons when you can see that pupils understood the subject matter. In other words, regularly irregularly. In connection with returning tests, the feedback on language proficiency level or skills is on a general level, nobody's individual performance is under scrutiny.</i>	<i>Every week spelling tests, 2nd graders write 10 sentences of spelled words. Non-formal almost daily, approximately every second month another written test. On everyone's weekly turn of show-and-tell type of oral situation. Once a year an oral situation one at a time with the teacher during the break.</i>	<i>I give feedback on English language proficiency almost solely in connection with formal language instruction. In my opinion, the most important thing in CLIL instruction is the mastery of contents, not language. The language is only a tool which each pupil uses according to his/her skill level.</i>
<i>not often and not regularly</i>	<i>when we finish each section in a textbook</i>	<i>oral feedback on weekly basis, in written form related to returning tests and monthly briefings</i>	<i>It is not relevant to my teaching.</i>
<i>a couple of times per school year, not systematically</i>	<i>at least every fourth week</i>	<i>daily, weekly feedback in the classroom</i>	

Some of the quotes in Table 22 reveal a negative outlook on CLIL assessment and provide further support to the finding that assessment in CLIL is not nearly as embedded in instruction as it should be according to the NCC (2004). The comment “It is not relevant to my teaching” indicates the tendency to leave language assessment matters to the EFL teacher in a similar vein as the longer comment above it (column miscellaneous).

As to informing parents of CLIL-related assessments, nine teachers (21%) admitted not conveying any feedback to parents. Out of the 33 informants who did that, five (15%) kept parents regularly informed, whereas 55% transmitted information on irregular basis, and almost one third of them (30%) seldom, as Figure 22 demonstrates. This means that 85% of teachers do not convey information to parents on regular basis.

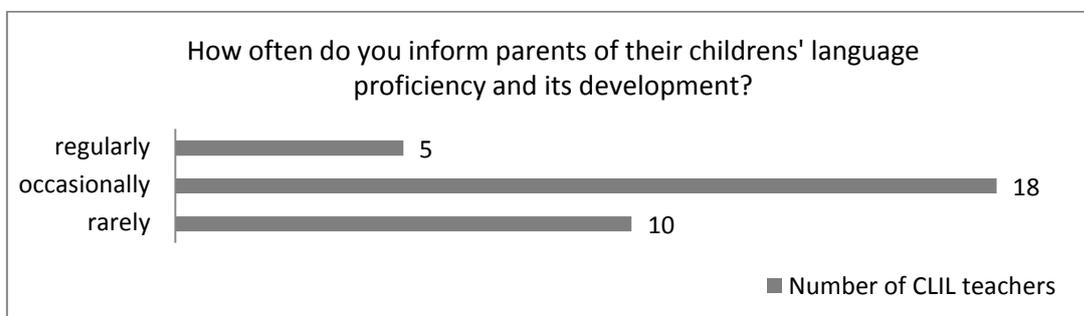


FIGURE 22. Frequency of teacher-provided assessment information

The free wordings in Table 23 exemplify how often, and how teachers take up CLIL language issues with parents. Some of the descriptions may refer to formal English (e.g. word tests), although they may also be subject-related.

TABLE 23. Teachers' descriptions of feedback on pupils' language proficiency or progress to parents

REGULARLY (n=5)	OCCASIONALLY (n=18)	RARELY (n=10)
<i>Once a year a development discussion; in connection with tests, projects and other productions self-assessment section and teacher's comment.</i>	<i>Every result on tests/exercises goes to an electric system which parents can check. At least once a month we have assessed something.</i>	<i>In connection with development discussion I mentioned to the parents of competent pupils that it is going well – 1st grade in question.</i>
<i>They see what I have written in children's production and homework what I have written when checking their notebooks and such, spelling tests every week. Then weekly bulletins through Wilma and reminders in case someone forgets homework etc. Sometimes in their 'bag book', but nowadays Wilma is so handy that almost all communication back and forth goes through it electronically and at least once a week. Additionally, I write general things via the school web site.</i>	<i>weekly word tests, larger tests after each study unit and term report cards twice a year</i>	<i>Always in developmental discussion, i.e. at least once a year. When applicable, I enter a mention of positive feedback into Wilma</i>
<i>Once a week, information on pupils' school work goes home in individual pupil diaries. English assessment is only a part of a larger weekly assessment.</i>	<i>Twice a year within formal report cards, but tests (also those of other subjects carried out in English) are always sent home and parents have the possibility to follow the development. I often ask parents to sign compositions etc.</i>	<i>Parents have the possibility to attend their child's development discussion twice a year. Additionally, we assess the language proficiency of the child very closely together with the parents before enrolling into the first grade.</i>

Teachers' own perceptions of regularity vary: one teacher's occasionally is another teacher's regularly and so forth. This is a source of minor discrepancy in the data. For example, the following quote is from a teacher who considered their practice to be as 'occasional feedback':

Profound development discussion once a term/year. Report cards twice a year with attachment of English language sub skills analysed at least once a year interlinked with tests, self-assessment forms and compositions approximately every second week/once a month (occasionally may occur more often if the closure of projects/tests/many essays concur). Information is verbal and/or based on numbers (now in the 5th grade, previously only verbal).

These differences in perceptions mirror teachers' individual professional theories in use which are based on, for instance, their background education, personality, experiences and beliefs.

Indirect and direct feedback

The word 'possibility' occurs in Table 23 and also in some teacher quotes implying that it is the parents' responsibility to decode or interpret how their child is managing in English in subject study and to which degree the language is developing. This approach to assessment requiring language expertise and activity from parents was also present in some teacher interviews. I will call this an implicit approach to assessment in CLIL. This phenomenon of favouring indirect over direct feedback may be linked to the sense of laboriousness of CLIL assessment.

I expect that parents surely monitor it [language proficiency] and make observations. If the parents are active and want to get information, they can open those notebooks [written in English] and investigate where we are going and how [proficient] the child is. I think that activity from the parents' part is important. I don't see that teachers' work load is added with reporting accountability on English language proficiency towards parents. I'd rather pass the ball to parents so that they can monitor. And if English has been used in tests, so there it comes as well.

We are trying to bring forth this assessment of language proficiency, but there's no going around the fact that parents see already in the [test] answers what's the level and what the situation is. [...] But my opinion is that if the parents would bother to read the notebooks and look what the children are doing, that would tell them what the level of language proficiency is.

I don't know how much they get [information] elsewhere than in the development discussions and seeing how successful the tests were. I have to admit that it is one thing I should do more, to inform parents, but I can't do everything.

The following teacher's account of 'remedy feedback' on language issues shows how parents are expected to participate into their children's study but how they are not always capable of noticing linguistic problems indicating that *direct feedback* as opposed to indirect is valuable.

Then I showed [pupils] what the problem was, saying that 'if you return these kinds of compositions to me, this language is at second graders' level'. Then I pointed out the problems and then we cried and bit the bullet. Then I gave the same kind of briefing to parents, that 'you haven't done your job'. The parents were very embarrassed that they had not noticed anything [...].

One teacher supposed that parents are not expecting to get feedback in CLIL.

I am thinking about the expectations of parents concerning this language development. I think that most of parents don't perceive what kind of feedback they get on those other things as very important, like how well the pupil masters the foreign language in a certain subject because I

think that parents are proud even of the fact that the children are in this kind instruction. It is valuable per se. It can be that when they are going to receive this kind of information more, it will be valued more.

These comments show that teachers are making assumptions on behalf of parents that may prove to be illusive and that they give credit to the linguistic expertise of parents.

According to the Basic Education Act (1988/628, 26§), the parent “shall see to it that compulsory schooling is completed”. No legal obligations regarding monitoring are placed on parents – moral obligations towards children are another issue.

SUMMARY

Language assessment in CLIL is not an established practice, and there is substantial variation in administering assessment in CLIL. Assessment of pupils’ language proficiency was less marked in the lower primary classes than in the upper classes but nevertheless, to varying degree, most teachers regarded assessment as important. A fifth of teachers (21%) did not gather systematic assessment information on pupils’ language proficiency or its development. This was regardless of the grade they were teaching. Some teachers also ignored the TL in assessment; the TL was used in instruction as a conveyor of content but not in the assessment of content knowledge.

CLIL class teachers favoured traditional-type assessment methods: teacher observation, tests or test sections, dialogic interaction and pupil self-assessment. The least commonly used methods were pupil peer assessment, simulations and language portfolios. Teachers have developed diverse practices in order to encourage and facilitate TL use in testing situations. They have created language use situations in which assessment could take place, and they appear to elicit both everyday-type language and subject-specific language. Collectively, teachers have adopted a wide variety of assessment methods from audio recordings to collaborative testing. Sometimes the use of English was seen as a bonus. The work of native teachers in CLIL was perceived as a valuable source of linguistic support and as a model for both pupils and teachers.

The majority of CLIL teachers neither imparted feedback to pupils nor to their parents regularly or often. The most common methods of giving feedback to pupils are informal, oral feedback on the observed or assessed linguistic matters and written notions in connection with tests, for example. Approximately a third of teachers reported giving feedback to pupils often, while 38% did that rarely. School year reports containing feedback on CLIL study or CLIL attachments were not frequently used methods to provide feedback. It appears that, when necessary, dialogue between home and school occurs in multiple ways. Teachers mediated linguistic assessment information to pupils’ parents irregularly and seldom; 85% of teachers conveyed feedback to parents occasionally or rarely. A fifth did not inform parents about linguistic issues at all. It was also discovered that some CLIL teachers expect parents to be self-reliant in interpreting pupils’ coping in English, level of language proficiency and its development. They relied on indirect instead of direct feedback.

6.2 Challenges in CLIL assessment

Teachers (n=33) were also asked to name problems and challenges hampering assessment in CLIL to investigate the research question 'What kinds of challenges are related to CLIL assessment?'. This section deals with those complications. Since there are a variety of approaches to CLIL, the problems encountered are not uniform, but vary by the type of CLIL provision. The identified problem areas, therefore, do not apply to every CLIL establishment.

I was able to categorise **a large number of recurrent factors complicating language assessment, or even hindering it**. I classified these factors into 11 problem areas: 1) lack of time, 2) lack of learning objectives and CLIL curriculum, 3) knowledge of curricular preconditions, 4) lack of assessment structure or criteria, 5) dialogue with regular English instruction, 6) diversity of pupils, 7) laboriousness of assessment, 8) assessment of oral skills, 9) lack of assessment tools, 10) diverse approaches to assessment and 11) distinct perceptions of the role of language in CLIL. Many of these areas of challenge overlap slightly which can be seen in the exemplifying quotes. The challenges are presented in random order, and no frequencies of the qualitative data were calculated.

Lack of time

Lack of time was one of the challenges teachers frequently brought into the foreground. Teachers named several issues contributing to the sense of rush and 'linguistic pressure.' They thought that there was not enough time for English instruction even in EFL. Some of the teachers represented the view that the language is learnt in EFL and then used in CLIL as the quotes show. The hectic school life is demanding in many ways, and teachers feel that other duties such as preparing materials and planning lessons take their toll on assessment which remains superficial and based on instinct.

There are not enough hours in the curriculum to sufficiently address the mistakes that CLIL students make. More time would be needed to practise the past tenses as these often cause problems. If one spends time addressing the linguistic issues, this is away from other things. Two hours of language a week is also insufficient to address the language issues adequately and to cover the curriculum.

It is pretty impossible that versatile language could emerge within 2 hours a week. Although for example history, science and arts are taught in English, the 2 hours is totally insufficient to support the rest of instruction. Language should also be mastered well – and one should also teach it! Merely studying in English is not enough.

The latter quote makes a strong point for focus on form instruction. This is an indication of dissenting views in the field in respect of implicit/explicit language teaching and CLIL. The administration of oral assessment is also considered as challenging to organise time-wise; teachers tend to keep the time tables and schedules, which is another issue to note. Rearranging instruction and using differentiating work methods might be a solution for this dilemma.

Lack of learning objectives and CLIL curriculum

When preparing and conducting this research, it became obvious that some CLIL-providing schools do not have any kind of CLIL curriculum, and the spectrum of existing CLIL curricula is wide. Many of these schools do meet the curricular requirements set by the NCC (2004). These profound deficits were confirmed by the survey and interviews. The frustration of teachers representing several schools came across markedly, as is notable in the quotes below.

It is challenging that there are no common criteria in setting the objectives. The teacher sets the objectives him/herself – comparison with the level of formal English pupils is hardly helpful.

When I don't know the objectives quite clearly myself either, then also the assessment is alike; sometimes I think to myself, what's the point in this. It is obligatory to assess; we should together talk it over and figure out what we are going to require of them. Because now we get comments from parents like "history is too difficult" and "that should be covered better" and so on. It would clarify the situation when we could show that 'these are the minimum objectives, these should be covered. Then, assessment would be much easier. Now it is really hard in my opinion.

In our curriculum, we just have sort of frames, what CLIL is, but we haven't separately and explicitly defined what should be taught in English in different grades. It is a decision for teachers to make. They consider class-specifically which project or theme includes English. So that varies year in, year out.

When there are no common objectives pertaining to language, or no contents or syllabi specifying what should be taught in English, also no common assessment criteria exist. This is an unbearable situation which also leads, as can be seen in the second quote, to misunderstandings and outbursts from the 'customers', pupils and parents, who are entitled to uniform and well-grounded CLIL instruction and assessment. This is also a question of equity which is a guiding value in Finnish basic education.

Knowledge of curricular preconditions

It also became evident that some teachers, in concentrating on their daily practical duties, have become estranged from curricular preconditions, or they are not familiar with the curriculum at all. The NCC and its prerequisites of CLIL instruction and assessment are not always known to teachers. This issue is sensitive, and it is not easy to admit. As a result, it is not easy to obtain explicit information on teachers' ignorance of the NCC or the local curriculum. The first quote provides evidence of this issue by implication while the second is more overt.

R: Are you aware of what the NCC says about assessment in CLIL?

T: Well, I haven't looked at it. For a while.

It would be good to cater for teachers with various possibilities to assess in individual ways; to give freedom if the curriculum says indeed that we should report. I had totally forgotten it although I have read it. I have taken liberties as a teacher. Or I have rather thought that parents should be active in this respect.

This suggests that teachers are not using the curriculum as the basic tool guiding their instructional decisions and solutions, but it is rather their own beliefs and preferences that drive

their teaching. Previous research (e.g. Luukka, Pöyhönen, Huhta, Taalas, Tarnanen and Keränen 2008) has provided evidence on such practices, including CLIL environments (Hüttner, Dalton-Puffer & Smit 2013).

Lack of assessment criteria or structure

The absence of a common structure and cohesiveness in assessment practices were issues that concerned some teachers in the data. It is likely that the lack of CLIL curriculum is mirrored in ambiguous, teacher-driven assessment criteria and non-existent common assessment structure or scheme. Teachers, as in the following passages, expressed the need for common ground rules in assessment:

It, [CLIL assessment], is an issue that we haven't had the energy to tackle so far. We have just developed CLIL; we have not had the time and energy; we have not been able to go into it.

Alignment of commonly negotiated assessment practices within the whole school would be extremely important. Should CLIL classes have report cards and assessment deviating from other classes?

And what's the system, then? Should we [assess] somehow systematically? That would certainly be one of those things that us teachers should talk over so that we all would do it in similar manner.

The first quote indicates that the establishment of CLIL as a system has been a venture that has taken its toll. Assessment issues have not been considered important enough to be incorporated and stabilised in the initial phases which is suggestive of pedagogical enthusiasm instead of thorough deliberation. It is not uncommon to start approaching education from the assessment perspective, thus acknowledging its profound necessity (Cf. Llinares, Morton & Whittaker 2012, 280).

Dialogue with regular EFL

The data disclosed that in many, but not all CLIL-providing schools, the CLIL class teacher teaches regular English to his/her own group. This practice increases the possibilities to bind BICS and CALP-type English together; thus, the teacher can form a more comprehensive conception of children's language proficiency and scaffold the development of subject-specific language considerably better. In contrasting cases, as was already noted in this chapter, the two lines of English tuition are unattached strands that have little, if anything in common. The first comment announces the benefits of collaboration; it is a matter of synergies and resembles the practice of collaborative teaching in Italian CLIL, for instance.

Cooperation with the language subject teachers is also a key factor here. When you know where they're at in language studying, you can take same things into consideration in subject teaching and vice versa.

I consider such practices as terribly good in which one of the [English] lessons is given to the subject teacher for normal language structure instruction.

The above comment describes a division of practice encountered while visiting other CLIL schools in Finland. This practice combines focus on form and focus on forms, but whether collaboration between the two teachers occurs is not clear. If the two teachers were to co-operate and align their teaching to support each other, the linguistic potential would be enhanced. Another aspect is that some schools emphasise English in CLIL instruction; instead of traditional two weekly lessons, CLIL classes have three, which also gives more room for differentiation (see e.g. Table 17). Additionally, as observed while conducting this research, there are teachers in the field teaching both CLIL and EFL simultaneously without any kind of linguistic studies.

Diversity of pupils

On the basis of the evidence obtained through this research, it seems fair to state that despite admission procedures, the pupils in CLIL classrooms are diverse. Pupils differ in their background, language skills and motivation, for example. Many of the CLIL schools admit pupils through tests in which linguistic skills (Finnish, English or both) are measured. It was also found that some schools accept every pupil in order to gain in numbers. The phenomenon called ‘school shopping’, i.e. selecting schools according to their supply, also appears to have arrived in Finland. One school admitted pupils to their CLIL classes through a lottery conducted amongst the willing study-place seekers which naturally adds to the heterogeneity of the pupils and poses challenges in assessment, as the following teacher quote portrays.

Tests differentiating enough in various levels do not exist yet (the level differences in the classroom may be vast).

The parents’ background questions regarding the motivation to gravitate towards CLIL revealed that language was the main incentive to apply for bilingual instruction for 88% of parents (see p. 113). Cultural issues were not a decisive factor; only 6% of parents claimed that introduction to Anglo-American culture was a motive for CLIL study. However, 14% of parents disclosed that the reasons behind applying to CLIL study were either solely or additionally non-linguistic: ensuring access to the nearest primary school, sibling(s) in the same school, provision of more intellectually challenging tuition and opportunity to go to school with the selected, “better” student material (quotation marks also in the original parental comment). This means that, hypothetically speaking, on average in every CLIL class there are two children whose prime motivator (or that of their parents) is not second language acquisition.

Laboriousness of assessment

It was a clear finding that assessment in CLIL is seen as arduous by teachers. The exigency of CLIL assessment becomes evident in these excerpts:

The documentation of the whole variety of linguistic competences is laborious and takes time: reading skills (speed, fluency and accuracy), writing skills, oral performance skills and mastery of structures. This all should be covered and one should also be able to assess them in a trustworthy and valid way.

This is not nice to say, but the teacher's personality is a factor that affects what you want to assess, how much work you want to put in, how much you'd like to monitor what's happening and how the child develops. I think that if you want to get off easily, you don't assess. It is always laborious for the teacher.

As can be noticed in the above quotes, the form of given feedback (written, oral) and form of language tested (written, oral) has an influence on the perceived arduousness of assessment. Assessment of oral proficiency was perceived as particularly challenging in terms of, for instance, time and situational organisation. Furthermore, in the absence of a CLIL curriculum, teacher personality and professional integrity are factors that have implications for actual assessment. Another factor influencing the perceived laboriousness of language assessment may be the lack of linguistic background knowledge about language learning and assessment – not all teachers had language education background or the level of it was not as high as stipulated.

Assessment of oral skills

The challenge of assessing oral skills seems to be partly related to a lack of time, classroom organisation and the fact that CLIL class teachers do not always teach regular English to their group which makes it difficult to organise amidst content teaching. The purpose of assessment of oral skills is not to detect mistakes but to notice progress in language development.

Assessing oral skills of a shy and sensitive pupil is very difficult.

If and when the linguistic skills are assessed in the first grade, only oral skills come into question, for most of the pupils cannot read or write when entering school. Arrangement of school hours is characteristic to the difficulty of assessing oral skills.

Giving feedback on pronunciation was not seen as relevant by some CLIL teachers; positive feedback is given when learning takes place. There is an accepting approach to pronunciation.

Lack of assessment tools

Uncertainty over language assessment methods appears to surround CLIL. However, teachers did not indicate explicitly, other than what is already stated in the previous pages and in section 6.4, what kind of assessment tools they would need. Collected from the whole sample of teachers, this study submits several innovative assessment methods and practices that may be an aid to diversifying or specifying assessment practices.

In tests, there are tasks in English, but they give evidence of only part of the skills. Measurement should be easy to administer. Could it be done by means of a web questionnaire?

In my opinion, one can manage with present assessment tools pretty well, but there's always room for improvement; for example, new possibilities arise with the help of new technologies (iPod, iPad).

As seen above, this uncertainty does not apply to everyone. Incorporating technology in learning and assessment is a megatrend, and Norrena (2013) calls teachers with personal intention, technology-friendly attitudes and future-orientation 'agents of change'. This teacher was an

exceptional example of an agent of change who had already experimented with various technologies in the CLIL classroom for assessment and learning purposes.

Diverse approaches to assessment and grading of language in CLIL

According to this study, the most overwhelming problem in CLIL assessment is the ambiguity of it which is displayed in diverse approaches to assessment. Teachers' stances towards assessment vary considerably as can be perceived in the passages below. The lack of proper foundation for assessment poses a severe threat to the success of the CLIL programme as a whole; the requirement of assessment stated in the NCC is not familiar and the pedagogical freedom justifies deviating actions.

Agonising! There are no clear [instructions] how to assess [the language]. We don't actually have any objectives either. I don't know what, in which subject or how much. But I can't know that when I don't know what I'm assessing. And the assessment varies; we have talked with each other. It is like night and day.

I would say that it [assessment in CLIL] varies tremendously from teacher to teacher. We have such a great freedom in acting as a teacher, also in assessment. There is nothing, I can't say that in certain schools in certain ways, but also in these CLIL issues the teacher has to have the freedom [to assess] according to how s/he perceives it, what kind of an understanding s/he has of learning and assessment and so forth. It [assessment in CLIL] is based on that.

I have to say that when I answered the questionnaire, it occurred to me for the first time that [the language] could also be assessed, the actual CLIL performance within every subject. [...] In that sense, I have not assessed the language in different subjects. The language use has been assessed in general in the formal English grade.

In the field, there are also dissenting views about in which connection the English language should be graded – if it is graded either in connection with the subject grade or EFL grade – if assessed at all at the end of the school year or otherwise. The majority of teachers in this sample included English in CLIL in the regular (formal) EFL grade, while a few teachers included it in the subject grade. The latter practice was also mentioned by a few pupils in the interviews. The quotes below make the diversity in practices as well as the confusion over grading clear.

I think that the main emphasis in assessing the language is on formal instruction – in the CLIL classroom the teacher sees to the learning of substances.

Where should it be seen? It is the report the parents are looking at. The general curriculum and CLIL curriculum; if the child meets the prerequisites of the general curriculum for the grade eight, say, in history, but let's take another who can't manage the English at all, what is the grade then? Where does the CLIL manifest itself? What is fair? [...] What is reasonable? Is it an independent grade or is it within the subject grades? It is certain that some emphasise it more, some see it just as a plus within the grade and some are punishing with it. What is right? The practice should be mutual, no doubt.

T: *Then the CLIL English, it is assessed within the subject.*

R: *So it is included in the subject grade?*

T: *In the grade, yes, because in our school it is a tool we are using, not a target. As to the school reports, [pupils] only get a tick that they have attended a CLIL class.*

In our school, it [the language proficiency] is assessed in the English grade. It is the English grade

given by the English teacher. The class teacher gives feedback, but isn't actually assessing, except for the report in the 2nd, 4th and 6th grade. Then, that's it.

It was not only the dissimilar approaches to assessment that cloud assessment in CLIL but also distinct perceptions of the role of language in CLIL, which is another decisive factor in approaching the CLIL assessment.

Distinct perceptions of the role of language in CLIL

Based on the evidence presented above, English language assessment in CLIL seems to be a controversial and ambiguous issue. An explanatory variable for this might be that, according to this study, teachers have dissenting, perhaps even unconscious views on the role of language in CLIL. I identified and classified three diverse approaches to the role of language in CLIL teaching that I did not look for but which materialised in the data. The views became essential because it seems that they are connected to how language assessment is perceived in CLIL.

Some CLIL teachers stress the role of language as an instrument or a tool, while other teachers represent a discipline which sees language as an equal aim of instruction besides content. A few teachers appear to see English as a supplementary benefit, or they even dissociate language and content. In other words, some teachers teach content through language (instrumental focus), some content and language simultaneously (dual focus) and some, although perceiving language as an essential part of CLIL teaching, let another aspect (e.g. affective factors) displace the importance of language proficiency as the target of assessment which is rather observation (eclectic focus). The effect of these diverse foci can be discovered in the following characterisations.

Instrumental focus

In an instrumental focus, language acquisition is a side product of content instruction, and the language acquisition is incidental and implicit, which justifies ignoring assessment of or monitoring language proficiency in CLIL subjects. Furthermore, English in CLIL is perceived as similar to the language taught in regular, formal English.

In my view, it is not necessary to assess language skills in, for example, mathematics, environmental sciences etc. Assessment of language proficiency takes place in formal [English] instruction. I don't want that learning school subjects transforms into pure language study. Language is only a means of studying, and every pupil uses it according to his/her own proficiency level.

If objectives and criteria for assessment were established, it would undoubtedly be fairer and more systematic. At this moment, I ponder which is more important: mastery of content or mastery of content in a foreign language. Understanding the content matter is always the number one thing for me; language is a bonus which is learnt through the content.

Dual focus

In a dual focus, language is intrinsic to classroom functions. Both language and content learning are supported, and the role of language in instruction and assessment is carefully premeditated. English in CLIL is perceived as different from English taught in language lessons, and their

complementary role is accepted. Assessment of language proficiency also occurs in CLIL and emphasis is placed on building pupils' language awareness and self-assessment practices.

In other words, I may stop in the middle of a lesson and just remark that now we have used this kind of expression or something else. Sometimes, in subject lessons, that may generate little streams so that they form a language flash.

In my opinion, however, the primary and secondary CLIL schools in the same area or city should be obligated to co-operate in assessment since they have parallel plans and methods anyhow. [...] At the same time, when common policies in assessing the language development (across the board) should be tried to be found, even more attention should be paid in pupils' self-assessment and longitudinal progress. Pupils' own reflection is central (portfolio-type work).

This approach is related to focus on form (see 2.2) which acknowledges the explicit teaching of the language as an essential means of linguistic competence-building.

Eclectic focus

These two examples are both from teachers of grade 1–2 which explains the emphasis on oral language skills. Assessment methods naturally need to be adapted to the target group, as these informants point out, but both of them mention affective rather than linguistic factors as subject to monitoring (motivation, courage, interest, enthusiasm and the performance of the class as a group).

Assessment of CLIL pupil's motivation is, in my view, more important than assessing progress in language proficiency. In a heterogenic CLIL class, there are all kinds of learners, and mastery of the foreign language is already versatile in the beginning of the first grade. Also, assessment tools should be very individualistic and they should point out very clearly the progress of the child compared to their own skills, not to their class mates. [...] Personally, I would like emphasis to be put on brave use of even elementary language. Language is nonetheless used only as a tool for learning in which case it is not, in my view, even meaningful to assess the level of language proficiency in the first place, not at least in the elementary instruction. It would be more important to assess whether the child has the courage to use even that little amount of language that s/he knows.

It is more essential to awaken interest and keep up enthusiasm. That is something you can assess with gut feeling when you've had the same group a long time.

The first passage also raises the point of one focal aspect in Finnish educational system: it is not comparative but rests on highlighting individual progress (criterion-referenced assessment of individuals).

SUMMARY

Eleven various, partly overlapping challenges identified in this research impede assessment in CLIL and frustrate CLIL teachers. The most substantial is the lack of proper referential foundation which is embodied in the lack of CLIL curriculum, i.e. learning objectives and criteria. Additionally, teachers do not always appear to be aware of the CLIL assessment prerequisites in the National Core Curriculum for Basic Education. Furthermore, lack of time, assessment tools, undefined roles of English (formal and CLIL) and especially teachers' diverse approaches to assessment

pose a threat to adequate assessment. There are varying practices regarding the grading of English in CLIL.

Moreover, the lack of a mutual assessment scheme increases diversity and causes perplexity among CLIL teachers. Assessment was seen as arduous. Three varying teacher perceptions of the role of language in CLIL were identified: instrumental, dual and eclectic. Some see language as a vehicle taking an instrumental focus, while others adhere to the dual focus (language as a vehicle and a target). An eclectic focus to language was present in the data as well insinuating that affective and motivational factors are prioritised instead of the development of language. These perceptions appear to have an influence on how teachers perceive language assessment in CLIL; teachers with instrumental and eclectic perceptions do not see language assessment as relevant in CLIL.

6.3 Adequacy of CLIL assessment

Adequacy is the quality of being good enough for a certain purpose. In section 6.1, I categorised various assessment methods and practices CLIL teachers employed for gathering information of pupils' English language proficiency and its development. The adequacy of those practices and collected information is largely dependent on how sufficient or informative this feedback is perceived to be by the recipients and how frequently it is given. This section is concerned with the adequacy of assessment in CLIL by approaching it from the angles of pupils and parents, thus providing answers to research question 2 *'How adequate is assessment in CLIL ?'* and 2.1. *'Do pupils and their parents receive information on pupils' English language proficiency and its development in CLIL contexts frequently and sufficiently enough?'*. The section first describes received feedback (6.3.1), then the frequency of feedback (6.3.2) and finally the sufficiency of feedback (6.3.3) as perceived by pupils and parents. Generally, the section points out that language assessment and feedback in CLIL is not adequate from the viewpoint of pupils and parents. This finding is reinforced in the section pertaining to future wishes (6.4).

6.3.1 Received feedback

This sub-section looks into how teachers' feedback and assessment information is received and perceived by the two other groups of assessment stakeholders, pupils and parents. First, I will investigate the importance of feedback as perceived by pupils and then continue to the means of receiving feedback as experienced by pupils and parents. The gathered data is mostly quantitative, and the stakeholder quotes are extracted mainly from interviews.

Importance of feedback

Pupils (n=108) on average considered feedback on their language acquisition in CLIL important. **Nearly half of pupils (49%) endorse that obtaining information on their language skills is of high importance to them.** Although closer scrutiny was not guided by the research questions, Figure 23 illustrates that 3rd and 4th graders perceived feedback as more important than 5th graders. In other words, **the younger the pupils are, the more important the linguistic feedback.** This result implies that pupils' age is a dependent variable in how significant feedback or assessment information is perceived to be, for a decrease in perceived importance occurs in the transition point between the 4th and 5th grade when the study of English as a formal school subject and the use of English in CLIL contexts has become more established.

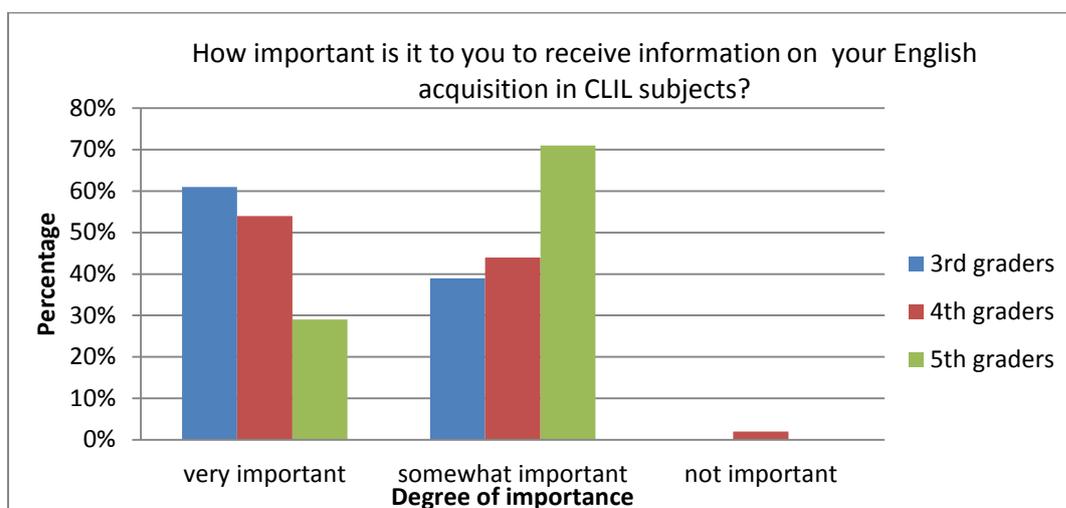


FIGURE 23. Importance of receiving information on English acquisition in CLIL subjects

In Finland, formal English (EFL) starts in the 3rd grade leading to more extensive language exposure and focus on forms; the subject-specific study should also gradually become a practice during grades 3–4. The language self-concept may have stabilised by grade 5. It is also noteworthy that practically none of the pupils (with the exception of one pupil) perceived information on language acquisition as insignificant. When examined according to gender (Figure 24), the results are even more striking, but also somewhat expected. **It is far more important for girls to receive feedback and information on their learning through English than it is for boys.** Approximately 60% of girls and 30% of boys value such feedback.

Forms of feedback

The study also asked about the means of receiving feedback from the teacher experienced by pupils (n=107). All grades combined, **the most common means of feedback was tests**, as can be seen in Figure 25. Whether pupils received separate language assessments in connection with tests or had to gauge their proficiency level alone based on their coping in the test was not defined. This finding is in contradiction with what teachers stated (Figure 21, oral feedback most common) and suggests that either pupils do not recognise the oral feedback they receive from

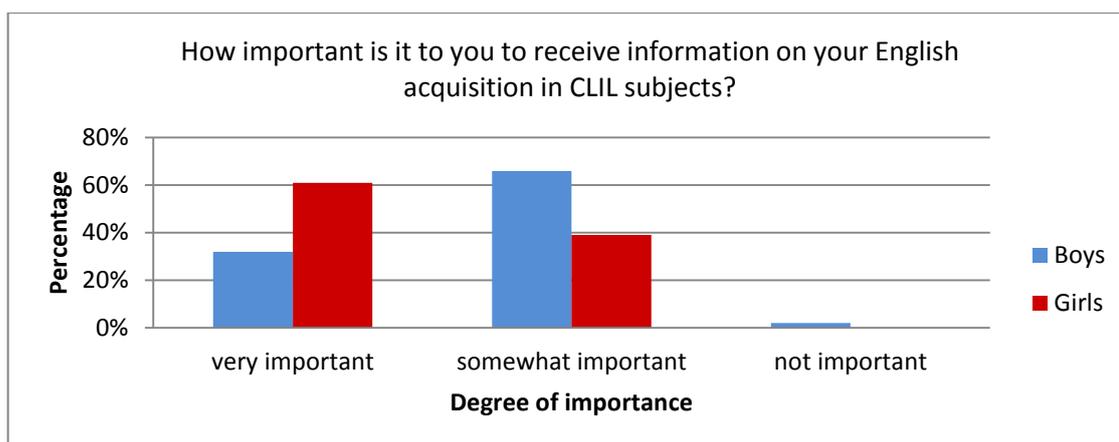


FIGURE 24. Importance of receiving information on English acquisition in CLIL subjects according to gender

their teachers, or their teachers do not give feedback or it is not explicit enough. Similar conclusions have been reached by Havnes and colleagues (2012). However, it is important to keep in mind that all the teachers and pupils participating in this study are not directly in pedagogical relationship with each other.

The area in which pupils would often like to receive more feedback is pronunciation.

One pupil (girl, 5th grade) mentioned in the interview that she would like to get reinforcement on things that are going well and how to improve in English. This extract also makes clear that some parents are willing to participate in the CLIL studies of their children but are not always experts of the TL (Cf. indirect feedback in 6.1.1.1). This pupil wishes to receive more developmental feedback:

I'd terribly much would like to know if I pronounce certain things right and if I write them correctly. Do I have a lot to improve or is it [the language] going well? Because with my mom, we wonder if this [test response] was quite right and how did the test go. And then we look at the test, and how it is, but it is never stated [what could be done]. It only says what is wrong.

The second most common form of feedback identified by pupils was coping with school assignments. This probably is a self-monitored means of feedback also representing indirect feedback, for teachers typically do not control every homework or school assignment personally. The pupils grasp their own potential or lack of it through how easy or difficult the assignments feel. **School report cards and teacher's oral feedback** (both 52%) **were the third most common forms of feedback**; just over half of the pupils claimed that they had obtained information regarding their coping in English in CLIL subjects from school reports. This is peculiar because neither of the participating schools provides any other information on CLIL studying in the school report than the sentence 'pupil has attended instruction in English foreign language' which does not make any judgment as to the pupil's language abilities, skills, language development, motivation or effort. It is possible that those pupils have thought about the grade in formal English or the subject grade.

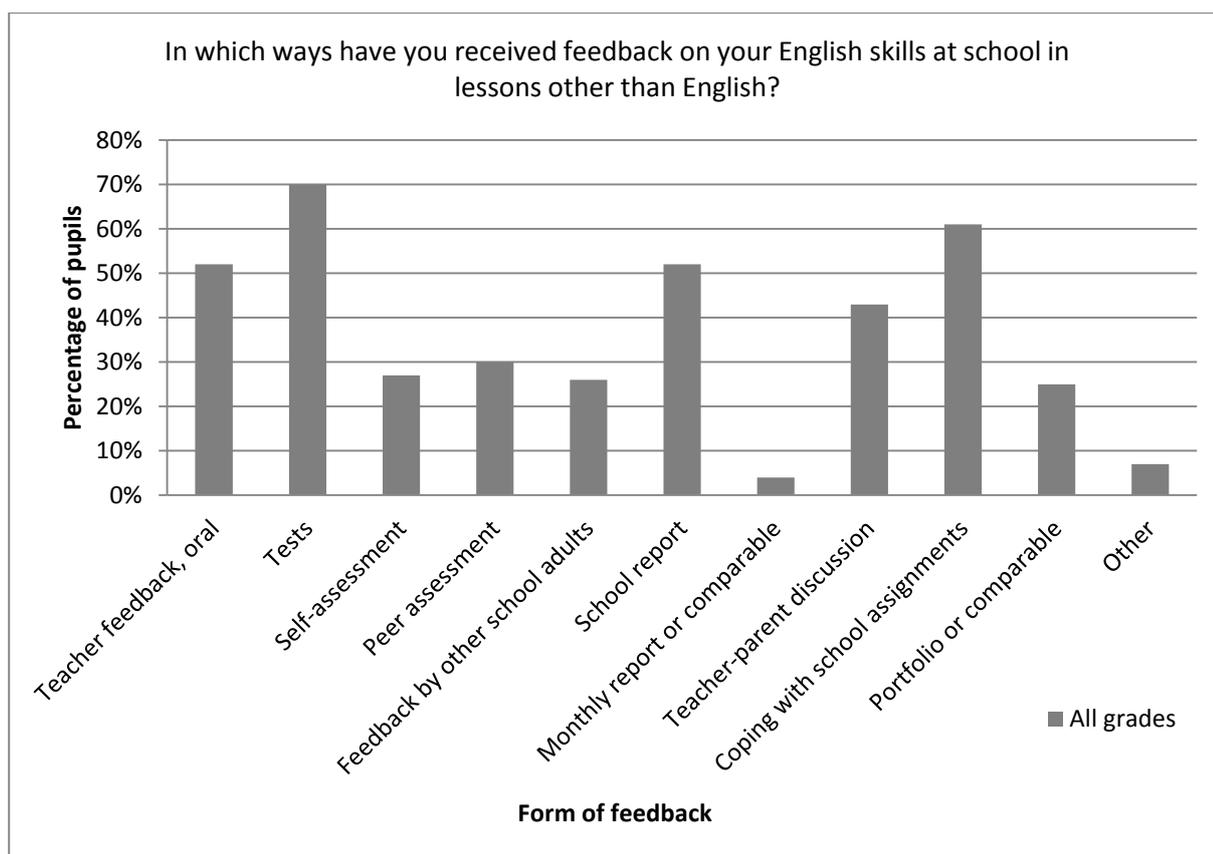


FIGURE 25. Forms of feedback identified by pupils

The pupils named the **teacher–parent development discussions the fourth commonest form of feedback** (43%). It is interesting to note that pupils' self-assessment was not acknowledged as a very frequent means of perceived feedback; only 27% of pupils stated receiving feedback via self-reflection. A portfolio was a fairly frequent method of feedback (25%) since both 3rd grade classes were engaged in some type of language portfolio work.

When these results were examined per individual grades, there was no specific difference except for teacher's oral feedback and portfolio work. **The younger pupils reported receiving oral feedback from their teachers more often than older ones.** Since 3rd graders are beginners in more systematic EFL instruction and contents are being taught more consistently in English, the teacher may be inclined to consolidate pupils in order to make them feel more reassured, whereas 5th graders are considered to be more established language learners and users.

Parents

Parents (n=87) also answered a question about the means of obtaining information on their child's language proficiency and its improvement (Figure 26). The results were parallel with the ones drawn from pupils. **Both parents and pupils perceived tests and coping with school assignments as the two key methods of receiving information;** parent percentages being as high as 81% and 79% in comparison to pupils' 70% and 61%, respectively. Parents apparently

have more experience of education and are therefore more capable of interpreting tests and school work in general.

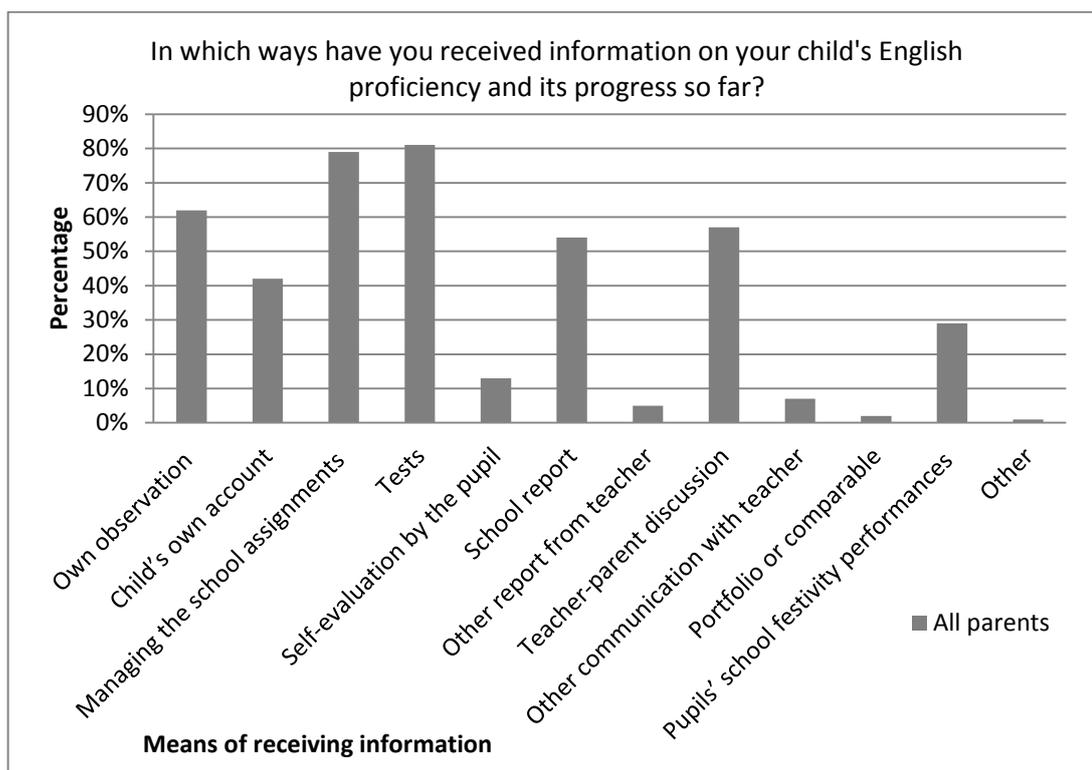


FIGURE 26. Means of receiving information on child's English proficiency and its progress

Teachers' oral feedback was along with the school report the third most significant means of feedback for pupils, whereas parents rely on their own observation (62%) of their child's behaviour and actions. **Teacher–parent discussion (57%) and school reports (54%) follow own observation as sources of feedback. The child is also seen as a significant informant.** Apart from teacher–parent discussion, the most frequent methods are all self-acting, i.e. requiring parent's activity and willingness to monitor the child's progression, as some of the following examples from parents' interviews depicting feedback (either implicit or explicit) show. The children's language skills also come across in extramural, non-content-related situations.

In my opinion, the child is mathematically very skilled. S/he has managed very well in it, but that assisting! Mathematically s/he understands, but the language is a challenge, yes.

The oldest and middle child [in CLIL instruction] received monthly [briefings] of what kind of vocabulary they have had, which songs have been sung. It was in the 1st and 2nd grade similarly as with the youngest one [...]. They had no book or there was some kind of an English notebook. Sometimes parents were asked to sign. But then, of course, in the 3rd and 4th grade and formal English study they start to use books more, so then become tests; a fairly small amount of self-assessments. There has not been any practice of any individualised or precise assessment of English proficiency such as oral skills or grammar.

The feedback in parents' evening is rather general. My sense of missing something must come from the fact that I have just one child, so I can't compare. I am somehow confused, I don't know. Maybe I'd like to have some more [feedback], but I know that I could be more active myself.

Pupils' self-assessments do not seem to be of high informational value to parents (13%), although approximately 40% of teachers claim to make use of self-assessment, and every fourth pupil states gaining information on their language skills in CLIL through self-reflection.

In my opinion, s/he doesn't trust [the language proficiency] and that is clearly visible in the self-assessment they had last month. In relation to an English test we received a self-assessment in which s/he had estimated all skills lower than the teacher had. But then again, I thought, maybe the assessments of the teacher were a bit encouraging because I have monitored it closely how much work s/he has to do in order to cope in tests. I'm not quite sure if s/he is that good in [English], but perhaps mummy is expecting too much.

Self-assessment by my child; s/he is comparing him/herself with others anyway. That's what they do. S/he has a very clear concept of other pupils in the class; those who speak well, who master [the language] well, then those who are at the same level with him/her and then, well, s/he never says that someone is bad at [English], s/he doesn't do that. Apparently everyone masters one area, something well. But they are ranking themselves.

This finding suggests that the true value of self-assessment lies in the reflection and self-examination it generates. It is possibly more suitable for building a child's language self-concept and language awareness unless the pupils are trained to assess specific language skills. Whether or not that is relevant depends on the group, type of CLIL provision and skills of the teacher. The first quote also shows that the assessment of English in both CLIL and EFL are tightly intertwined in the minds of parents. More than pupils' self-assessments, parents seem to appreciate school performances (Halloween and Christmas parties, days of open doors etc.) as an information source. Several parents commented on pupils' school festivity performances.

Small drama plays, performances, songs etc. are important. Children must = are allowed to use the language in a concrete way.

My son must have been in the 5th grade [...]. Then I was amazed how easily they used the language [in mathematics]. Maybe such [days of open doors] could be arranged. During my daughter's studying there haven't been such demonstration lessons for parents so that parents could come and see how [CLIL] is. [...] I was very impressed how well they could use the language.

These school festivities, there I have enjoyed children's language proficiency and seen it in action. It is always amazing that they are able to learn long lines by heart and it is surely useful.

In the above quotes it becomes evident that **parents appreciate concrete demonstrations of language proficiency and direct contact with school life**. It is possible that teachers' linguistic insecurities prevent them from opening doors to parents. Instruction in Finland is public by law. Therefore, parents have the right to observe instruction without specific invitation if they wished to do so, but this right is hardly used.

6.3.3 Frequency and sufficiency of received feedback

This sub-section specifies the frequency of assessment information from the viewpoint of pupils and the sufficiency from the perspective of parents. Pupils were asked how often they receive feedback at school from their teachers, and parents were asked whether or not the feedback

obtained from school is sufficient for them. **The findings show that pupils do not perceive the feedback received from their teachers as frequent.** As Figure 27 depicts, only 8% of all participated pupils (n=109) actually feel they are getting frequent feedback on their language proficiency.

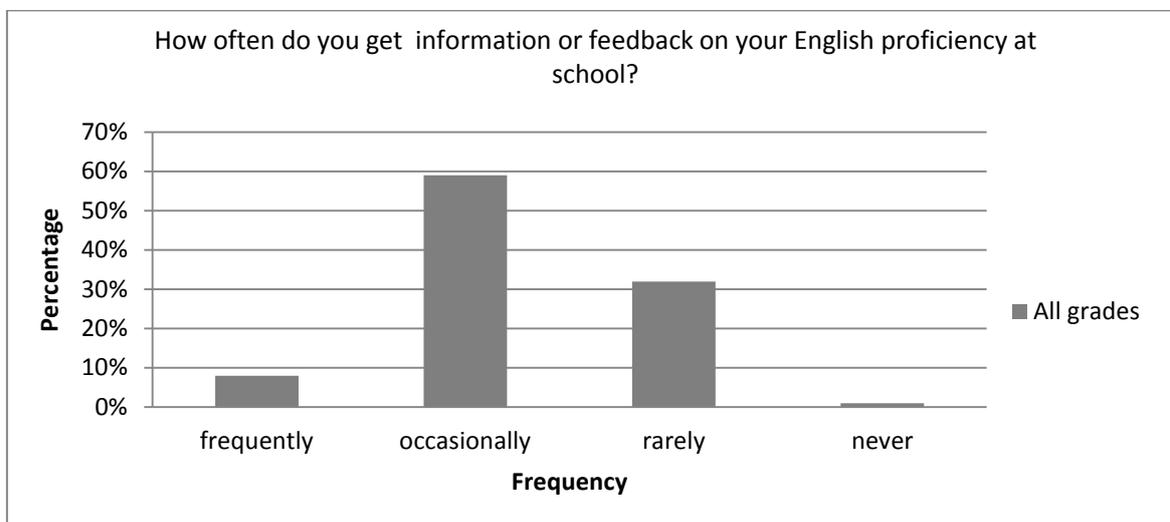


FIGURE 27. Frequency of feedback on English proficiency experienced by pupils

Most pupils stated receiving feedback occasionally, while a third signalled getting feedback rarely. This is in line with the teachers' statements; 61% of teachers declared that they give feedback only rarely or not very often, although at regular intervals. The excerpts from various pupils' interviews also consolidate this finding and transmit a sense of melancholy.

R: Do you get feedback on your English proficiency from your teacher?

L: *Well, I don't get terribly much any kind of feedback from the teacher. (boy, 5th grade)*

R: I'm not quite sure if I already asked this, but do you get enough information on your language proficiency at the moment?

L: *Well, maybe a little too little, I'd probably like to get a bit more. Usually it's school reports and self-assessment and that's it. You don't get that much from that. (boy, 4th grade)*

R: Do you get information of the development of you language proficiency now that you have studied almost four years in a CLIL class?

L: *At the moment, principally, I don't get any from the teacher.*

R: Where from do you get [feedback]?

L: *Mostly I notice myself that I have learned things, so from there [...]. (boy, 4th grade)*

R: Do you get positive feedback?

L: *Not much. (girl, 3rd grade)*

This research result, the pupils' notion of not receiving regular feedback of their language skills – especially if and when language is a vehicle of learning in CLIL – is of paramount importance. Studying in a bilingual class is their own choice for many pupils, as announced by 46% of parents when asked about the gravitation towards CLIL. Therefore, it is also reasonable to assume that

learning language and progressing in it is important for children – as is receiving feedback (Cf. Figure 23). The fourth grader's remark on how insignificant grades are in inability to communicate the diversity of skills has been also noted by the CLIL research community (Zydatiss in Dalton-Puffer 2008, 143).

The investigation into the sufficiency of feedback revealed that parents' perceptions of receiving sufficient assessment information or feedback are polarised. Parents' opinions (n=97), portrayed proportionally in Figure 28, were almost divided in half.

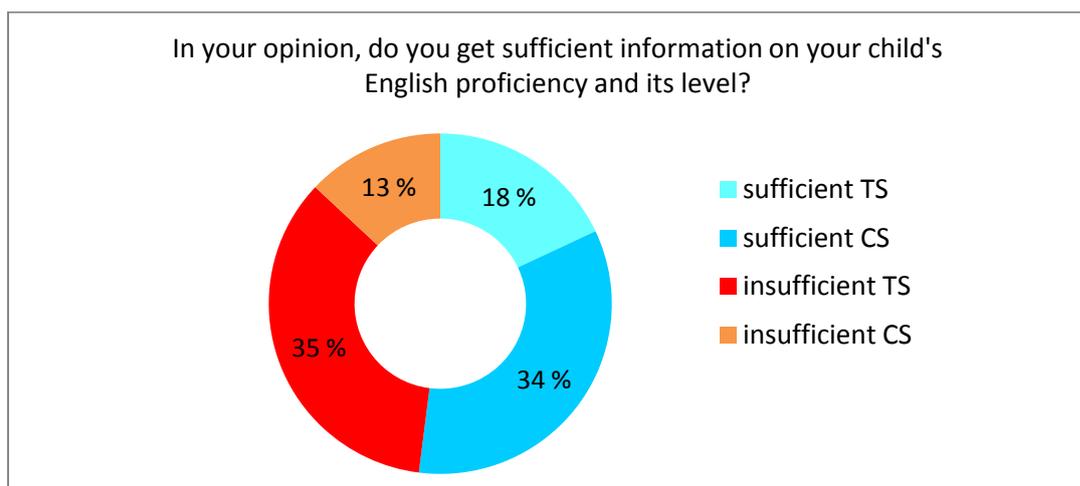


FIGURE 28. Sufficiency of information on child's English proficiency and its level in parents' opinion

A slight majority of parents (52%) were content with the quantity of feedback pertaining to the level of language proficiency. Moreover, although the parents of the TS were considerably dissatisfied with the obtained assessment information, the parents of the CS had the opposite perception. There may be multiple reasons for this, one being significant differences in the school profiles (see Table 17). The most marked differences are the ethnography of pupils, qualification of teachers and years of CLIL provision.

A couple of parents added a comment in the questionnaire paper regarding the sufficiency of assessment information, both for and against, although negative remarks outnumbered positive ones. Examples of such remarks attained from both schools are listed in Table 24. Any references to teachers or pupils have been omitted.

TABLE 24. Parents' comments on sufficiency of assessment information

Sufficient Information	Insufficient Information
<i>No expectations – no pressure on the child.</i>	<i>I'd like to hear more feedback from the teacher.</i>
<i>Development discussions are a good tool to inform about the school attainment of a child.</i>	<i>We get no information what so ever; we don't receive any kind of assessment of language proficiency.</i>
<i>I haven't missed information particularly on language proficiency, for language <u>studying</u> starts in the 3rd grade.</i>	<i>We don't know exactly at which level the studying of English should be in a given year.</i>
<i>In my opinion, coping with homework and success in tests tell, ok.</i>	<i>Only now from the X grade onwards we get allusions from assessment of English as a school subject.</i>

Parents who were content with the quantity of current assessment information indicated, for instance, that study of EFL provides sufficient information on English proficiency in form of tests, grades in reports and home work. These notions clearly suggest that CLIL study is perceived as similar rather than complementary to EFL. The underlining in the above parents' comments also insinuates that CLIL is somehow less serious than EFL or that assessment of language proficiency is only allowed in EFL contexts. Quite negative, even harsh comments were made by discontented parents who revealed that linguistic issues are not brought up, and nor are objectives of CLIL elucidated. The EFL study with standardised assessment practices communicates proficiency issues and assessment for the first time in the 3rd grade. In the interviews, the parents, who all were from the TS, were expressing mixed feelings regarding sufficiency which validates this finding.

When asked about the **sufficiency of information regarding progress in language proficiency, the proportional distribution remained quite similar**. In the TS, 66% of parents found they did not get sufficiently information on their child's linguistic progress, while in the CS, 28% of parents had parallel thoughts. It thus seems that **parents in the CS, on the whole, are more content with the quantity and quality of assessment information they receive from school**. Some remarks on the topic, sufficiency of assessment information, were again made in the questionnaire both for and against, as presented in Table 25.

TABLE 25. Parents' remarks on sufficiency of assessment information on English progress

Sufficient Information	Insufficient Information
<i>I feel that I can ask the teacher when needed. The child takes care of the school rather independently.</i>	<i>Regarding this I would like to have sub-goals and feedback during the term instead of the mere school year report.</i>
<i>Development discussions, word tests and tests provide us sufficient information on the direction of our child's progress.</i>	<i>Since we haven't received any feedback, we have assumed that [the language proficiency] is sufficient to follow instruction.</i>
<i>It is one school subject among others. Information has been the same as in other subjects.</i>	<i>It seems to depend on the teacher how much you get feedback.</i>
<i>As necessary</i>	<i>Based on test ratings only</i>

In a similar manner as in Table 24, the idea that the child manages school work independently appeared a few times in the data, in which case the parental duties in monitoring the child's schooling would be reduced. Therefore, it is possible to contact the teacher when complications arise, and if the teacher does not give any specific indications of problems, the parents assume everything is going smoothly. The notions in Table 25 show that more detailed objective-setting and reporting would be appropriate in some parents' opinion, and that parents have experience of distinct assessment practices between CLIL teachers. Both Tables 24 and 25 contain examples of how English in CLIL is perceived comprehensively and as strongly associated with EFL.

A parent's letter

One parent, who had observed CLIL tuition during several years through his/her many children, wrote an attachment letter or rather an essay about his/her insights on CLIL. The parent speculates that the general linguistic skills of the class have a bearing on how the instruction can be organised. S/he also questions whether the motivation of some families to apply for CLIL has been appropriate since “many see CLIL very difficult and challenging for their child”. S/he derives from this that parents seem to know fairly little of CLIL methodology, “if anything”. Furthermore, s/he had noted that teachers’ outlook on CLIL appears to vary, and s/he calls for levelling of the linguistic demands between teachers within a school. Sudden changes in teaching materials (e.g. one year the mathematics book is in English, the next it is not) s/he denounced as contradictory. Although the parent sees teachers’ pedagogical freedom as excellent, s/he still persists with the unification of CLIL.

SUMMARY

Assessment in CLIL seems to be infrequent, insufficient, based on self-reliant acquisition methods and therefore inadequate. It appears that teachers, who do collect assessment information on the TL, do not always use it for feedback. For pupils, it is important to get feedback from the teacher on their English language proficiency and coping in the TL in school subjects. The younger they are, the more important the feedback is perceived to be. Girls considered it more important than boys. However, only 10% of pupils reported receiving feedback frequently.

The most common ways for pupils to get feedback are tests, coping in the school assignments and school reports, whereas parents receive feedback mostly through tests, the child’s managing in the schoolwork and their own observation. These means are based on one’s own activity and ability to interpret the material and evidence available; it is indirect, implicit feedback. Almost half of the parents considered the received feedback as insufficient and 63% of pupils would like to obtain more feedback on their English proficiency in CLIL. Based on what was noticed in this section, it can be claimed that parents’ expectations and sometimes weak knowledge base of CLIL also hamper CLIL assessment.

6.4 Preferences and future visions

This section paves the way into future practices in CLIL assessment. It reflects the preferences of pupils and their parents and encapsulates the developmental future visions of CLIL teachers. The research questions 1.3 ‘*How should CLIL assessment be developed?*’ and 2.2 ‘*Which methods and practices do CLIL pupils and their parents prefer in receiving information on pupils’ English language proficiency and its development?*’ will be addressed in this section.

Pupils

A considerable majority, 63% of all pupils (n=107), expressed a need to receive more information on their coping in English in CLIL subjects. When differentiated according to gender, it was apparent that such information is somewhat more important to girls (68%) than boys (58%). There is no reason to assume anything other than Boud and Molloy's (2013, 1) statement in a general sense: "Learners care about their work and care about how it will be judged". Pupils spend a considerably long time in school and develop a relationship with the teacher which can be as close as the one with parents. What and how the teacher appraises, matters.

The pupils were also asked in which ways they would like to receive feedback. **The most preferred means of receiving feedback were rather traditional**, as shown in Figure 29: teacher's oral feedback (64%), school report (52%) and tests (51%), whereas self- and peer assessments as well as more regular written feedback did not gain support.

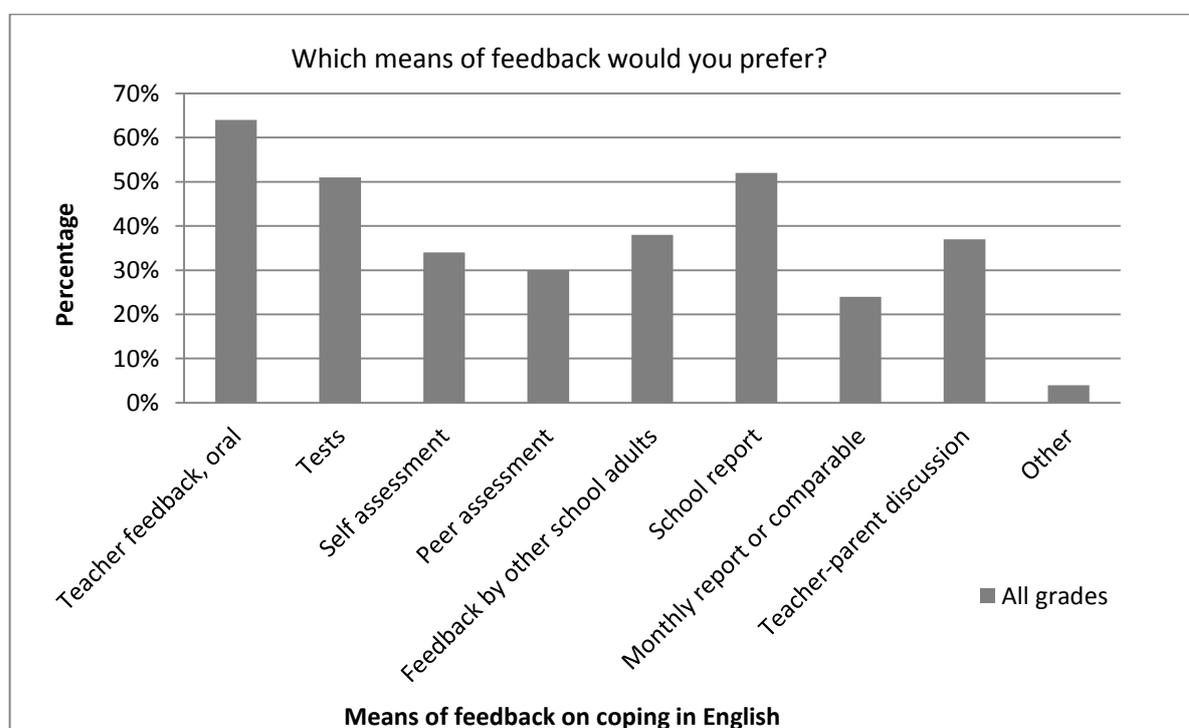


FIGURE 29. Means of feedback preferred by pupils

According to the theme interviews, it also appears that **feedback from an adult, even someone other than the teacher** (e.g. native teacher, teacher trainees, language assistants), **is more important than self-assessment or peer assessment**. Pupils may find this kind of feedback more reliable because the adult is perceived to be more knowledgeable and linguistically competent (Cf. ZPD). Pupils also seem to think that it is the teacher they are being assessed for rather than for themselves. Examples of such ideas are in the following quotes.

For my part, it is much better that the teacher is assessing [language proficiency] because s/he [...] says as it really is. (girl, 5th grade)

Some think that the downside is that you, for example, overestimate yourself and the teacher underestimates you, then it's quite a hassle. I think it is good that the teacher gets to know what I think of my own skills. (girl, 4th grade)

S/he masters English and then school mates wouldn't dare to say that "you are better than I" but the teacher can say that "you are really skilled". (girl, 3rd grade)

School reports were pupils' second most preferred means of getting assessment information. Neither of the research schools, however, provide a report with more specific information on CLIL study, as one pupil (boy, 4th grade) points out.

Those reports, nowadays there is only 'excellent', 'satisfactory', 'good' and so on. Principally, when it says 'good' or 'excellent', that's all you get; actually it is only the overall result.

The disappointment of this boy may be derived from the fact that he considers himself as skilled, and would like to be praised for his abilities. As seen in prior studies (e.g. Kärkkäinen 2011), children's belief in their own learning potential and abilities is higher than that of teachers and parents. The information that is mediated (or not mediated) on their academic potential thus serves as building blocks for their academic self-concept (ibid.).

Self-assessment seems to be regarded as more reliable when it is combined with teacher assessment. Its purpose is to ignite further improvement or to target study in areas of need. One pupil brought up an inarguable statement in favour of self-assessments which is why s/he would prefer them over others.

Well, I think [self-assessments] are good because no one can say that "hey, that's wrong". (boy, 3rd grade)

If there, [in self-assessment,] would be teacher's assessment aside and then you would compare your own and teacher's [assessments], then you would know better how to improve it [language] (girl, 5th grade)

In comparison to the feedback practices as perceived by pupils (Figure 25), children would like to obtain more immediate oral feedback from teachers (+13%) and other school adults (+12%) such as teacher trainees, native teachers or comparable. Self-assessment would also be appreciated by pupils (+7%), whereas somewhat less desirable are tests (-19%) and teacher-parent discussions (-6%). Attending developmental discussions may manifest itself as threatening to children who are unsure of the content of the discussion in advance. Therefore, the more constructive feedback the pupils receive, the higher the trust in the 'benevolence' of the teacher. As was noted in the theory section, assessments are, unfortunately, sometimes also used to sanction on other than study-related matters.

Parents

Approximately half of parents were discontented with the quantity of information they received about their child's language proficiency and its development in CLIL, and on that account it is logical that **76% of all parents** (n=99) **wish to receive more feedback** (Figure 30). This wish was identical in both schools with almost exact numbers (39/51 in the TS and 36/48 in the CS),

although the parents in the CS were much more pleased with the quantity of information they received in the first place. What this suggests is that both research schools would benefit from developing a CLIL assessment strategy or restructuring the practices currently in use. A proposition for such an assessment scheme is presented in chapter 8.

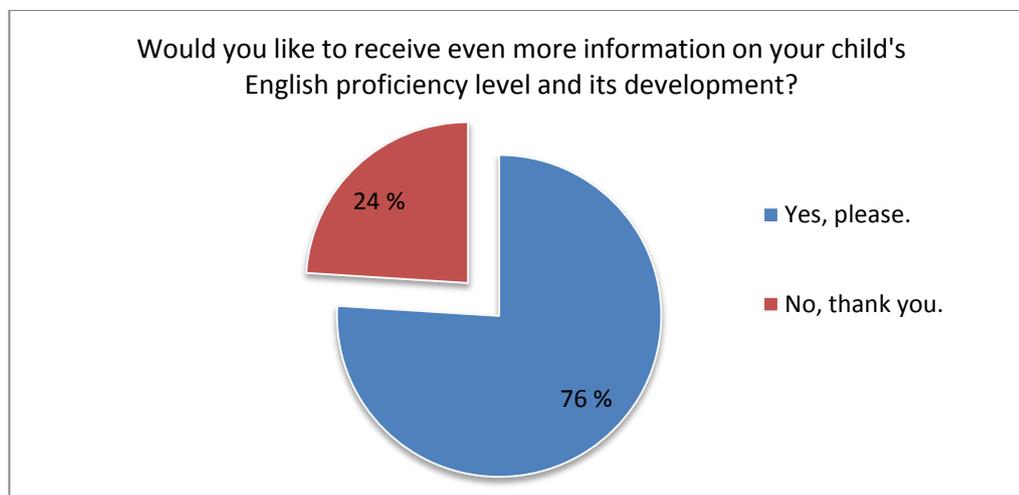


FIGURE 30. Parents' wish to receive more information on the English proficiency of their child

Parents also passed additional comments, some of which are exemplified in Table 26. These comments demonstrate how, as already noticed before, language acquisition is not the primary motivation for all parents. Although additional information is not perceived as necessary, it would still be accepted when provided. Those parents wishing for more information express their desire for language objectives and general aims of TL acquisition. Parents' desire for more CLIL-related linguistic feedback is obvious. The accuracy of that feedback was not considered to be as important as the overall quantity. However, **almost seven out of ten parents (68%) wished to receive more precise and detailed information on the language skills of their children as well as its development.**

TABLE 26. Examples of parents' comments on additional assessment information

Yes, please.	No, thank you.
<i>Absolutely!</i>	<i>Is not necessary.</i>
<i>We would also like to know what is required in each year in a CLIL class.</i>	<i>Is not necessary as long as school attainment is good (it is not the language that is the main thing but the normal school attendance).</i>
<i>It would be good if the teacher would collect all tests and word tests to store them and at the end of a school year return them in a portfolio for each pupil.</i>	<i>I don't have any specific expectations of language learning. Language is just one part of school attendance which the child handles him/herself according to his/her abilities.</i>
<i>Separate feedback on this would be good.</i>	<i>It is certainly always nice to hear if the child advances in his/her studies or language proficiency.</i>

Some parents also commented on the necessity of more exact information; both objecting and favouring remarks were made. Examples of these remarks are collected in Table 27. The

comments chiefly replicate the ones given in Table 26, but were not always given by the same persons.

TABLE 27. Examples of parents' remarks on additional, more detailed assessment information

Yes, please.	No, thank you.
<i>It might be nice to have personal achievement goals for the child.</i>	<i>Not necessary.</i>
<i>It would be interesting to know more about the objectives. However, during these X years this has gradually become clearer. In the 1st grade, I felt that the CLIL method did not "open up" in a concrete way.</i>	<i>X [pupil's name] has done well at school. If something changes, then a more specified discussion is in order. Now everything is ok and X gets along independently → well.</i>
<i>Information would make things better and perhaps also motivate the child.</i>	<i>On the other hand, maybe some kind of feedback would be nice</i>
<i>If not too complicated to arrange.</i>	<i>Is not necessary.</i>

Which, then, are the means through which parents would prefer to gain more information on linguistic CLIL issues? These preferences are presented in Figure 31. **Most** of the 87 **parents opted for teacher–parent discussion** (59%). The next most favourite means were other reports from the teacher (48%) and other discussions or messaging with the teacher (38%). Apparently, parents wish to obtain individualised feedback concerning their own child only, but parental evenings are typically instances for general discussion and feedback concerning the class as a whole.

The following interview quotation remarks on how a general feedback instead of individual, more specific feedback may raise more questions than give answers and evoke feelings of insecurity.

R: Now there's practically three years of CLIL study behind. How does the growth of language proficiency manifest itself? Or does it?

P: *This is probably the thing I'd like to get more feedback of. I somehow was shocked in the parental evening when the teacher praised the class "a very good class", but I don't see it in my own child. There are pupils in the classroom that watch the Simpsons and their English is so good that they notice if it is translated wrongly. I don't see that kind of level in my child, no way. It feels so strange and terrible and I think "help, if my child is supposed to be that skilled". I don't recognise that kind of a level in [his/her] language at all.*

Parents, similar to their children, thus appear to prefer immediate and individual contact with the teacher. One parent had added an adjustment regarding teacher-parent discussion in the questionnaire:

Perhaps a face-to-face discussion with the teacher only. The parents' quarter is not very suitable for [receiving information on child's language proficiency] because the child is there and there are also many other things to discuss. And there's that one quarter of an hour per year only.

This implies that **more time is necessary to converse** about school, studying and language-related topics. Other means of feedback comprised 8% of all choices.

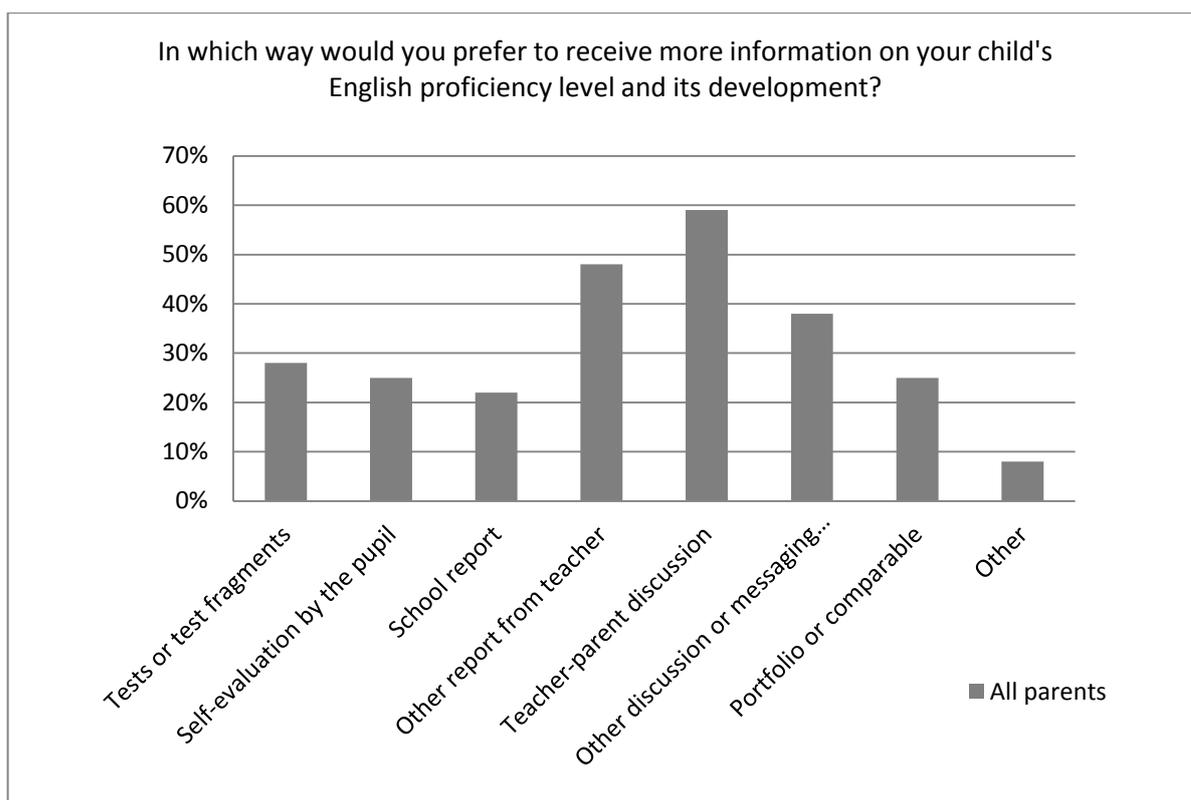


FIGURE 31. Parents' preferred means of receiving information on their child's English proficiency level

In the additional comments (open space for comments in the questionnaire), three parents propounded that a complementary attachment should be included in the school report; one of them listed suggestions for aspects to assess: "Specified components for example in school report or development discussion: oral, listening, written, grammar". Another parent called for markedly separate assessment of the two types of English:

I would like to see that the English in English class would be assessed separately. How well the child has acquired the English instruction given in a CLIL class. Not just the school subject grade.

Such comments were not common among parents, for English in CLIL is most often associated with general study of English. The third parent in favour of a CLIL English attachment wished for a written, textual attachment. Some schools offering CLIL instruction actually do issue complementary CLIL attachments but neither of the research schools does. Some parents hoped for more tests, for instance as follows:

I'd like to have more homework examinations (written random tests), test sections and explicit homework in English (what you need to master actively and what passively and so forth). [Underlining in the original]

The underlining in this parent's comment is likely to originate from the confusion caused by "contradictory" and unestablished practices one parent (see a parent's letter on p. 168) had discerned. Although parents make observations and have manifold requests, they seem to respect teachers' work and do not wish to increase their work load as in the following quotation.

I think that teachers have enough work already, so no monthly reports are certainly needed. But maybe some kind of scale or measurement system for the development of language four times a year: once in the middle and at the end of each term. There you could get feedback were they are going.

The **idea that assessment is also feedback for teachers** on their own success has apparently been realised in some families, as the following quote depicts exemplifying accountability and formative views on assessment:

I know that there are also families, such guardians that don't think like this but they see that assessment tells how the teacher has taught and whether the teacher has taught well or badly. But in my opinion assessment tells where the child is in his/her leaning. I think the perspective is a bit different if the starting points of assessment are the objectives that have been defined for the instruction. Assessment then specifies how those objectives have been achieved.

This excerpt also appropriately crystallises the essence of assessment and the educational tripod: objectives, instruction and assessment.

Teachers

The query regarding future prospects generated a wide variety of suggestions, ideas, visions and opinions – both for and against pupils' language assessment in CLIL. In this study, a minority of the participant teachers felt that no specific assessment tools are needed; the **majority of teachers are inclined to develop and embrace various new tools for assessment purposes**. Teachers made remarks on CLIL-related assessment issues both generally and specifically. From the data, I generalised classes that need readjustment in the teachers' opinion. I will focus on these issues first. Understandably, many of the developmental aspects stem from issues CLIL teachers had perceived as challenging or problematic (see 6.2) and are therefore partly overlapping.

- **uniform foundation** (existence of CLIL curriculum and assessment criteria)

Teachers feel that due to the absence of proper CLIL curriculum stating the mutual and explicit linguistic objectives and specific contents to be learnt through the TL, there is no coherence in CLIL tuition.

Perhaps explicit recording of grade-specific proficiency levels, concrete objectives and development of assessment instruments for them.

Class-specific objectives related to language learning are different according to teacher.

The objective needs to be clear. Otherwise you are just scrabbling without knowing which direction to go, what to do [...] Objectives should always be SMART: specific, measurable, achievable, realistic and time-bound.

- **coherence** (uniform tools, methods and practices within the school)

This issue is closely connected to the uniform foundation of CLIL and a result of it. It also has implications to learner equality.

- **basic vocabulary**

Basic vocabulary for CLIL was seen as advantageous insofar as it contributes to coherence in many levels: teacher, school and national levels. A few school material publishers have attended to this and provide subject-specific terminology in the teacher's guide. This is an innovation that could be found online and be helpful specifically for novice CLIL teachers struggling with many other challenges. The following quote is an example of a demand for generic vocabulary acquisition tests.

For assessment purposes, it would be very good to have scaled tests for vocabulary of different school subjects. Naturally the vocabulary varies from school to school and teacher to teacher, but assessment of the mastery of basic vocabulary would be very important for older children. This would clearly be wake-up-type assessment in grades 4–6.

- **establishment of various tests** (admission, norm-referenced or standardised tests)

Different types of tests were suggested for CLIL purposes, as seen in the quotes below. In Finland, entrance in specialised classes such as music classes almost without exception requires passing an admission test. Teachers recognise the benefits of preventing linguistic learning difficulties. Standardised tests may have a different connotation in Finland than in countries with accountability policies; such tests are seen as generally valid indicators of language proficiency rather than absolute measures of it.

There's room for improvement especially in pupil selection. In our school, there are no criteria for pupils applying for admission. Lots are drawn for pupil admission.

At the moment, there are no standardised tests in use, but everything is adapted from English materials or prepared by the teacher. It would be interesting if there was at least one general test package which would be administered once a year per grade. Designing it would be challenging, of course, because of the diversity of CLIL instruction in our country.

- **use of the CEFR proficiency levels**

Apparently, the CEFR is not very well known among the class teachers; it was brought up by a few teachers, but where exploited, as the first remark demonstrates, there is variation in the use.

Assessment tools (prepared by the teacher/school) as such are fairly sufficiently in use, although their linking with the proficiency levels is up to the teacher.

In my opinion, the proficiency levels of the European Framework are a brilliant tool, also in accordance with the curriculum. There's no need to adopt any other tool alongside.

Plenty of problems and challenges were distinguished, but many teachers had also contemplated how to meet these challenges, and some had even found concrete solutions for the trials and tribulations of assessment, as was shown in section 6.1. Both **pupils' self-assessment and language awareness were topics that were rather often mentioned as future signposts**. These signposts are also in line with the NCC draft (2014) for the NCC reform. Moreover, teachers visualised dialogic, accepting, encouraging and positive future assessment that would be more informative than a single grade. Activity and effort should be recognised as well.

Instead of interpreting the grades given by the teacher, pupils would obtain a genuine, personal conception of his/her own skills (linguistically and content-wise) and improvement issues.

Dialogues in which the areas [of language proficiency] would be discussed and the objectives would be set together, what are the next goals. And then we would assess if those are achieved or not.

Teachers also suggested various approaches to assessment. Finding meaningful ways for pupils to show their skills was seen as one important aspect that should be better taken into account in the future. **Participatory assessment approaches seem to be a future trend according to the visions of CLIL teachers.** Language development or progression can be made visible or audible by recurring tasks, with criterion-referenced inferencing and rubrics being another way for students to participate in assessment when the objectives are defined together.

Evidence-based assessment approach would be an interesting idea. It could also be a motivating factor. The language study gets a meaning when you have to show your skills. It could be a project work you could do to show your proficiency and then present it in written and oral form, both. The topic should be interesting to the child and s/he could do a larger project work.

I think it would be a good idea to create criterion-based assessment together with the class for one year and then follow that. In the autumn the objective criteria would be created, what we would like to achieve, and then mid-assessment and in the spring we would go back to that and see how we succeeded.

The teachers were also rather productive in bringing forth assessment methods they would need. Many of them pertained to oral language assessment which was perceived as a problematic skill to assess time-wise and due to a lack of instruments. One teacher suggested that at regular six-month intervals the same book or picture could be shown to the testee and the utterances would be audio-recorded for later examination, a variation being in written form. This way, the progression of language production and vocabulary should become evident.

SUMMARY

Both pupils and parents would like to receive more assessment information from the teacher – explicit, direct methods of giving feedback were preferred over implicit and indirect means of receiving information. Conversations are preferred over written feedback, but a combination of both might please the majority of parents. Both pupils and their parents wish to have personal contact with the teacher, parents also wished for this contact more regularly. School reports were considerably more favoured by pupils than by parents. Teachers named multiple areas to be improved, of which many were already mentioned in relation to challenges hampering CLIL assessment. Such areas were, for example, a need for uniform foundation, cohesive assessment methods and the use of the proficiency levels as defined in the CEFR. The vision statements of CLIL teachers included raising pupils' language awareness through self-reflection and making language progression more salient.

7 COMPUTER SIMULATIONS

In addition to describing the state of art of language assessment in Finnish primary CLIL classrooms, I have experimented with two prospective CLIL assessment methods: language portfolios and LangPerform computer simulations. Within the scope of this research report, I will submit the most pronounced results obtained from the computer simulation experimentations implemented within the project PROFICOM (2013), while the portfolio experimentations will be subject to a separate publication (Wewer forthcoming) because the experimentation is still ongoing. The implementation of the two computer simulation experiments is described in sub-section 5.4.2 and computer simulations are theoretically addressed in 4.2.3.

The focus of this research was not to scrutinise pupils' language proficiency, but to test the appropriateness and affordances of the computer simulation as an alternative assessment method for language proficiency in CLIL contexts in order to become prepared for the technological changes in education. As a result, this chapter attempts to answer the third research question '*What are the key issues and advantages in using computer simulations as an alternative assessment method in CLIL as perceived by pupils and their parents?*' with its two more specific sub-questions '*What kind of information do computer simulations yield on pupils' English language proficiency?*' and '*What kinds of thoughts and experiences are produced by the simulation experiments in pupils and their parents?*'

The results yielded from the two experimented simulations will mostly be presented as an integrated whole, but will be dealt with separately regarding the two participant groups: pupils (7.1) and their parents (7.2). Both sections will be completed with a concise summary. The results submitted here basically describe pupils' and their parents' experiences and represent more elaborate reflections of the simulation piloting than mere "I tried it and liked it" statements (see Chin, Dukes and Gamson 2009, 554). Teachers were not part of this research phase. The appropriateness of the LangPerform simulations in primary CLIL will be further discussed in chapter 8.

7.1 Pupils

Altogether, 146 pupils participated in the simulation pilot. The results I present here are based on video documentations taken in the CS sessions, pupils' questionnaires and video interviews, and they are categorised into the following units: test anxiety, comparison of simulations, effort and endeavour, coping in English, four basic language skills, perceived linguistic difficulty of simulation tasks, appropriateness of a computer simulation in assessment of various areas of language in CLIL, examining the performance, overall comments on the simulation and coverage of the simulation contents at school.

Test anxiety

The results show, logically, that **it takes some time to get used to a new assessment system and adopt a new technology**. Simulation as an assessment method was new for all pupils. Therefore, afterwards, they were asked to estimate their anxiety level at the beginning of both simulations. Test anxiety weakens test performance (e.g. Chapell & al. 2005) and may also affect the way the testees perceive the test itself. It was emphasised in the beginning of each session that the simulation was not a test for the pupils but a piloting situation, and the simulation performance would not affect their grading.

As shown in Figure 32, 64% of pupils (n=74), most of them girls, were feeling unrelaxed to varying degree when starting the simulation, while in the second simulation, the corresponding percentage was only 49% (n=35) regarding 4th graders who piloted the simulation for the first time. This may be because the 4th graders had already heard from their elder peers about the simulation and its characteristics, and therefore already knew what to expect. Also, 51% of the 5th graders (n=37), who were in the pilot for the second time, and were in that sense already experienced, reported feeling less stressed than during the first time.

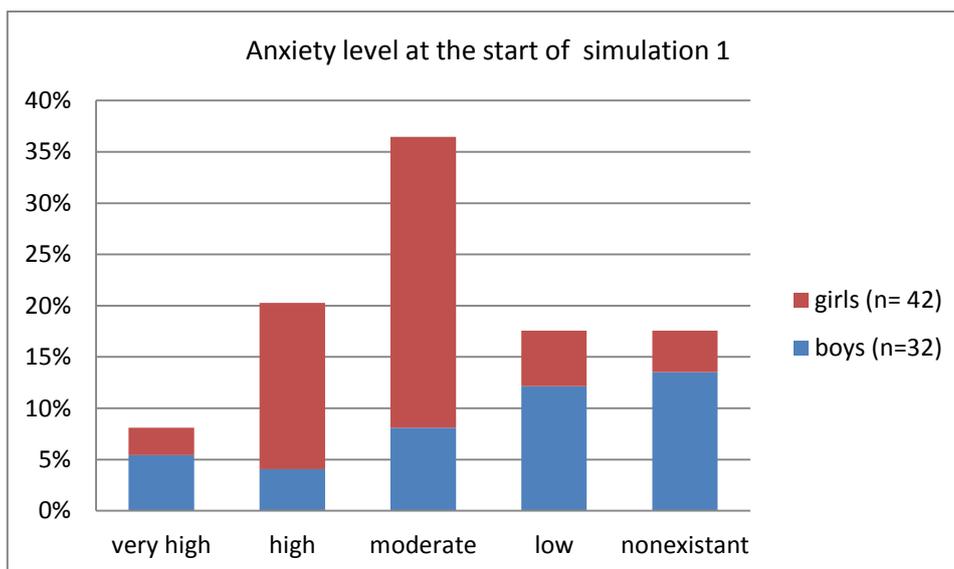


FIGURE 32. Pupils' anxiety level in the beginning of simulation 1

The sensations and fears that the pupils felt were causing this test anxiety pertained mostly to exposing oneself linguistically to others, the difficulty of the English language in the simulation and facing the unknown, as the following quotes depict.

I was nervous that everyone's listening to me and will start laughing and that the people in the computer will see me. (girl, 4th grade, very high stress level)

I was quite excited whether it is easy or difficult and how much I have to talk. (girl, 5th grade, high stress level)

In both simulations, most pupils (six out of ten) relaxed during the simulation performance.

Because I noticed that I can [manage] well enough. (girl, 4th grade)

I was less tensed when I noticed that the simulation English was not that difficult. (boy, 6th grade)

They noticed that the linguistic difficulty was not as high they had expected and they were able to cope through the TL.

Comparison of simulations

The 5th graders, who had experienced two simulations, were asked which one they preferred. **Most pupils (59%) preferred the second simulation**, which was situated in an international school; the first location was an American home in Michigan. Almost a third (30%) liked both simulations equally. Their justifications revealed that in the second simulation, pupils were functioning within their own comfort zone, because a school environment is familiar to them; they know what kinds of things to expect in such a setting and the communication occurs mainly with other children or adults whose roles are clear to pupils. Also, the following quotes demonstrate that the mode of assessment was becoming more familiar to pupils.

There were English tasks more to my liking (boy, 5th grade preferring the first simulation)

I knew by and large what to do. It was also a bit easier. (girl, 5th grade preferring the second simulation)

I liked them because it was nice to talk and listen to British accent and American accent (girl 5th grade liking both simulations equally)

In the 5th graders' (n=35) estimation, both simulations were linguistically equally demanding; 49% considered the first, and 51% the second simulation to be more difficult language-wise. This suggests that the sensation of difficulty is utterly individual and personal. It is also possible that the content material affected the gauging of the linguistic difficulty. The more linguistically diverse tests can be produced, the better, in order to cater for a wide variety of pupils.

Effort and endeavour

In both piloting rounds, pupils (n=142) were asked to gauge how much they invested effort in the simulation (Table 28). **The majority of pupils (74%) reported having tried their best.**

I didn't necessarily understand everything, but some. I reached my own ability level. (boy, 4th grade)

According to their own estimation, every fourth pupil did not reach their own potential, which was also noticeable in the sessions. Although participating in the experiments was voluntary, not all pupils seemed to be in earnest.

TABLE 28. Pupils' effort and endeavour in the simulation sessions

Effort and Endeavour	Girls	Boys	Total	Percentage
I tried my best.	67	38	105	74%
I could have done better if I wanted to.	16	21	37	26%
I did not try in earnest.	0	0	0	0%
Total	83	59	142	100%

This may also be a matter of keeping up appearances among peers; it is safer to ruin the performance than to let others notice possible weaknesses in one’s language proficiency. Such behaviour represents avoidance strategies. Some pupils also ‘froze’ in the situation; they were not able to produce an utterance at all at first. Girls tended to put more effort into the simulation performance than did boys; 81% of girls stated trying their best, whereas only 64% of boys did the same. If the simulation had been implemented in private, the result in this respect could have been different, as one pupil stressed:

*I could have probably been able to speak clearer if there weren’t others doing the simulation.
(girl, 5th grade)*

Furthermore, the technical difficulties briefly described in 5.4.2.2 must have affected pupils’ motivation and performance.

Coping in English

Pupils (n=142) viewed their coping in English in various BICS and CALP type tasks (see Appendices 11 & 12) rather positively; an express **majority (93.5%) of pupils in all piloting sessions estimated that they managed at least moderately, if not even better** (Figure 33).

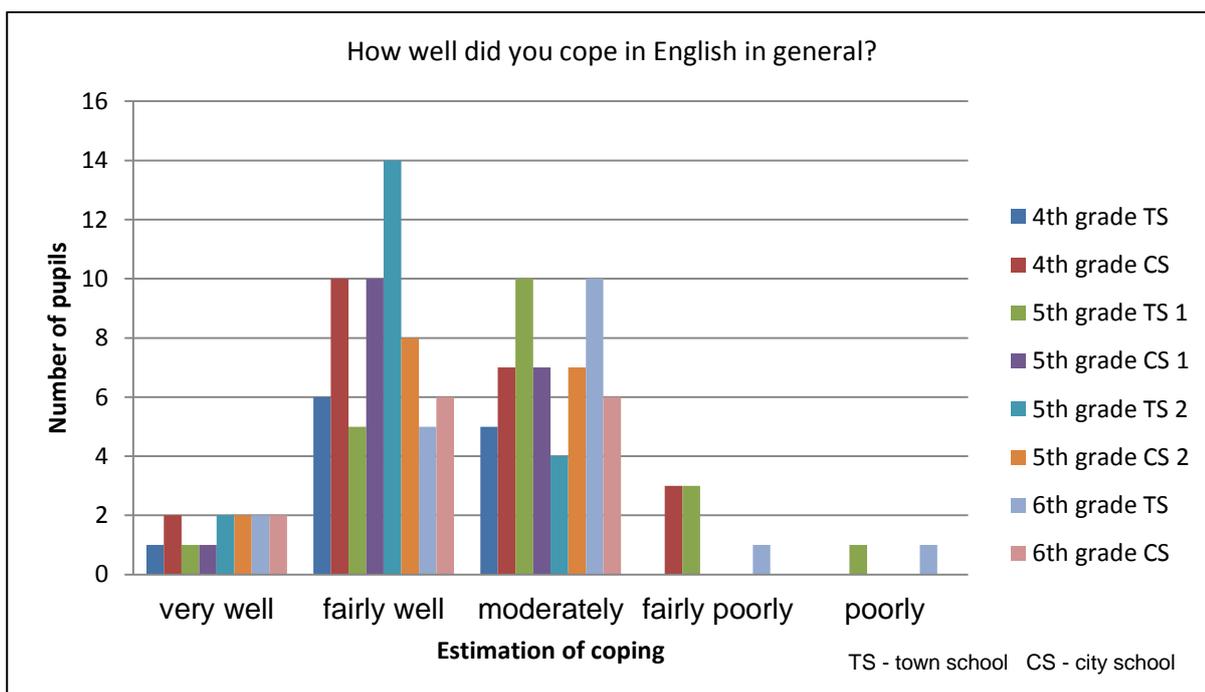


FIGURE 33. Pupils’ estimation of their coping in English in both simulations

The graphic presentation also demonstrates how 5th graders had gained more self-confidence in between the two simulations: many pupils in the TS perceived their managing in the second simulation as reasonably better than in the first (5th grade TS 2 in comparison to 5th grade TS 1). When interpreting the graphic, one should not compare the columns as such because the number of pupils varies in each class (see Table 21). The graphic merely implies general trends in the children’s perception of their own coping in English. **Boys deemed their coping in English in**

the simulation more positively than did girls in the two highest categories (very well and fairly well) as Table 29 displays. It is also notable that only a relatively small amount of pupils (6.5%) gauged their coping in the two simulations as poor or fairly poor while **every second pupil believed they coped very well or fairly well in the two simulations** (Cf. Kärkkäinen 2011).

TABLE 29. Percentages of pupils' general and gender-specific coping in both simulations

How well did you cope in English in general?	Number of girls	Percentage of girls	Number of boys	Percentage of boys	Total number of pupils	Total percentage
very well	7	8%	6	10%	13	9%
fairly well	35	42%	29	50%	64	45%
moderately	37	44%	19	33%	56	39.5%
fairly poorly	4	5%	3	5%	7	5%
poorly	1	1%	1	2%	2	1.5%
Total	84	100%	58	100%	142	100%

The occasional comments made by pupils reveal four primary reasons on which they blamed their unsuccessful coping in English: technical (sound) problems, anxiety, lack of (reaction) time and their own language proficiency (Table 30).

TABLE 30. Pupils' comments on their coping in English in simulations

Very well (9%)	Fairly well (45%)	Moderately (39.5%)	Fairly poorly (5%)	Poorly (1.5%)
<i>I knew what to answer but I was tense which disturbed the task performance.</i>	<i>It was ok but because the microphone didn't function so it was like ☹.</i>	<i>I think I'm doing ok, but it happened so quickly that I didn't have time to understand them.</i>	<i>In my opinion pretty badly because there wasn't enough time although sometimes a lot.</i>	<i>Because I understood the question but I somehow froze.</i>
<i>I understood everything but I couldn't hear all the questions so that might be the cause of mistakes.</i>	<i>I did quite well because English in my view is quite easy.</i>	<i>When I started to get nervous there were many 'erm' and 'I guess' moments.</i>		
<i>It was ok.</i>	<i>Sometimes I didn't know some words.</i>	<i>I couldn't answer all questions and a couple writing tasks remained unfinished.</i>		

As can be seen in Table 30, pupils were inclined to explain their coping through negatives ('I wasn't able to...', 'I couldn't...', 'I didn't...') which is why assessment methods that document skills rather than judge them are more positive in accentuating things pupils are able to do rather than things that they are not.

Four basic language skills

The four basic language skills are receptive skills of listening comprehension and reading comprehension and productive skills of speaking and writing. In the analysis stage of simulation 1 I stated that the question wording was weak in the sense that pupils were forced to choose the

most difficult skill for them. Many pupils were reluctant to do that, thus a number of answers had to be ignored due to arbitrary replies. The question mechanics was changed in the simulation 2 questionnaire (Appendix 8). Regardless of this alteration, in both simulations, pupils (n=128) answered the question 'How did you manage in the following language skills in your own opinion?' similarly: **listening comprehension**, i.e. understanding spoken English, **was perceived as the area where the majority of pupils coped best** (46% in the first simulation, 73% in the second). This self-evaluation is in accordance with what has been found of language competencies favourably affected by CLIL (see Table 11). The dialogue in both simulations was designed to be easy enough for primary pupils to follow and participate in.

Reading, i.e. textual comprehension, was estimated to be the second easiest skill. There was some fluctuation in estimating coping in the productive skills: pupils in the first simulation considered speaking easier than writing, whereas the result was reversed in the second simulation. It is noteworthy, however, that in each group there were pupils who reported that all four basic language skills were equally manageable for them. In general, receptive skills were perceived as more effortless to cope with than productive skills.

Perceived linguistic difficulty of simulation tasks

The topics in the simulation tasks were chosen from common curricular themes, both from regular English (BICS language) and other subject areas (CALP language) to ensure that the tasks were manageable for pupils. The perceived linguistic difficulty of tasks in the simulations is shown in two separate graphics, because the tasks were not identical (Appendices 11 & 12). Moreover, the wording of the given question was slightly altered for the second questionnaire. In both questionnaires, however, it was stressed that the pupils were expected to estimate difficulty according to the language used, i.e. how difficult it was for them to manage in English in those tasks, not the task per se. The results obtained from simulation 1 are presented in Figure 34 regarding the most difficult and second most difficult tasks. Figure 35 is concerned with the very difficult and difficult ratings for the simulation 2.

Simulation 1

The first simulation was piloted by 5th and 6th graders (n=74). Concerning the perceived linguistic facility or difficulty of tasks **in simulation 1, pupils (n=69) rated tasks requiring use of subject-specific CALP-type language most difficult and tasks featuring BICS-type language most easy**. The easiest task was presenting oneself (59%); the second easiest tasks with equal ratings (15% each) were reading the Michigan text aloud and telling the time. These are tasks one encounters regularly in conventional EFL instruction.

As can be seen in Figure 34, the three most difficult tasks in simulation 1, regardless of the manner of interpretation and calculation, were: 1) explaining how to calculate a multiplication in columns; 2) a gap filling exercise in which pupils were asked to drag and drop the geographical and related terms into right places in a text which was the same as in the previous task (reading

aloud: introduction of Michigan); hence the terms had already been encountered once; and 3) the black bear text. In the black bear text, certain key words (e.g. hibernation, cub etc.) were introduced within context and underlined, and pupils were to read the text and then explain the underlined terms in Finnish. The instructions were not clear for the pupils, or they did not understand them: many pupils started to explain the words in English. This confusion might have affected the task's rating as the third most difficult, as inferencing the meaning from the context might have been challenging.

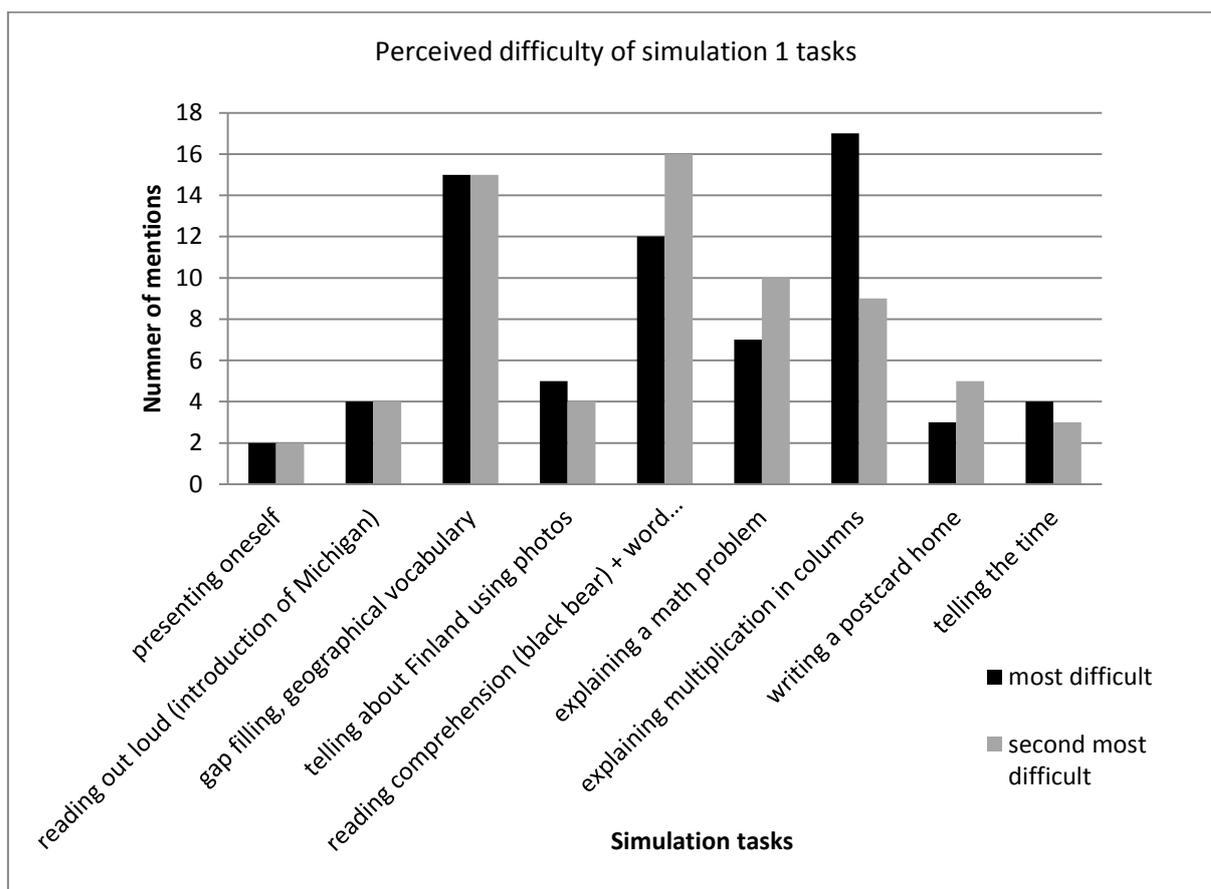


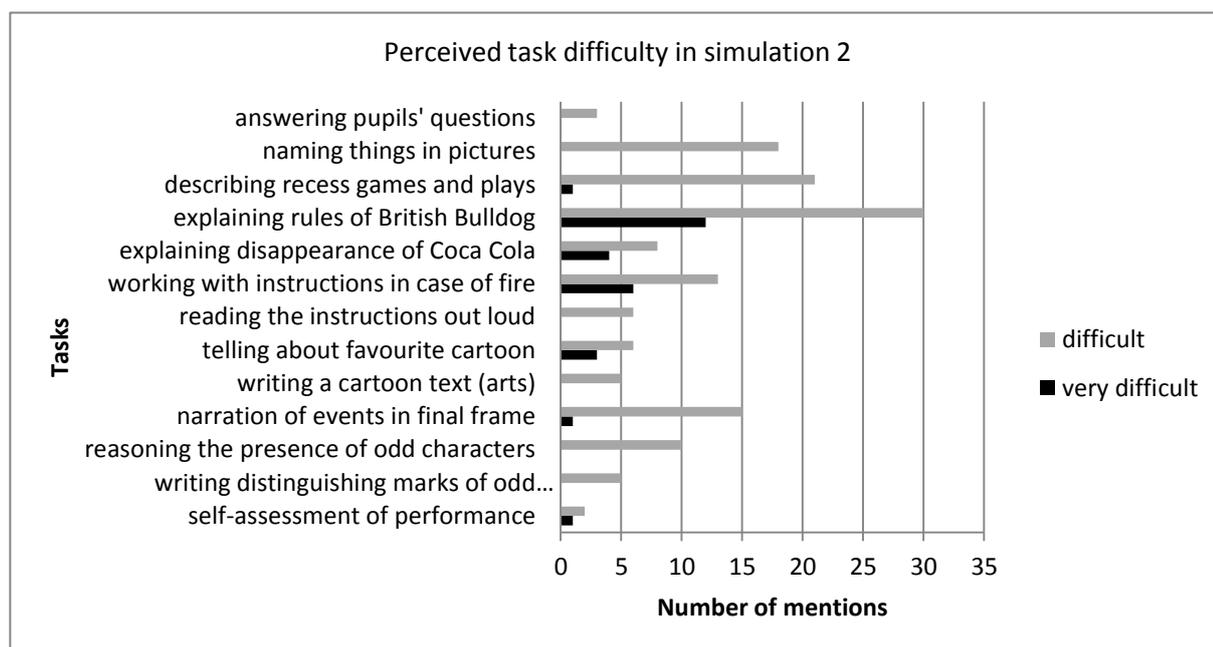
FIGURE 34. Perceived linguistic difficulty of tasks in simulation 1

Interestingly, the distribution of the perceived difficulty of tasks was levelled: **every single task was rated either as difficult or easy by at least one pupil**. Their opinions on the task difficulty thus varied considerably which, from the designer's point of view, is a desirable quality for a language test, since the wide distribution of ratings reveals that the simulation serves the diversity of pupils' language proficiencies. Another interesting finding was that particularly one class, the 5th grade of the CS, perceived the multiplication in columns task, in comparison with other classes, as markedly less challenging. This particular task was rated as the linguistically most difficult task by a quarter of the pupils. Only one pupil in the 5th grade CS class rated the multiplication task as the most difficult and another as the third most difficult – otherwise it was seen as easy or it did not receive any ratings at all from this class.

The class teacher, when I asked for possible explanations for this, gave five likely explanations: 1) the class has studied using an English-language mathematics books from the first grade onwards; 2) English is significantly present in every mathematics lesson; 3) pupils are very learning-oriented and motivated as learners; 4) everyone's right to peaceful work is respected and guaranteed; and 5) mathematics is viewed as fun. For example, mathematical extra assignments are perceived as a reward. The pupils were thus very accustomed to using and expressing ideas in 'mathematical English' which implies that they have been exposed to rich input and have had plenty of opportunities to practise subject-specific output.

Simulation 2

The second simulation was piloted by 4th and 5th graders (n=72), and the linguistic difficulty level of tasks increased gradually according to Bloom's taxonomy of learning objectives (Appendix 12). Again, **pupils considered the tasks requiring subject-specific language to be more difficult than tasks eliciting BICS-type language**. Such tasks were explaining the rules of the game British Bulldog requiring sports-related vocabulary and sequencing of actions, which 70% of 4th graders and 50% of 5th graders estimated as either very difficult or difficult, and working in different ways with instructions 'in case of fire' (40% of 4th graders and 14% of 5th graders). It is understandable that the 4th graders with less English study perceived the tasks as more difficult



than their older peers.

FIGURE 35. Perceived linguistic difficulty of tasks in simulation 2

Occasionally, the tasks I had striven to compose as more demanding and elaborate were estimated as easy by pupils. Such an example is the Coca-Cola task which was established by a cartoon animation. The possibility cannot be excluded that sometimes pupils may have actually estimated features other than linguistic difficulty, for example the mechanics of the task (e.g. telling, drag and drop) or the visual representation. Furthermore, the anticipated subject-specific

language (e.g. solid, evaporate) may have been compensated with BICS-type expressions (e.g. ice cubes, go away) which explains why the task was not perceived as demanding. Compensating as such is resourceful and a sign of language deficit strategies in which CLIL pupils appear to be skilled (see e.g. Rahman 2012). Fully explaining why the liquid Coca-Cola with ice cubes disappeared would have entailed vocabulary related to different forms of water. On one hand, the task design was not successful if the anticipated language was not elicited; on the other, it is possible that children had not studied forms of water in such a way that they would have been able to recycle the vocabulary.

Appropriateness of computer simulations in CLIL assessment

The pupils' (n=143–146) position towards computer simulation as a method for assessing various aspects of language was rather positive as displayed in Figure 36 in which the estimations of pupils from both simulations are combined.

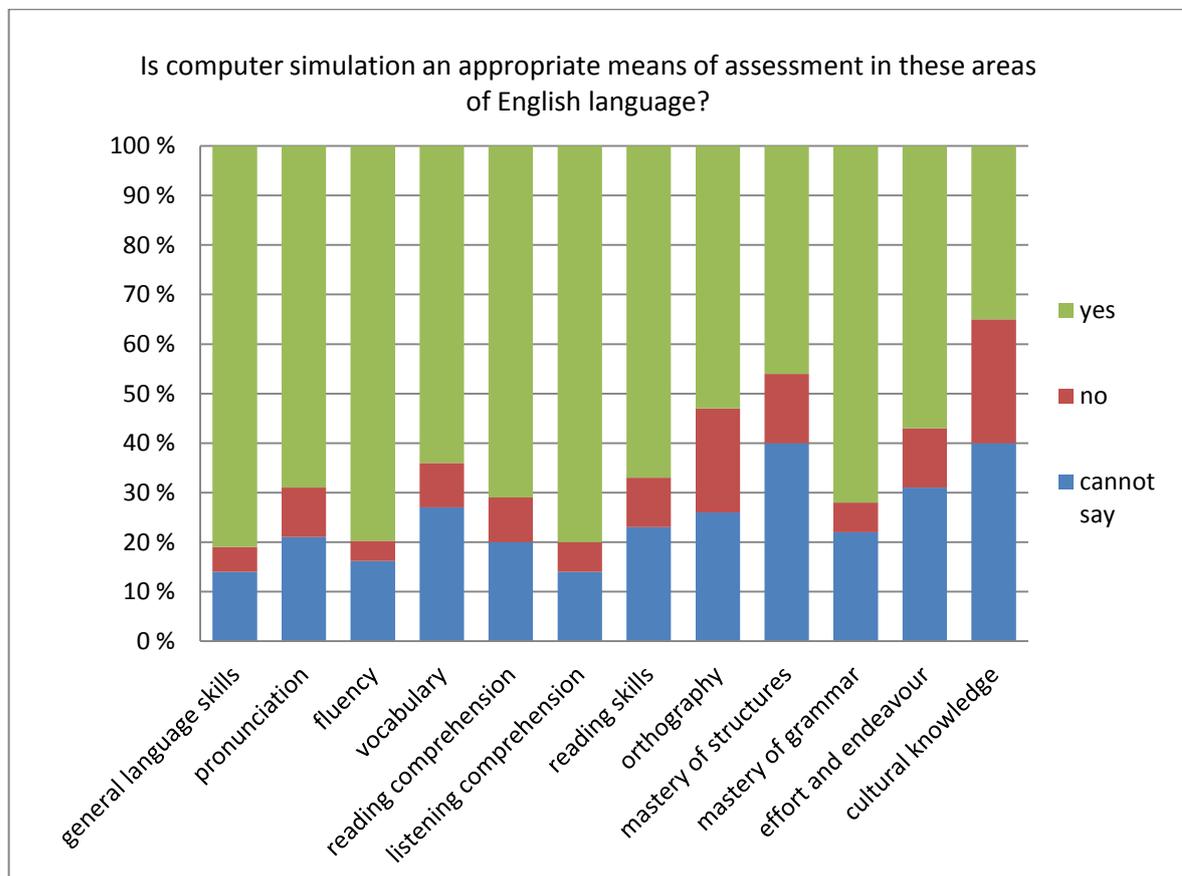


FIGURE 36. Appropriateness of a computer simulation in assessing aspects of language (pupils)

According to the pupils' perspective, **computer simulation is an appropriate means of assessment especially in gauging general language ability (81%), listening comprehension (80%) and fluency of spoken language (80%)**, although a number of other aspects also received high appropriateness ratings. A computer simulation is, according to pupils, least appropriate for assessing cultural knowledge, orthography and effort and endeavour. Reading

skills denote ability to read English texts adequately, the letter–sound correspondence of which differs significantly from that of the Finnish language.

The question did not touch upon every single language aspect. For example, sociocultural aspects and register were ignored due to the intricate nature of the concepts. As for cultural knowledge, neither of the simulations contained overt instances of culture-related linguistic behaviour other than English conventions of politeness, which is rather a sociocultural aspect. The complicated concepts of structure and grammar may have caused some confusion for younger pupils – the amount of ‘cannot say’ choices was fairly large in the simulation 1 questionnaire. This issue was taken into account in the simulation 2 questionnaire: before each query session, I explained the concepts to pupils with examples both orally and in writing on the blackboard. The concepts may still have remained abstract, for this exercise did not change the situation considerably; pupils still tended to choose the alternative ‘cannot say’.

In the early years of formal EFL instruction, grammar is not substantially present; its role is rather permeable and transparent. However, structures, are, and should be practised from the beginning onwards, but it is possible that teachers, for that matter, do not mention concepts such as ‘structure’ that often. In the simulation, there were no grammar tasks, but mastery of grammar and structures becomes essential and explicit in connection with independent production (e.g. writing a postcard home or filling in the speech bubbles of a cartoon).

The appropriateness of computer simulations in assessing *content knowledge through a foreign language*, was also appraised by pupils (n=143). On average, **65%** of them **considered a computer simulation to be an appropriate method to assess how pupils master subject matter in English**, while 28% were indecisive (Figure 37).

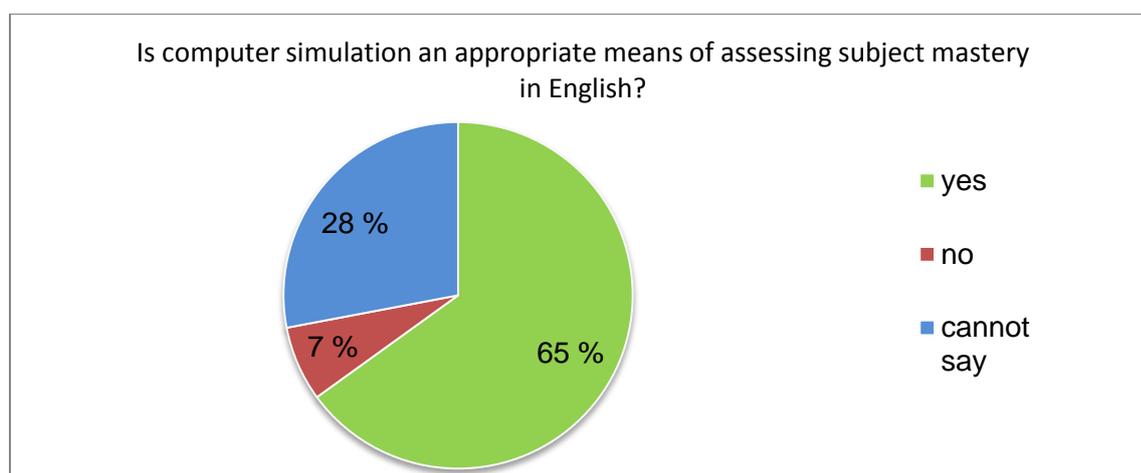


FIGURE 37. Appropriateness of a computer simulation in assessing content in English (pupils)

One out of every ten pupils was opposed to computer simulations as an assessment method. When the two simulations are compared, **the appraisals in the first simulation were slightly more negative than in the second**: 60% vs. 70% ‘yes’ answers, 11% vs. 3% ‘no’ answers and 29% vs. 27% ‘cannot say’ answers.

In both piloting rounds there were severe technical problems that frustrated both me and the children. This misfortune may have affected pupils' appraisals. Regardless of this, the general opinion is in favour of computer simulations as a test type. Some pupils wrote additional comments to explain their choices regarding the appropriateness of computer simulations in content-based language assessment. Among the 'yes' comments were the following pointing out the variety of tasks and the pragmatism of the simulations:

Because different topics were asked in English (girl, 4th grade)

Yes, because in different simulations different subject questions are asked, for example, environmental sciences, physical education... (girl, 5th grade)

This simulation shows how a pupil can use English in a practical situation (girl, 5th grade)

Only two 'no' comments were obtained from the whole group, but they were not very descriptive in explaining their objection.

More fun than on paper. (boy, 5th grade)

Because... I just don't know. (boy, 4th grade)

The 5th graders were piloting their second simulation which can also be seen in their answers; they were starting to grasp the overall idea. Simulation as a test concept fascinated the majority of pupils, for **58% of the simulation 2 testers would opt for simulations rather than paper-and-pencil tests**. Boys were more traditional in this sense than girls; 61% of girls would prefer simulations over paper tests, whereas only 52% of boys concurred. Those who would choose a traditional paper test argued for their familiarity, operational reliability, shorter duration, having the option not to speak and anti-technological attitudes. Among these arguments were the following:

It is just more familiar and you don't have to stress that much (girl, 4th grade)

It is nicer to do it quietly on your own when you don't have to speak. And if you do it on paper, you get time to think longer because in the simulation, you have to answer almost straight away. (girl, 5th grade)

Pupils in favour of computer simulations took the position that simulations are fun and a variation on traditional tests as in the following quotes:

Because taking a test by means of a computer is nice change. (girl, 4th grade)

Because when you have to write a lot, your hand starts to hurt but not with the computer (girl, 4th grade)

It was more relaxed; it doesn't feel like a test. (girl, 5th grade)

The characteristics of alternative testing (see p. 86) are echoed in these quotes. The use of computers was not, surprisingly, familiar to all of the children. One would expect that using a keyboard would not be a challenge for digital natives, as this generation is occasionally referred to; this was not always the case.

Examining the performance

After the simulation experience, the pupils whose performances were successfully saved in the language laboratory (see 5.4.2) could examine and assess their performance stored both in audio and written forms. In the first simulation questionnaire, the issue of self-assessment and the examination of one's own performance were not explicitly addressed as a question, but it was included in the second piloting round. Pupils were asked 'What did you notice of your language proficiency when listening to and looking at our answers?'. Unfortunately, this query could not be answered by all of the pupils, because a substantial number of the TS performances were not saved due to an application problem (see 5.4.2.2). Out of 21 TS 4th grader performances only two were saved, and out of 20 TS 5th graders 15 could not answer this question properly, which is a disastrous result. The frustration surrounding this issue erupted in the questionnaire:

I COULD NOT LISTEN BECAUSE MY SIMULATION DID NOT WORK! (girl, 4th grade, capital letters in the original)

In this sense, the CS pilot was more successful; **pupils' self-assessment notions reveal various strands: improvement in language proficiency, native-like accents, pronunciation issues, areas in which practise is needed as well as disappointment in own linguistic performance or content knowledge**, for example. The quotes below represent such strands.

My voice sounded strange and I noticed even myself that I already have a little bit of a British accent. (boy, 4th grade)

Basic things are well mastered. In practical situations speaking and understanding main points succeeds well. (girl, 5th grade)

My English vocabulary is large, but it contains mainly everyday words. Pronunciation of some words wasn't too successful. (girl, 5th grade)

I can [speak] English at least that much that I can communicate with English-speaking people. (girl, 5th grade)

I was pretty bad ☹️ (girl, 5th grade, had drawn a tear on the frowning face)

A computer simulation may thus be a means that helps pupils in noticing various aspects, especially in their spoken language; it works in a similar way to recordings if pupils get to produce and listen to them. A teachers' assistance may be needed to concentrate pupils' attention on more precise aspects of language so that pupils could gain primarily positive feedback, for example, on content mastery, pronunciation or the use of past tense.

Overall comments

Pupils' overall comments were gathered in both questionnaires using a technique in which they were asked to complete a sentence, the beginning of which was already given. The seven sentence beginnings were designed to tolerate both negative and positive continuations. In the following, I will provide representative examples of pupils' continuations, both positive and negative, and also give classifications of and observations about the answers. **Overall, pupils'**

comments were very positive and they mostly viewed computer simulations as fun, exciting and revealing. The opportunity to speak was especially valued.

a) I think that the simulation was...

- *a nice experience, a bit like a video game*
- *fun*
- *thrilling, but at the end totally WONDERFUL!*
- *well done, but annoying*
- *really challenging*

An overwhelming majority of the comments were positive; nice, fun and interesting were the most used adjectives.

b) Regarding simulations, I would like that ...

- *simulations were improved, e.g. you could choose the level how difficult it is*
- *there were more simulations*
- *they were rare*
- *there were different kinds of simulations*

In respect of this statement, there was a major deviation in wishes; some wanted more simulations, others fewer or none; some easier simulations, others more challenging; some longer, others shorter; some with more opportunities to speak, others with less speaking. The majority, however, would like to experiment with simulations in the future. This again shows how important it is to assess in multiple ways, using a variety of tasks and difficulty levels in order to cater for everyone and to give possibilities to excel in one's own zone of comfort and specialty.

c) My simulation experience was...

- *different than a normal English lesson / exercises in the book*
- *scary*
- *a great way to learn what is progressing well and what isn't*
- *interesting and I found out again how well I speak English in practical situations*

Pupils mostly used similar praising adjectives in this sentence as in the first one: their experience was exciting, fun, distinct and nice. A few pupils mentioned the staring and talking heads; the simulation contained several close-up film clips where the interlocutors (actors) paused and looked at the camera (see Figure 19) as if they were listening to the test taker. These were the instances where the test takers were supposed to react by, for example, replying and saying or explaining something.

d) The best thing in the simulation was...

- *development of my own language proficiency*
- *to get to speak*
- *to (sort of) be there yourself*
- *to listen to English and notice that you understand*

A marked number of pupils noted that the best thing was to actually speak and use the language. Also, the experience of authenticity was mentioned – the children were immersed in the virtual

surroundings and received the sense of being in an English-speaking environment. The frame story of simulation 2 particularly sparked the pupils' imagination; the odd, 'out of context' characters lurking in the school corridor were mentioned in this item a few times. The characters were planted in the script precisely for that purpose, to introduce a game-like element.

e) The most horrid thing in the simulation was...

- *nothing*
- *it was annoying to wait for the time bar to proceed*
- *listening to my own failure*
- *that although you didn't say anything, they would still answer as if you said it right*

Although the setting and frame story were seen as authentic, the simulation concept does not allow for totally authentic reaction – the interlocutors in the simulation do not talk back or answer any questions. Thus, the LangPerform simulation could be described as semi-authentic. The problem of improper response patterns was taken into account in the script phase, and such reactions as 'well done' were avoided. Instead, a line such as 'thank you for your effort' was incorporated.

The pupils identified several flaws and weaknesses in the simulation structure and procedure of which the following were the most notable: lack of proper reaction and answering time (pressure to be fast), duration (either too long or too short), level of difficulty (problems in linguistic or content-related understanding) and technical failures (sound and saving deficiencies). In the simulation 1 questionnaire (Appendix 7), there was a query pertaining to further development of the simulation concept especially for primary pupils. This question elicited a number of suggestions for further improvement which were useful for internal project use and were considered when designing simulations 2 and 3.

f) With the help of simulation I noticed that...

- *I should practise pronunciation*
- *there are also entertaining ways to learn about your skills*
- *I should practise mathematics in English*
- *I need more time to think*

This question was very similar to the one in the second piloting questionnaire pertaining to examining the performance in the language laboratory assessment environment. The results were also partly parallel. The pupils' answers can be categorised into following main themes: more accurate perception of one's language proficiency, need for and focus on further practise, and the level of content mastery in English. The final sentence prompt attempted to elucidate how the pupils would word their simulation experience to their peers.

g) I would tell my friends about the simulation that...

- *it was the best thing of my week*
- *they could try it as well but they shouldn't have too high expectations*
- *it was difficult but fun*
- *it is an amusing adventure in English in which you have to be able to answer everyday questions*

Again, the pupils' comments were primarily very positive. It appeared that they appreciated taking part in such an experiment and they acknowledged the uniqueness of the experiment. However, this in turn may cause the Hawthorne effect, which is discussed in section 8.4.

Coverage of simulation contents at school

Finally, in order to better understand the stance of pupils towards simulations and their simulation experience from the viewpoint of subject instruction, pupils who experimented with simulation 2 were questioned about how extensively the simulation topics were instructed at school, and whether they remembered learning the simulation contents through English at school. As Figure 38 depicts, the **pupils' recollection of the topics studied through English is vague, and not all the topics touched upon in the simulation, as far as pupils remembered, were studied in the TL.** This result was not unexpected, because, as accounted for in section 2.4, the provision of CLIL varies in Finland tremendously and as such the models are not comparable. Therefore, it is practically impossible to create a simulation that would apply for every CLIL classroom syllabus context.

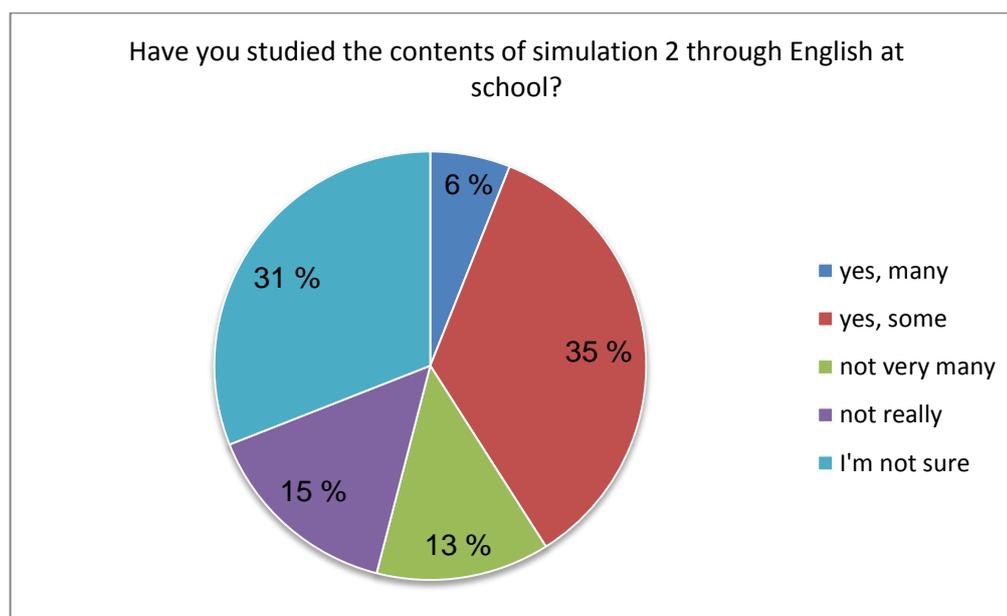


FIGURE 38. Pupils' recollection of simulation topics studied through English at school

The relevant class teachers are the only ones who could provide a definite confirmation of whether or not the topics had been studied through English and to what extent, but the teachers were not participants in this research phase. However, the more pupils are involved as active agents in their own learning, and the more meaningful and tied to the pupils' lives the instruction is, the better the topics will be imprinted in their minds.

7.2 Parents

Parents' were given the possibility to monitor the child's performance at home from their own computer. Their role in the PROFICOM simulations assessment scheme is depicted on page 127 and Figure 16. The amount of data received from parents remained low, 27%, due to the loss of data for various reasons (see 5.4.2.2). It follows from this that the obtained research results are highly tentative. One also has to bear in mind that these results are, similarly to pupils' results, subjective experiences and opinions. I will present the obtained results of the two simulations together when possible. The aspects measured by the parents' are pupils' general language proficiency as estimated by them, the discussions generated by the simulation, appropriateness of computer simulations in CLIL assessment and overall comments.

Pupils' general language proficiency level estimated by parents

Parents were asked to estimate the general language level of their children by using the CEFR scales. It is not known whether the parents (n=39) actually appraised the level based on the simulation, on their own observations made during a longer period of time or both. Figure 39 portrays the estimates made of 4th, 5th and 6th graders' language proficiency levels. **Most parents estimated the proficiency level of their primary-aged children as considerably high (B1)** (Cf. CEFR 'glass ceiling' for YLLs in Little 2007 and Hasselgreen & al. 2011).

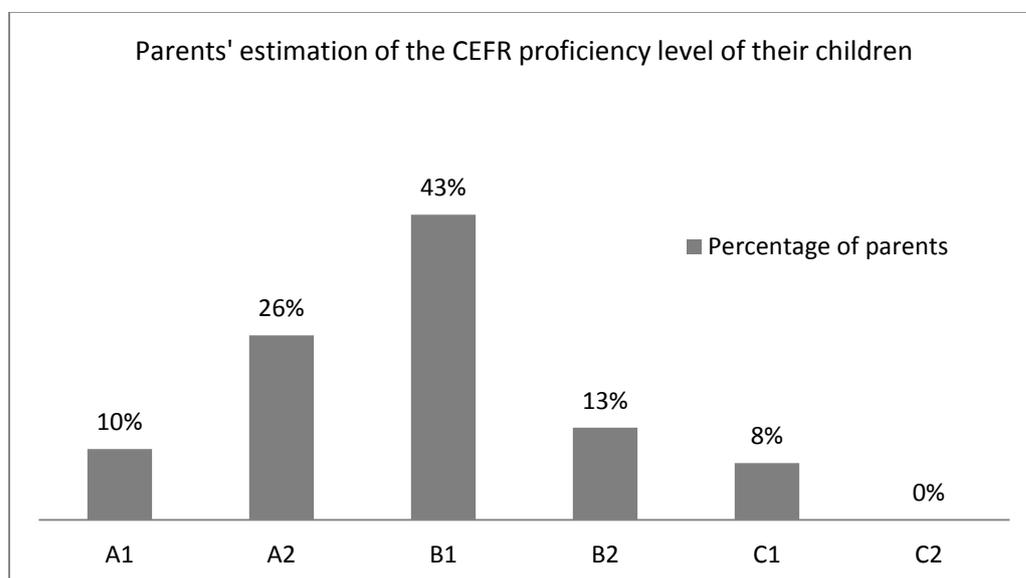


FIGURE 39. Parents' estimate of the CEFR proficiency level of their children

The estimation is, however, considering CLIL pupils, somewhat credible, although in normal EFL circumstances this estimation would be highly unlikely. Assessment in the language laboratory (see 5.4.2) is also based on the CEFR scales, which is why it is necessary to familiarise parents with it. The laboratory was not originally developed for primary pupils and their parents, and therefore, in its current stage, it is not applicable for young language learners.

When the parents were asked about the observations they had made of the English language proficiency of their children, the remarks (n=11) were practically divided in two categories: positive and negative. Underneath is an example of both.

Spoken language appears to be at everyday language level and production sounded natural. The extent of vocabulary surprised, but also limitations were surprises. The emphasis of language study has already been shifted towards vocabulary, but the development should be reinforced.

His/her language proficiency is still very passive. There have not been enough authentic, interactive situations in the foreign language, and it takes practise to be able to react to speech one doesn't quite understand. Also the anxiety in the recording situations will be alleviated through experiences, but this time [that] affected the performance negatively.

In normal circumstances, the simulation concept does not require parents to act as raters. When the teacher or external rater (see Figure 16) assesses the performance, the assessment will be seen in the language laboratory both for the parent and the learner who can, when needed, add their own comments. It is natural that the ability of parents to assess and gauge linguistic performance in the language laboratory is limited, but as for content knowledge demonstrated through the TL, the language laboratory provides default answers (which can be compared to the ones given by the child).

Pupil-parent discussions generated by the simulation

The parents' questionnaire in simulation 2 contained a question asking whether parents examined the language laboratory together with their children. **Nine out of 13** claimed having done so and **reported that they had particularly discussed the following topics: difficulty of tasks, task instructions and recorded speech**. The quotations underneath exemplify some of these topics.

Why the child had answered as s/he had and what perhaps should have been answered.

In child's opinion, it was difficult to understand the questions and react to them quickly. The recording situation made him/her tense up. The child thinks that a simulation is a good means to assess spoken language and comprehension of the foreign language. Studying a book is easier. The child did not understand what to do in different tasks.

The simulation performance may work as a gambit to provide opportunities for fruitful conversations regarding language learning, content learning and school issues between the learner and their parents. It is likely that spoken English, especially subject-specific spoken English produced by the pupil does not manifest itself in domestic situations that often. Simulation performances may thus cater for a new type of development discussion for pupils and their parents. It is also possible to include simulation sections in official development discussions.

As was noticed in connection with the pupil simulation questionnaire, **the simulation performance is not always an enjoyable experience for the child**. One parent reported that their child was embarrassed by his/her performance and another that their child did not wish to return to the simulation because s/he regarded the performance as not good.

Appropriateness of computer simulations in CLIL assessment

Regarding the four basic language skills, the **simulation is most appropriate in measuring speech production** according to 49% of parents (n=35) and **listening comprehension** (40%). Most parents (89%) thought that writing is a skill that a simulation mostly fails to measure. This result, although approaching the four skills from a diverse perspective, is partly in line with what pupils estimated to be easy for them; most pupils stated that receptive skills (listening and reading comprehension) were more natural for them than productive skills (speaking and writing).

A more detailed enquiry into the areas of language proficiency that would be applicable for simulation assessment revealed that, in parents' (n= 35-37) opinion, **computer simulations are most appropriate in assessing fluency of spoken language** (94%), **general language skills** (92%) and **listening comprehension** (92%) as well as **pronunciation** (Figure 40). The least appropriate computer simulation is in gauging orthography and cultural knowledge.

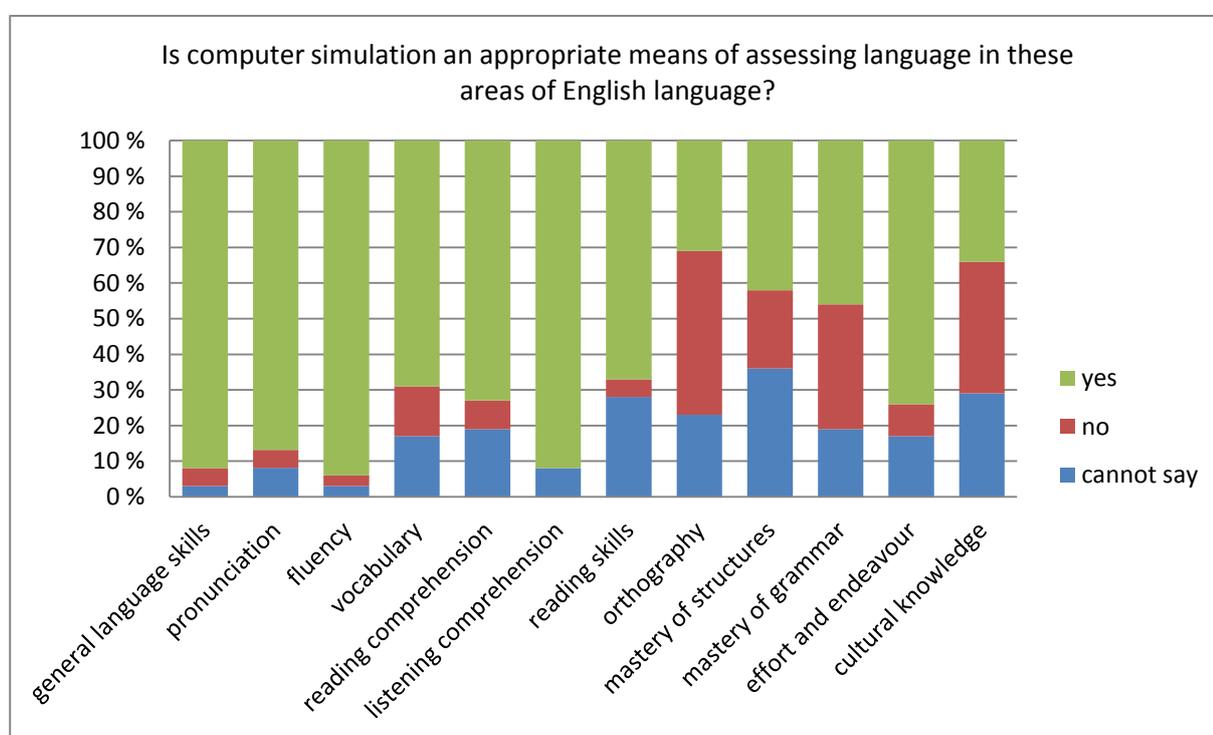


FIGURE 40. Appropriateness of a computer simulation in assessing aspects of language (parents)

The key question, 'Is a computer simulation an appropriate means of assessing content knowledge (e.g. science, P.E.) in English' divided parents (n=35) so that **the majority** (66%) **agreed with it being an appropriate method for assessment of content knowledge through the TL**, 17% of parents disagreed and another 17% were indecisive (Figure 41). One parent had chosen both options, yes and no, and added a comment: "A simulation equals with a verbal test so it emphasises oral skills. In the studying, however, vocabulary and written production are more important". This comment shows that simulation 2 manifested itself to this parent as one that elicits mainly spoken language.

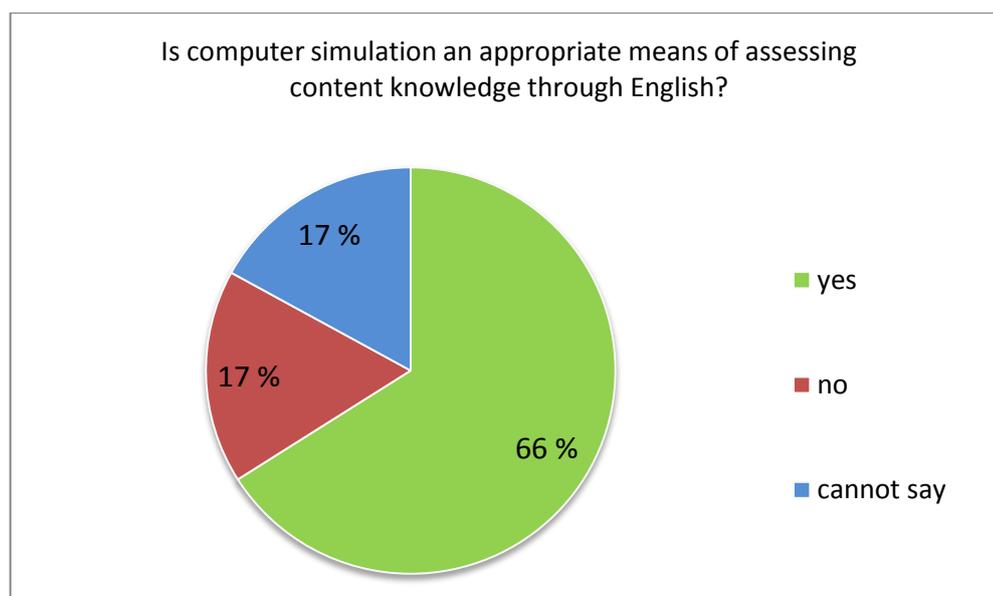


FIGURE 41. Appropriateness of a computer simulation in assessing content in English (parents)

Parents in favour of computer simulation wrote additional notes pointing out, for instance, that the method needs to be established properly and it is also appropriate for practising purposes. None of the parents objecting to computer simulations made a remark or justified their dissenting opinion.

The study thus indicates that **parents, on average, are sympathetic to computer simulations as an assessment method in language learning in general and CLIL specifically**. However, when parents were approached in the simulation 2 questionnaire about whether or not, with the help of the simulation performance, they were able to build some kind of an understanding of their child's mastery of contents in English, five out of 13 replied that they had been able to do so, six that they had not and two chose the option 'cannot say'. One parent had underlined the words 'subject-related issues' and added the comment "It appears that s/he doesn't master any kind of vocabulary yet...".

When analysing the questionnaires, I started to question whether the concept of CLIL as an approach to language learning was quite crystallised for all parents. In both questionnaire rounds, there were parents who wrote annotations or footnotes indicating either specifiers or disapproval of the simulation qualities as well as the wording or phrasing of the questionnaire. Such comments included. 'audio missing – relevance?' in connection with pronunciation, 'narrow subject matter' regarding vocabulary or 'What is meant by this?' referring to reading skills. Some of these comments are certainly justified, as will be discussed in 8.4. One parent repeatedly stressed expressions referring to subject-specific English by underlining them (e.g. 'English used in other subjects than English') and making remarks which implied that CLIL is confused or associated with EFL instruction, this is English as a school subject. Such behaviour reveals that **the subject-specificity in CLIL is not known to all CLIL parents**.

Overall comments

Further notes were elicited by using the same sentence completion technique as with pupils, although the pupils had seven sentence prompts, but the parents only five. However, the sentence openings conveyed the same meaning as those for pupils. The following quotes feature typical examples of parents' sentence endings gathered from both questionnaires 1 and 2.

a) In my opinion, the simulation was...

- *a new and different method to assess language proficiency. When functional, it enables an increase in the communication between home and school, but still technically clumsy.*
- *an interesting approach. As for my own child, I guess that the new situation caused anxiety. I didn't quite understand all the questions in the simulation.*
- *had been further developed, but gave a partial picture still. Familiarised quickly with, the big picture remained unclear.*
- *well implemented and realistic*
- *a good opportunity. It mediated the level of English proficiency X [pupil's name] has reached so far.*

'Interesting' was the most frequent word used in this item; also adjectives such as 'useful' and 'fun' were often filled in. Overall, the parents' stance was cautiously positive. In comparison with pupils' overtly praising characterisations, parents were more analytic.

b) Regarding simulations, I would like that...

- *they would be utilised more in school.*
- *they would be established. It would be interesting to see the results, say, after a year (revisited).*
- *the simulation [performance] could be seen as a video recording. It remained unclear whether part of the questions were posed out loud or were all in textual form.*
- *they would be developed to cater more for the children themselves.*
- *they would be utilised in the instruction as well and that pupils could compare their production to a model (pronunciation).*

A substantial majority of parents expressed that there should be more simulations more frequently. As the quotes portray, many parents suggested alterations and improvements. Some of their comments also revealed that the familiarisation at home had occasionally been superficial because the suggestions concerned functions or aspects that were already inbuilt in the simulation environment. This alludes to the idea that the assessment approach requires considerable familiarisation from parents; it is time-consuming and entails concentration. Parents should also be aware of the CLIL topics and syllabi in order to form an understanding of content-related performance. One needs to remember, however, that in normal circumstances, parents would be involved only after the teacher assessment phase in which case the teacher's rating would already be available from the language laboratory and serve as a reference for parents and the pupil (see Figure 16).

c) The advantage of a simulation is...

- *that the pupil needs to use the foreign language ex tempore, as it is often in real life as well*
- *its truthfulness and multifaceted possibilities for self-assessment and developmental planning*
- *its motivating nature and stress on own speech production; listening comprehension is brought into a authentic situation*
- *in multimedia and audio files*
- *that you can listen to it at home and get information on a child's oral language proficiency*

The most often mentioned advantages were the elicitation of spoken English and how the simulation mediates both assessment data and information between the parties of assessment. Parents had also captured the interactive and activating essence in the simulation as well as authenticity and correspondence to real-life-like circumstances.

d) The disadvantage of a simulation is...

- *this novelty and that pupils may be nervous about the situation itself and perhaps understanding instructions may be troublesome in the beginning*
- *time limits in the replies; not understanding one thing may ruin the whole task*
- *artificial interaction, because the interlocutor cannot react to what you say*
- *that the pupil may become distressed if they don't understand and they know that they will be assessed.*
- *none*

Quite many parents had left this item blank, drawn a dash or written 'no disadvantages'. A fair number of parents also mentioned technical setbacks (e.g. sound problems and system meltdowns) as well as the rush caused by time limits. One parent noted that simulations are disconnected from other schoolwork, which again reveals that the subject-specific nature of CLIL is unfamiliar. Affective factors, i.e. test anxiety issues caused by assessment, lack of understanding or insufficient language proficiency, were rarely mentioned.

e) With the help of simulation I noticed that...

- *his/her English proficiency is much better than mine at the same age (12 yrs)*
- *s/he pronounces well but makes rather many grammatical mistakes*
- *it is surprisingly poor. Pronunciation was ok, and s/he had the courage to speak, but basic grammar was lost and the scarcity of vocabulary surprised me*
- *based on this, s/he wouldn't know anything. And that isn't true.*
- *this a splendid way to monitor the development of child's language proficiency. S/he understands English well and is able to reply, but needs time for that*
- *written production using the keyboard is not good*
- *the language proficiency is surprisingly good!*

The computer simulation provided most parents with positive discoveries of the language proficiency of their children or the development of it. Some parents imagined that the language use or content knowledge of their children is not equivalent to the perception they most likely

obtain from paper-and-pencil tests and EFL grades. Several parents foregrounded how the simulation laid bare the spoken language which they do not get to monitor in normal home circumstances.

SUMMARY

LangPerform computer simulations were perceived by both pupils and parents as an appropriate means of assessing general language proficiency, listening comprehension and fluency – parents even more so than pupils who were, in this respect, more indecisive. Similarly, a slight majority of both pupils (65%) and parents (66%) concluded that computer simulation is an appropriate assessment method in CLIL, i.e. assessing content knowledge through English. An impressive majority (93%) of pupils considered that they managed moderately, well or very well in the simulations; boys assessed their performance more positively than girls.

Pupils tended to view BICS-type tasks including communicative, everyday English far less demanding than CALP tasks requiring subject-specific language. The opinions, thoughts and experiences of the participants varied and were polarised. The simulations were criticised for, among other things, causing test anxiety, being artificial in the interaction and not leaving enough time for reaction. Technical failures characterised or even hampered the performance of a significant number of pupils. Simulations were approved of for providing possibilities to practise and show abilities in spoken language, authenticity and bridging the assessment information gap between home and school.

The simulation questionnaires also provided information on how the nature of CLIL is not clear for all parents. It appears that CLIL is associated with the study of English as a school subject. Simulations as an assessment concept appears to be functional and attractive, but, in order to become fully in use, it requires more familiarisation from both pupils and parents.

8 DISCUSSION

The purpose of this study was to form an overall understanding of how language assessment in primary CLIL instruction is organised in Finland; the more specific aims were to look into the assessment practices and methods of CLIL class teachers, identify problem areas in CLIL assessment and verify whether or not the assessment practices are adequate enough for the stakeholders. Furthermore, computer simulations as an alternative assessment method were experimented with to examine the affordances of technology in CLIL assessment and to orient towards future assessments.

This study shows that assessment of primary pupils' English language proficiency in CLIL is not an established practice. The key results provide convincing evidence that in the Finnish CLIL field, there is a need for an elaboration of CLIL fundamentals, a stabilisation of language assessment practices and awareness raising of the linguistic objectives in CLIL and CLIL characteristics for all stakeholders of assessment: teachers, pupils and parents. This investigation suggests that language assessment is not always well-grounded, evidence-based and adequate enough for meeting the expectations of pupils and parents. Comparable results have also been obtained in other SLA contexts (e.g. Gattullo 2000; Hunt 2009). Assessment research in the disciplinary area of CLIL is rare, as I have pointed out, and consequently it is difficult to place these findings into the context of other research. Therefore, when applicable, I will rely on recent findings in CLIL, SLA and second language assessment research in my discussion.

During this research process, it became even increasingly obvious that assessment does not occur in a vacuum; it is always closely linked to the objectives of learning (curriculum) and actual instruction. Assessment is the mirror in which learning objectives and instructional strategies are reflected. Figure 42, modified from Miramontes, Nadeau and Commins (2011, 79), depicts the desired state of the total alignment of curriculum, instruction and assessment in CLIL as well as their interaction. It places the curriculum centrally and assessment as the overarching, encompassing feature in education. Therefore, assessment is an integral, inseparable and interwoven part of any instruction – including bilingual instruction (see Figure 10).

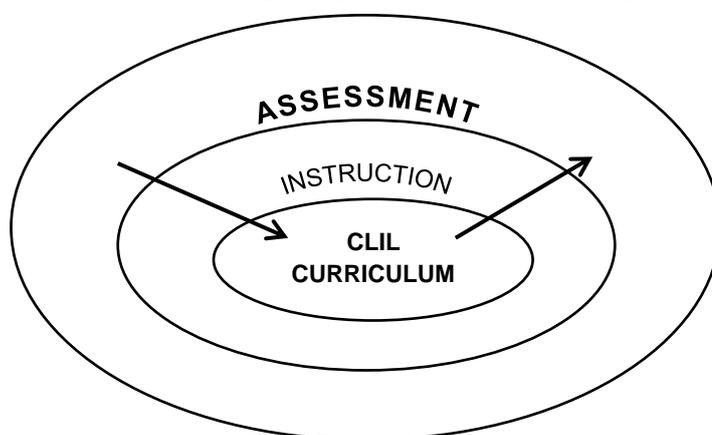


FIGURE 42. Desired alignment of curriculum, instruction and assessment in CLIL

In conducting this research, and from the data obtained, I was thus lead to the fundamentals of CLIL assessment. As a result, I am not able to fully address assessment in CLIL without taking the other elements, especially curriculum, into consideration. Therefore, I have to begin this discussion by foregrounding the fundamentals that provide the basis for solid CLIL assessment: CLIL curriculum and objectives, the lack of which was obtrusive in the data. After contemplating all relevant fundamentals and challenges of, or obstacles to, assessment in CLIL in section 8.1, I will proceed to the assessment methods and practices and their adequacy in 8.2, after which the computer simulations are discussed and evaluated in 8.3. Section 8.4 is concerned with the validity and significance of this study, and I will conclude this research report by proposing an assessment scheme for primary CLIL in 8.5.

8.1 Fundamentals and central issues in CLIL assessment

This section deals with the most crucial issues that lay foundations to assessment in CLIL: curriculum and language objectives. In addition to these fundamentals, which deserve profound contemplation, I will discuss the CLIL knowledge base, the 2016 curriculum reform and teachers' perceptions of language assessment in CLIL. The role of English and the persistence of views supporting implicit language acquisition are important issues that affect both CLIL instruction and language assessment. I will argue for a focus on form approach and suggest an angle of CLIL English which is far more extensive than may be conventional. I will also address the balance of content–language integration and touch upon teachers' varying assessment approaches. The qualifications and language proficiency of class teachers are potentially relevant to language teaching and assessment in CLIL; therefore, those issues will be similarly addressed. Finally, I will justify the need for a mutual assessment scheme.

CLIL curriculum and language objectives

Premeditated curricular objectives are the benchmarks and roadmaps of instruction and assessment, without which there is no direction or reference point. It was very alarming to discover that there are CLIL-providing schools without no CLIL curriculum, linguistic objectives for CLIL instruction or specified content, as the current Finnish National Core Curriculum for Basic Education (NCC 2004, 270–271) necessitates. According to the NCC draft (2014) for the NCC 2016 reform, these prerequisites will not be altered or attenuated. The local-level CLIL curriculum, designed within the frames of the NCC, works as a path finder for teachers, pupils and parents; it is easier to navigate when the route is clear and the destination known.

In a few schools in my data, curriculum issues were attended to as stipulated, but none of the primary schools of the interviewed teachers had defined the desired language *level* in CLIL. This result was replicated in an Internet search. It seems that schools are cautious about setting

linguistic objectives – perhaps in order not to promise too much to parents, perhaps out of ignorance. This caution or ignorance is reflected in the total absence of a CLIL curriculum. After all, the basic objective of bilingual instruction is to cater for “firmer language proficiency than in lessons reserved for language in normal instruction” (NCC 2004, 270). It is noteworthy, however, that language objectives are not absolute but relative goals; they represent educational aspirations framed by the NCC.

In the NCC draft (2014), the desired language level is no longer mentioned, but the language objectives must still be determined, and the CEFR scales are recommended for that. According to the current NCC (2004, 270–271), the linguistic CLIL objectives should be defined in the four basic language skills as well as cultural skills, after which the education provider is to specify the core contents and subject areas that should be taught through the foreign language, – in correspondence to the linguistic objectives. The starting point in CLIL is, thus, language. This is a matter that has also gained attention elsewhere, as Dalton-Puffer (2007a, 295) has so aptly queried “But why should we be doing CLIL at all if there are no language goals present?” Indeed, we should not.

These specifications, language objectives and predefined bilingual contents, should set the benchmarks for either integrated or discrete language assessment in CLIL. The first condition for enabling teachers to carry out assessment in CLIL according to the prerequisites of the NCC is to make adequate CLIL curricula available for them. This in turn is a task of educational decision makers, head teachers in schools and chief education officers in the municipalities who should initiate and facilitate the design process. An existing CLIL curriculum only can be adopted and used as a tool by CLIL teachers.

Some of the interviewed CLIL class teachers implied that they were not aware of the curricular preconditions set by the NCC. This suggests that the core curriculum per se may not be a tool for teachers, which is in line with the study of Luukka, Pöyhönen, Huhta, Taalas, Tarnanen and Keränen (2008, 68-69) according to which teachers of foreign languages value their own experiences and principles more as a reference in instructional objective setting than the NCC or the local curriculum. A similar conclusion was reached by Hüttner, Dalton-Puffer and Smit (2013) in Austrian tertiary CLIL: in the absence of “language management”, classroom CLIL is driven by teachers’ beliefs of language learning.

CLIL knowledge base

The lack of objectives and proper curriculum was also reflected in the parents’ questionnaire and interview data. On the whole, the parents in the two research schools seemed to be pleased with bilingual content instruction, but too many parents appeared to be ignorant of the philosophy of CLIL and seemed to associate CLIL study with EFL. Parents disclosed relatively often that they were not very familiar with CLIL–provision and that the aims of CLIL were somewhat unclear. Although both schools arrange information events for those considering applying for CLIL, as well as thematic happenings and parental evenings for those already in the programme, CLIL issues

do not seem to come across clearly. It is understandable that not everyone is able to attend these gatherings, and that those who do, will not always grasp the true nature of CLIL, because pupils' presentations often (and I am not claiming that this is wrong) represent more casual, EFL-type English.

Therefore, parents' weak CLIL knowledge base needs to be addressed. The first remedy is to ensure that a proper curriculum with language objectives and TL contents is available for parents and pupils; the curriculum should be clear, interpreted and written in general terms. The CLIL learning path or map should be evident for both pupils and their parents. Furthermore, it is appropriate to state the CLIL objectives of the given year during each parental evening and perhaps send them home in a short, written form. In addition to this, it might be worthwhile to invite parents into classrooms to observe plain, everyday CLIL lessons and let them see how language is used by teachers and pupils in any school subject. One central notion to be made clear is that CLIL study is not similar to the study of English as a school subject. I will return to this later in this chapter. Shortly put, we should help parents realise what CLIL means in practice, especially because it is manifested in different, incomparable ways in different schools. When national requirements and CLIL objectives are transparent, the assessment of their achievement and progress in language is more tangible for all parties involved, and language assessment will not be questioned.

Core Curriculum reform 2016

The NCC reform process will certainly draw attention to curricular issues in CLIL-providing schools. It is one remarkable step in increasing curricular ownership and partnership. The launch of the renewed National Core Curriculum for Basic Education will take place in 2016.¹⁶ The process of drafting the new NCC 2016 is unlike any prior curriculum reform in Finland; it is public, participatory, and highly future-oriented, and it takes place in multiple stages. The first curriculum outlines have already been published in 2012. Various interest groups and stakeholders, teachers, pupils and their parents, as well as lay citizens, have had a possibility to publicly comment on this outline.

Invited expert panels have drafted the texts of various chapters which have also been reviewed openly during spring 2014. The undertaking of the CLIL (bilingual instruction) panel has been to define the purpose and task of various programmes of instruction in a foreign language (immersion, CLIL and language-enriched programmes) as well as the distinguishing features of pedagogical practices and assessment. The NCC 2016 will thus be co-constructed through three diverse rounds of comments, which should increase the sense of ownership of everyone concerned. At its best, the new NCC reflects the educational will of the whole nation. On basis of

¹⁶ Core information on the reform in English available from http://www.oph.fi/english/education_development/current_reforms/curriculum_reform_2016

the final NCC 2016, which will be published by the end of 2014, the local curriculum work will start in 2015.

The forthcoming, renewed NCC 2016 is improved in that it will, more explicitly than the current NCC (2004), list specifications of issues to be decided upon and dealt with at the local level. There are no utterly drastic changes in the NCC draft (2014) regarding bilingual education – it merely itemises issues to be noticed. The draft does not mention the four language skills any more, and nor does it directly compare the language skills acquired in bilingual education to EFL instruction. However, it emphasises subject-specific, academic language use and highlights the parents' role in supporting the language development in both languages of instruction.

The tools originating from EU policies, the CEFR and the ELP are named as helpful in objective setting and assessment, thus reinforcing the interconnection between Finnish and European language education. Language assessment in bilingual education is more markedly brought into the fore in the new draft; monitoring the growth of the TL in diverse subjects is highlighted. This is a significant change in comparison to the current NCC. The draft does not, however, give any recommendations on how this assessment information should be mediated to the other stakeholders, but it gives instructions on how final assessment at the end of basic education should be addressed in order to ensure pupil equality when applying for secondary education. The new core curriculum will hopefully steer more consistent local CLIL curriculum work and elucidate the foundations of both instruction and assessment in CLIL.

Teacher perceptions of language assessment in CLIL

As already mentioned above, according to research, teachers' own experiences and ideas, even course books, inform planning more strongly than curricula (Luukka & al. 2008). The same principle also appears to apply to assessment in CLL. It was discovered in this research that some teachers carry out language assessment regularly, while many teachers do not gauge or gather evidence of pupils' language proficiency and its progress systematically, if at all.

One reason for the diverse stances towards assessment in CLIL, in addition to individual beliefs in language teaching, learning and assessment, is most likely that the CLIL curriculum guiding the pedagogical practices does not exist in every CLIL establishment, which leads to eclectic practices. One teacher described assessment in CLIL as "agonising" due to a lack of foundation. CLIL assessment seems to evoke confusion and even distress in teachers. They are unclear what to assess (to what extent assess language, which content through the TL), when, how and why. Moreover, they are unsure what the roles of continuous, formative and summative assessment are in basic education.

Nevertheless, one point is clear: the current NCC (2004, 273, boldfacing mine) states indisputably that assessment of language proficiency within CLIL must take place:

*[A]ssessment **must** give the teacher, pupil, and parents or guardians **adequate information** about the pupils' language proficiency **in relation to the objectives.***

The NCC (ibid.) also stresses that the “growth in comprehension of a foreign or immersion language” needs to be monitored closely, especially in the beginning years of CLIL. This statement refers to grades 1–3. In the participating teachers, the group of grade 1–2 teachers was especially prominent. At the first glance, one might think that this factor has most likely affected the obtained results. However, most of the teachers in this group also considered language assessment in CLIL rather, very or highly important, but in practice, many of them emphasised other than linguistic issues in assessment. The distribution of teachers who were not collecting assessment information was even across all grades. Thus, it is rather the focus, extent and methods of assessment that are affected by the grade level.

Although teachers seem to perceive language assessment as important, this perception is not always reflected in their actual assessment practices. Language observations are not necessarily translated into actual feedback, as this study shows. One teacher explained in the theme interview that the establishment of CLIL in Finland has taken an enormous amount of pioneering work which was started from scratch, and while energy was placed in creating instructional practices and stabilising CLIL as a system, assessment issues have been minimised or ignored. This explanation is plausible, and is what Marsh (2013) refers when discussing the CLIL trajectory having not yet reached its peak (see also 2.3 on the three decades of Finnish CLIL). The curriculum reform will most likely unify teachers’ perceptions of assessment in CLIL.

The role of English in CLIL and its connection to assessment

Another fundamental issue having an impact on how assessment of language proficiency in CLIL is either accepted or rejected is the distinction between teachers’ perceptions of the role of English in CLIL. In this study, three different teacher foci on language were detected: instrumental, dual and eclectic. It appears that teachers who see language purely as an instrument tend to be reluctant to assess it, because, as they justify the perception, language is “just a tool” for learning, and they do not wish to let language overpower content instruction. This reflects the controversy in the CLIL field, but this ‘language is just a tool’ view is nowadays often – but not always – seen as outdated. The belief in purely implicit language acquisition (see Tella 1999 and Dalton-Puffer & Smit 2007) most likely has its origins in the first decade of Finnish CLIL when the system was adapted from the Canadian immersion which, in turn, is based on ideas of implicit language acquisition. Furthermore, the early CLIL research leaned on the immersion research that was supported by Krashen’s Input Hypothesis. For this reason, it is very important for CLIL researchers to be very aware of the ideological background of the studies they are drawing from. In sum, an instrumental focus, i.e. inclination towards implicit learning, seems to be a relic of early CLIL in Finland still persisting today.

When making comparisons between different models of bilingual education, most notably immersion and content-based instruction, it is also important to realise that the conditions in Finland are totally different from Canada or the United States (see 2.2). In these countries, the learners are immersed in the target language both in and outside school. The TL is thus the

language of the surrounding community and schooling, whereas in Finland this is not the case. Although the status of English is strong in Finland,¹⁷ it is not a publically used language – nor is it the language of schooling on any large scale. Therefore, it is not reasonable to expect similar or parallel results of CLIL if most of the instruction and work in the classroom occurs in the language of schooling (Finnish), as the input is not very strong. In order for natural language acquisition to take place, English needs to be substantially present in the classroom, with the learners as active producers and users of the language rather than passive recipients. In such cases, CLIL contributes to active bilingualism rather than passive bilingualism (see Ricci Garotti 2007 and Cf. Table 10).

I claim that the input in Finnish CLIL, in many cases, is not extensive enough to support implicit language acquisition. The instrumental focus on language thus deserves to be updated to a more contemporary dual focus, which not only accepts but also emphasises the role of language both as the medium and object of instruction – a basic principle also stated in the NCC (2004, 270). This dual focus is the national CLIL norm; it is acceptable and even recommendable to support language acquisition by teaching language forms necessary for more accurate content learning and knowledge manifestation.

An eclectic focus on CLIL underlines affective factors, which are naturally also important in language acquisition, as was shown in the theoretical part of this study, but they should not be the only focus in language assessment. Nevertheless, it is important to consider highlighting affective factors along with linguistic issues which need to be taken into account in relation to the linguistic objectives stated in the curriculum. The courage to use language, the joy of using language for a real purpose and collaborative construction of content meaning are all factors that should be taken into consideration when assessing language proficiency in CLIL. Positive feedback on effort, motivation and attitude, for example, may boost the linguistic self-concept of learners whose linguistic performance is somehow deficient or who are still in the early stages of their learner language development.

The finding of teachers' having different language foci in CLIL is partially in line with the current findings of Hüttner, Dalton-Puffer and Smit (2013), who discovered beliefs in both incidental and explicit language learning in Austrian tertiary CLIL teachers, although explicit learning was only mentioned in connection with vocabulary acquisition. The three foci detected in this study also corroborate the results obtained by Serragiotto (2007) in Italy, where teachers who were team-teaching secondary CLIL content and language assessed 1) both language and content, 2) primarily content or 3) primarily language, depending on their own educational background. Language teachers tended to focus more on language issues, whereas content teachers on content. Following this logic, it is unsurprising that many CLIL class teachers,

¹⁷ See e.g. the National Survey on the English language in Finland: Uses, meanings and attitudes (2011) in <http://www.helsinki.fi/varieng/series/volumes/05/>. The study was conducted in the University of Jyväskylä.

originally trained as content teachers, are inclined to reject the idea of language teaching and therefore also language assessment.

From implicit language learning to focus on form (dual focus)

We need to remember that the slogan of CLIL is to get two for the price of one, so the language aspect carries heavy weighting in CLIL - there is no CLIL without the language aspect. The foreign language must be given added value also because the majority of parents and pupils, as the background questions in this study showed, have opted for CLIL due to the possibility to learn more English and of higher quality. If the TL exposure is not marked and the language learners are not immersed in the TL environment, as in the lower end of the bilingual continuum (see Figure 2), the means to guarantee the progress in academic language development have to be questioned. If the overall goal in CLIL is active bilingualism, then a focus on form supporting the growth of academic English is particularly needed.

In order for the language to become a functional tool for the learners, they have to have construction blocks (lexis, phrases, clauses, sentences, collocations, cohesive rules, connectors, etc.) and models of how to use them successfully in academic contexts, including studying at school. There is ample research evidence showing that focus on form enhances language acquisition (e.g. Ellis, Loewen & Erlam 2006, see also section 2.3), which is another reason why CLIL instruction, and also its assessment, should display a dual focus. CLIL lessons should certainly not become pure EFL (focus on forms) lessons, but the current view is that focus on form, instead of implicit language acquisition, is a desired, advantageous praxis in CLIL classrooms. Linguistic scaffolding is not equivalent to explicit language teaching, and neither does placing more explicit emphasis on language, i.e. putting focus on form, undermine the importance of content learning.

CLIL permits focus on form situations which include context-embedded “language flashes”, as one teacher interviewee described these focus on form moments. I argue for a more competent, balanced and aware language instruction within CLIL. When pupils learn to use the target language better, assessment or documentation of that language use, language proficiency, becomes a more natural process. The fear that CLIL lessons would turn into language lessons is ungrounded, because English as a foreign language is still a separate subject supporting the learning of casual, everyday English. However, the question of how to proceed from EFL to subject-specific language still remains unanswered. I encourage investigation into the nature of subject-specific, academic language which is the language needed in CLIL study – and the language to be assessed in CLIL. Academic language in CLIL is thus a subject for further investigation.

CLIL English = EFL + subject-specific academic English

If EFL is mostly teaching *about* the language and learning to communicate in everyday situations, then CLIL primarily is teaching and learning how to *use* the language in subject study. Considering instruction and assessment in CLIL, it is essential to realise that CLIL teaching and learning requires subject-specific, academic-type English which is different from EFL-type English (see 3.3) which in turn rather represents, at least at primary level, casual, everyday English. For this reason, I doubt whether it is appropriate to treat these two types of English as one, as it seems to occur. It is obvious that in the first years of CLIL, the emphasis is on casual English and the accumulation of general vocabulary, but later on, as the language needed for study becomes increasingly subject-specific, EFL-type English and English in CLIL start to diverge from each other.

Accordingly, I see that these two are two sides of the same English language coin, and following from this, I argue that these two sides require different assessment practices. In mathematics, the pupils should learn to ‘speak mathematics’, whereas in biology they work with scientific genres. The need for subject-specific language in schooling has been increasingly acknowledged in general education. Every teacher should contribute to the pupils’ language development, in particular those who are not studying through their mother tongue. The same principle applies to CLIL: every CLIL teacher is a language teacher who should help pupils to accumulate their language proficiency – especially the academic proficiency needed for studying subjects other than languages. This trend is also visible in the NCC reform draft which emphasises language awareness and multiliteracy. The NCC draft (2014, 14, my translation, my emphasis) states:

*The multiliteracy of pupils is **cultivated from casual language towards mastery of language in diverse disciplines** and manners of representation. Development [of multiliteracy] requires rich a textual environment, **pedagogy that makes use of it** and collaboration between different school subjects.*

Subject-specific language is also mentioned in the draft chapter concerning bilingual education.

Academic language proficiency consists of knowledge of academic linguistic features, knowledge of subject matter, learning strategies as well as sociocultural and psychological components (Cf. Krashen & Brown 2007; Scarcella 2003; Snow & Uccelli 2009). In order to teach and assess through subject-specific language, CLIL teachers should be aware of such subject-specific language features, different genres and registers. Moreover, knowledge of language functions, chunks, phrases and vocabulary are essential in recognising features of subject-specific language. In order for pupils to demonstrate their content knowledge through the TL (also in assessment contexts), they have to be taught to interpret and use subject-specific language. Especially understanding and producing longer stretches of academic language requires more than mere vocabulary building. Various projects (e.g. ConCLIL 2013; ECML 2013; eCALLMS 2013) that concentrate on elevating language and literary development of bi- or multilingual students, in addition to CLIL teachers’ linguistic studies, are of use in achieving this.

My view is that CLIL English is a conglomeration of English in EFL and all the different, subject-specific academic English that the learner is exposed to in the school environment. The English encountered in extramural contexts naturally has an impact on the language proficiency demonstrated at school. Figure 43 illustrates how CLIL English is the sum of its constituent parts and therefore substantially more than EFL alone, which is why I strongly argue for detaching EFL and subject-specific English in assessment practices.

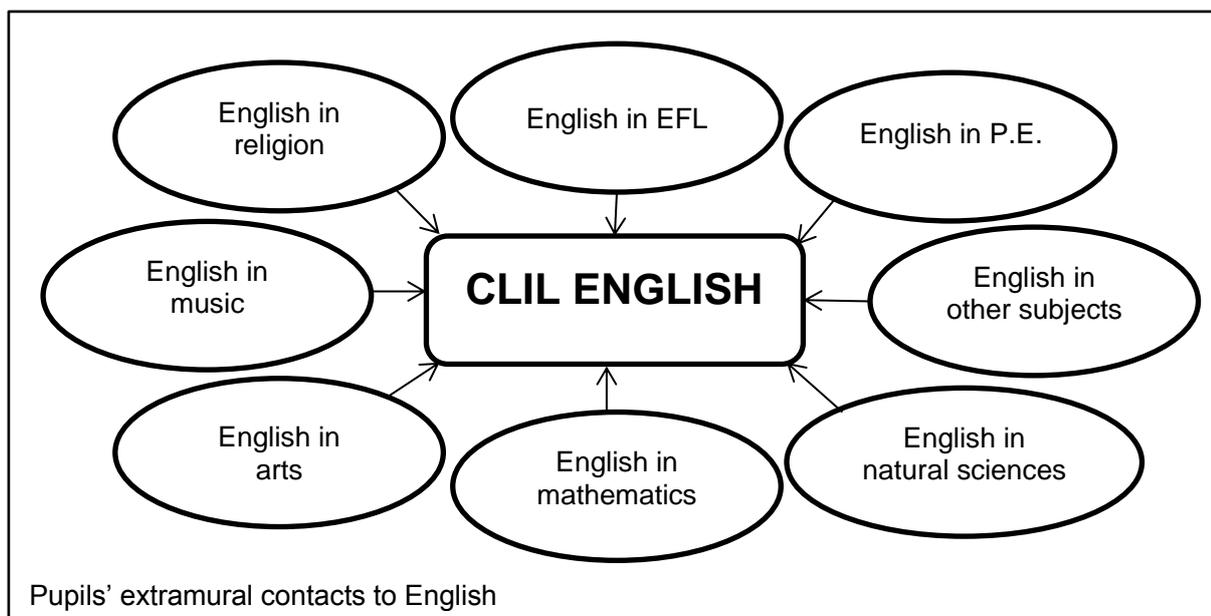


FIGURE 43. CLIL English as the sum of its constituent parts (adapted from Wewer 2013c, 1)

This view of CLIL English is completely contrary to the understanding of Austrian tertiary teachers and students in a technical university; they perceived CLIL English as an extension to EFL (Hüttner & al. 2013, 277). In addition, my data contained a number of indications of perceptions according to which I am able to conclude that EFL is often perceived as synonymous to English in CLIL. If CLIL pursues EFL-type English only, then subject-specific language needed for content study does not develop. Such instances were also seen in the computer simulation experiments. Some pupils became frustrated due to their inability to express content knowledge in English while performing the subject-specific simulation tasks. As one 5th-grade girl noted of her lexis with the help of simulation: “My English vocabulary is large, but it contains mainly everyday words”. Such vocabulary is not helpful in expressing content knowledge precisely.

The level of integration

One bias in CLIL assessment is the ‘rivalry’ between content and language, which is reflected in the level of integration and the balance between them. Many CLIL scholars admit that content comes first: the primary aim is to learn content. This principle is also emphasised in the NCC draft (2014): regardless of language of instruction, the content objectives stated in the curriculum must

be achieved. According to the dual focus, CLIL is also about learning the target language – otherwise we would not have any need for CLIL. The level of integration needs to be specified before the onset of CLIL to erase all possible inconsistencies and to clarify the role of language in respect to content – again, an issue to be addressed in the curriculum and one that should be transmitted to parents. Pupils should also be aware of the role and extent of TL in the content instruction. Awareness of such matters helps pupils and parents to relate to the teacher's pedagogic solutions and anticipate the presence of the TL in studying as well as in the assessment of content knowledge through the TL. Parents reported inconsistency in the TL integration during both study and assessment. The emergence of English within a pupil's study and assessment should be clearly grounded in the curriculum.

Labelling bilingual instruction as more extensive or minor, as proposed in the NCC draft (2014), is a welcomed corrective to the current core curriculum: programmes with less than 25% English-medium instruction are defined as language-enriched, while those with 25% or more content instruction through the TL could be defined as CLIL-type provision. Such division compels schools and municipalities to define their bilingual instruction with more precision and it also makes the level of TL integration markedly more transparent to parents and pupils as well as teachers, who are obligated to conform to the curricular specifications.

The overly strong a status of language in relation to content in CLIL may also provoke counter reactions. In the annual CLIL conference organised in Poland in 2013, a few CLIL scholars expressed their concerns over how language has started to drive CLIL and content issues have been ceded. I would rather voice the opposite concern regarding Finnish primary CLIL: this study is indicative of the fact that the language aspect may not be as integrated in CLIL as it could or should be, – which in turn has implications for language assessment in CLIL

In CLIL literature, integrated curricula (Marsh 2013, 137) or language curricula for CLIL instruction (Dalton-puffer 2007a, 295) have been suggested to reinforce the status of language; counterbalanced instruction was proposed in immersion contexts (Lyster 2011). I see many advantages in such curricula. They would clarify the status and weight of the languages in CLIL and offer reference points for language assessment. Integrated curricula would also help to build a structured, consistent learning path from the beginning of CLIL to upper levels of bilingual education. When teachers, pupils and parents are aware of the linguistic learning path, in particular what has been learned before and what will be covered in the future, the quality of CLIL will be improved. The documentation of pupils' progress in their individual learning paths becomes easier when teachers know what kind of evidence to look for.

I would also like to explore the synergies between EFL and CLIL language curricula: they would underpin each other better if they were more tightly linked and interrelated. The order of morpheme acquisition, for instance, is well known through SLA studies. This should be taken into consideration in CLIL study in order to capitalise on pupils' sensitive periods. Additionally, CLIL pupils often already grasp the basics of typical EFL syllabi (e.g. names of colours, numbers and animals as well telling about oneself, school and family) during the first two grades – although,

depending on the type of CLIL provision, often in spoken form only. When EFL officially begins in the 3rd grade, these syllabi are essentially repeated instead of offering more challenging and new content, collocations, phrases and bridging the EFL syllabi to content study or vice versa. Similarly, because EFL instantaneously goes to the sentence level, that sentence level should be introduced in CLIL as well right from the beginning in order to aim for more than mere vocabulary expansion level. The EFL curriculum could be structured totally differently for CLIL pupils than for learners in ordinary classes.

Varying assessment approaches

It was discovered in this study that assessment approaches also vary: sometimes CLIL pupils' English proficiency was, in the school year report, rated in the subject grade, sometimes in the EFL grade, sometimes it was ignored. This result resonates with McMillan (2003), who demonstrated how teachers' own beliefs and values often conflict with external factors (e.g. parents' expectations, educational policies) in assessment. Also Hill and McNamara (2011) provide such evidence. The EFL teacher is expected to assess language according to the EFL criteria; CLIL teacher should resort to the criteria or objectives specifically created for CLIL. We are returning to the curriculum again: in the lack of external guidelines, teachers have been compelled to create their own practices that vary due to distinct teachers' own theories-in-use. Furthermore, some CLIL teachers assumed that the EFL teacher takes care of the language assessment which denotes that CLIL assessment is externalised to someone who necessarily does not work with the pupils in subject study at all. And vice versa: in EFL lessons, the proficiency of in subject-specific genres does not necessarily come across at all.

The bundling of assessment for EFL and subject-specific English is more challenging if the CLIL class teacher is not teaching EFL to the class. In such cases, co-operation between the two teachers is needed, but that does not always seem to occur, as my data suggests. Assessment is even more challenging if the teacher and assessor of EFL and/or English in CLIL has no language training. Whether a single grade or statement is representative enough of CLIL pupils' language proficiency is also a relevant issue to raise with educators in addition to calling the question of the fairness of varying grading practices. If the English teacher and the CLIL class teacher are not the same person, then it would be important to know how the information of the linguistic achievement and progress in CLIL subjects is translated to assessments in the school year report.

The objectives of EFL do not take CLIL-type English into consideration. If my view of CLIL English as an overarching aspect in bilingual primary instruction is embraced (Figure 43), then assessment practices and especially feedback and reporting should also be reconsidered, because CLIL English is much more diversified than a single EFL or CLIL subject grade. This is what some of the research participants, even pupils, also voiced: one grade is only an ambiguous generalisation or a mean value of a complex phenomenon which reveals nothing else than how the pupils manage on average both in CLIL English and EFL studies, which are not comparable.

According to assessment literature and research, grades do not motivate learners to enhance their performance, but rather engender performance-avoidance strategies (Pulfrey, Buchs & Butera 2011), whereas non-graded tasks elicit higher intrinsic motivation and task interest (Pulfrey, Darnon & Butera 2013). The study of Mäensivu (1999) concluded that verbal assessment supports the development of primary pupils' school self-concept and helps to demerge the contents into more manageable pieces. Furthermore, verbal assessment, rather than numerical assessment helps pupils to set their own learning objectives, make choices, take risks and evaluate their own possibilities (ibid.). An older study (Butler & Nisan 1986, 215) also concluded that routinely given grades "may encourage an emphasis on quantitative aspects of learning, depress creativity, foster fear of failure, and undermine interest" and "no negative results ensue from the use of task-related individualized comments". These studies underpin the use of alternative assessments as well as diversified means of feedback and challenge the idea of reconsidering grades as the only method of school year reporting.

Pupils and parents in both research schools, regardless the different school profiles, put forth that they wished to receive more assessment information on CLIL language matters. I believe that CLIL pupils deserve a more detailed description than one grade or a description containing one word or a few such as 'needs practice' or 'varying skills'. The description could entail not only their skills in using the TL in diverse subjects, because the competence may alter in different disciplines, but also the effort they place in learning and the attitude they show while working with the language even though the outcome may not even be near adequate (see also McMillan 2003, 39 on including student effort and motivation in assessment). How this would be achieved is a decision that needs to be mutually agreed within each CLIL-providing establishment. This research report should be helpful in suggesting some solutions to this issue. I will return to this again later in the next sub-section 8.2 under the heading *School reports and official reporting*.

On CLIL teachers' proficiencies

Research shows that the overall quality of teaching plays a significant role in the achievement of primary pupils. The achievement difference between a primary pupil with a low-performing teacher and a high-performing teacher may be as much as 53% (Sanders & Rivers 1996) or, as Hattie (2003) credits, approximately 30% of a students' achievement variance accounts for teachers. There is no reason to assume otherwise in CLIL. The impact of CLIL teachers as language models, facilitators of language use and composers of content-language integrated tasks is immense. Simply put, the more quality exposure to the TL, the better the linguistic outcomes.

In order to maximise exposure to the TL and natural, fluent input in CLIL classrooms, the CLIL class teachers should be competent, confident language users themselves. This unfortunately is not always the case, as the study of Pihko (2010) highlighted. CLIL is a specialised field within SLA education and therefore having specialised teachers with high

language proficiency is a basic requirement (Cf. the linguistic preconditions for CLIL teachers issued by the Ministry of Education in 2005 and Jäppinen 2004). Also, assessment of language proficiency is a specialised field within SLA, which CLIL teachers need to be familiar with. Hasselgreen (2005) underlines that class teachers, when in a position to assess language, should at least be aware of the basic criteria and methods. Whether or not teachers with little or no linguistic education are able to support, monitor and assess the development of language proficiency in EFL or CLIL competently is a question that has to be posed in this connection.

The educational background of teachers presumably has an impact on how primary CLIL class teachers see the role of language in CLIL and therefore also language assessment. Class teachers' training is mainly comprised of content studies. CLIL class teachers without any language training may more likely neglect the language aspect in CLIL and therefore also assessment of language, because teachers with different backgrounds and training emphasise various aspects in their teaching, as the CLIL assessment study of Serragiotto (2007) has proven. Lorenzo and Moore (2010) report on the tentative conclusion they made about the foci of different teachers in CLIL: content teachers tend to stress instruction of disciplinary vocabulary, whereas language teachers centre on grammar and sentence level issues reflecting the stance of the traditional language instruction. This bias is noticeable in the comments of the participants of the CLIL assessment survey and in the variation in their ideas about the roles of language in CLIL (instrumental, dual, eclectic). The effect of teacher background on teaching and assessment practices in CLIL is an interesting issue for future study.

Brookhart (2011) enlists essential educational assessment skills for teachers, and foregrounds knowledge of learning of the content area as number one. In CLIL, that also entails knowledge of second language learning. To avoid doubts of inexpertly language assessment in CLIL, I argue that language competences and studies in addition to basic class teacher education should be the decisive instead of secondary factor when recruiting CLIL teachers. A class teacher with double qualifications (language teacher + class teacher) is an ideal candidate for a CLIL teacher position. In the current situation, linguistic merits do not seem to be accentuated by teacher recruiters (Kangasvieri & al. 2012). Unfortunately, this practice indirectly signals that the language aspect is not seen as valuable in CLIL.

Assessment strategy: from implicit to evidence-based assessment

A need for structured, cohesive assessment strategy is an issue that was raised by a number of teachers. A similar concern was raised in an Italian case study in EFL contexts (Gattullo 2000). A CLIL assessment strategy would increase credibility and improve CLIL provision. However, it is important to remember that Finnish teachers are largely independent agents, who are not accustomed to imposed restrictions. A teacher's right to make individual assessment decisions at a classroom level must be maintained and appreciated, but the larger assessment frame, including formative, classroom assessments leading to school year reporting, needs to be mutually agreed upon. Common ground rules of, at least, the type and form of school year

reports, self-assessment forms and development discussions would give the appearance of a cohesive assessment scheme which also contributes to the equal treatment of pupils. This is a matter that can be determined in municipalities or schools.

The Finnish basic education system considers school year reports within grades 1–7 as a form of continuous, formative assessment, the task of which is to promote learning. From this perspective, grade ratings, as discussed earlier, are not appropriate, and they should be either replaced with a verbal account or accompanied with a verbal attachment. Assessment in Finnish basic education should be encouraging, but truthfulness and fairness have to be maintained as well. Quality assessment is informed by diverse evidence rather than gut instinct, which is, according to a few participant teachers, the basis of CLIL assessment. How each teacher gathers evidence for analysis is their own decision, and can be grounded on, for instance, the attributes of the group, handling of the subject topic and linguistic matter. It might be useful to discuss within the school's CLIL community whether or not it is necessary to include principles of bilingual testing, scoring and collection of audio data in the assessment scheme. Parents, although very tolerant, appear to make comparisons and remarks on teachers' different teaching and assessment practices, as became evident, for instance, in one parent's letter (see p. 168). Variation is richness, but an assessment strategy, albeit one providing room for spontaneous, resourceful and individualistic gauging or information gathering is professionalism.

8.2 CLIL assessment practices

After devoting a fairly high number of pages to the fundamentals of CLIL assessment, which are the bedrock for the actual assessment practices, I finally turn to assessment methods as disclosed by teachers nation-wide and their adequateness as appraised by pupils and parents in the two research schools. I hope to have shown that the 'infrastructure' of assessment is quintessential to successful assessment in CLIL. Assessment is not only about assessment methods and practices, it is also about substantiating and contextualising the inferences and judgements achieved through miscellaneous assessment methods, the implementation of which contributes to the trustworthiness, fairness and representativeness of the assessments, and increases the equality of pupils. Assessment is an act of power on the teacher's part and therefore high ethicality is required. In this study, it was discovered that teachers employ a wide variety of assessment strategies, often without the curriculum as a reference point, and the strategies and practices appear to be, for the most part, infrequent, incidental, implicit and based on impressions. Indeed, language assessment in CLIL is not an established practice.

Teacher observation and feedback

The most widely used CLIL assessment method in the assessment survey was teacher observation, a form of continuous assessment. Observation as an assessment method represents implicit classroom assessment that is often used for deciding, for example, how to proceed with teaching. As Bachman and Palmer (2010, 29) note, both learners and the teacher may be unaware of the assessment process, which is why observation ultimately serves learning facilitation and immediate decisions. Without any specific focus, observation may fall into the category of the previously mentioned 'gut instinct' assessments. Another challenge with observation is that it does not necessarily translate into actual feedback, and the target of observation may also be allusive.

In order for the observation to centre on language use, there has to be plenty of language use situations in the classroom in which the learners, instead of the teacher, are producing language or using it. This appears to be a challenge in language classrooms. Communicative language learning has been the dogma in SLA for decades, but, according to Eurydice survey report (2012, 12) on language teaching in Europe, EFL or equivalent students in most participating countries claim that the TL is not usually used during lessons by their teachers. This result might be in connection with teachers' language proficiency, for a CLIL classroom study revealed that, when encouraged by a teacher proficient in the TL, students were activated to participate in the teacher-orchestrated classroom discourse (Nikula 2010, 110). Dialogic interaction, or teacher-initiated classroom discussion, the third most often used assessment method, may thus provide input for teacher observation, but in order to succeed it requires a competent TL speaker teacher, and one who also is the organiser of student group discussions and collaborative work in the TL.

Teachers' immediate feedback is normally grounded on observations made during the lessons. McKay (2006, 195) names observation as "a central tool" in assessment of young learners' oral language; assessment through the other three basic language skills will only "deny the essence of young learners' language learning" (ibid., 177). Observation is thus inherent in oral language assessment, which in turn was characterised as challenging by teachers. Since observation is a substantial assessment method, (immediate) feedback generated on the basis of observation is desirable assessment information. In the research literature, feedback is one of the most efficient means of promoting learning and it is considered to be one of main qualities of an expert teacher (see e.g. Hattie 2003; Hattie & Timperley 2007).

Teachers named oral feedback as the most often used method for informing pupils of their language proficiency in CLIL subjects, although this statement was slightly contradictory to their own temporal definitions of feedback frequency, according to which feedback was given rarely by over a third of teachers. This finding is congruent with what is reported in a study by Havnes, Smith, Dysthe and Ludvigsen (2012) in the Norwegian secondary education context: teachers' formative feedback for students is rare. Their study also concluded that the feedback practices are highly teacher-dependent and feedback will more likely be given in workshop-type situations

than in academically oriented settings. This leads to the question of whether the working methods in CLIL classrooms are such that they are supportive of giving and receiving feedback, or how related this outcome is to teachers' personality or even sociocultural factors.

The reasons why teachers do not seem to convey assessment information or give that much feedback to pupils may be manifold. One reason for a lack of teacher feedback may be the perception that English does not need to be noticed and assessed in CLIL contexts due to its instrumental role in the instruction. Another reason could be that assessment of language proficiency does not take place in any particular manner or form in CLIL classrooms. A third possibility is that there are no adequate assessment tools available for particularly novice CLIL teachers who balance with numerous requirements and challenges such as language control, age-appropriate material production and the quest for relevant sources.

An additional reason may be the lack of theoretical knowledge of language assessment and language learning, as linguistically competent teachers are not always available for recruitment. Accordingly, the teacher elected for a CLIL position or a substituting teacher may not have any language education or insights into CLIL instruction. One interviewee, a new teacher in CLIL, stressed that the school administration does not allow substitute teachers to participate in in-service training courses which in turn leads into an unbearable situation where pure enthusiasm may fall short. Finally, a very human cause for not conducting assessment – and therefore also not giving feedback on language issues – is the principle of applying the least effort. Assessment is arduous and time-consuming; these two attributes of assessment were also frequently found in the data.

The fact that teachers do not provide plenty of feedback was also supported by pupils' equivalent characterisations: one of the main findings in this research was that the pupils in the two research schools felt they were not receiving enough feedback from their teachers. Boud and Molloy (2013, 1) note that “through feedback teachers communicate what they value and do not value in what students do”. Following this line of thought, communicating very little or no feedback signals to learners that what they do is not valuable. This affects motivation which has, according to motivation studies (see e.g. Ortega 2009), a reciprocal relationship to learning.

According to this research, one third of pupils reported receiving feedback rarely or never, and over 60% wished to receive more feedback, in particular oral feedback directly from the teacher. Pronunciation and issues of speaking concerned them considerably more than other aspects of language use. This implies that organising opportunities to speak in the TL is highly important in addition to designing opportunities to receive feedback on pronunciation. Proper linguistic models are also important. Pupils perceived teachers' or other adults' (e.g. native teacher's) feedback in any form as more reliable and important than self-assessment or peer assessment. This indicates that pupils depend upon the professional judgement of adults, who they trust are better linguistic experts than themselves or the 'kid at the next desk'.

Feedback seems to be more important to CLIL pupils in the beginning stages, and girls valued linguistic feedback more highly than boys. It is probable that pupils' linguistic self-image

becomes more established after the outset of EFL in the 3rd grade in connection to which the English language is assessed in a more formal manner – but from different angle, as I argued in the previous section. Only 8% of pupils claimed that they get feedback on their language proficiency frequently. My hypothesis is that these pupils are the outspoken ones, who are linguistically talented and succeed without any specific effort to demonstrate their language proficiency. If English is markedly present, actually used in the lessons and produced by pupils, teachers could give spontaneous, targeted, individual and instant feedback for pupils on their coping in English to embolden them and boost their linguistic self-esteem. Research has shown that half of EFL learners, of various ages at diverse levels of education, profit from corrective feedback (Havranek 2002). Corrective feedback in turn is closely linked to a focus on form approach which is, in CLIL research circles, promoted as one step closer to quality CLIL (see 2.3).

Quality assessments are based on multimodal evidence; observation alone is not sufficiently trustworthy. Teachers create impressions of pupils as TL learners and users through observation, but these impressions may be deceptive, as the case recounted by Rahman (2012, 107) reveals. Here, the teacher had created an impression of a pupil being “shy and quiet in the classroom”, but the interview audio recording evidence collected by a native teacher surprised her: the pupil took an “active and capable role” in the face-to-face interview showing “good skills in his English” (ibid.). In different circumstances pupils take different roles. Without the recordings and the native teacher’s contribution, the language skills of the pupil may never have become known to the teacher, thus resulting in a falsely negative impression or assessment. How extensively teachers use recordings for actual assessment or feedback provision remained unclear in this study.

Furthermore, the risk of observing and assessing features other than language (e.g. personality, classroom behaviour) – especially if the observation is unsystematic and unplanned – is always present, as the study conducted by Llosa (2011, 370) shows (see also Bachman 2004, 155). Additionally, the CLIL assessment research conducted by Hönig (2010) is a good reminder of the fact that teachers are not always assessing what think they are assessing. The point to be taken is that we should never be content with one assessment practice only, but should strive to implement various assessment methods, thus gathering more diversified evidence on which we could base our decisions or judgements on, for instance, support instruction, differentiation or reporting. Multifaceted assessment driven by multiple methods is the key to valid and more trustworthy inferences rather than gut feelings based on subtle observations.

In order to use teacher observation for well-grounded assessments, observation sheets, as proposed by several experts (e.g. Short 1993, Quartapelle 2012), may help teachers in focusing their observations to intended language material. This, as stressed above, needs material for observation – which is assumingly why dialogic interaction was mentioned as the third most employed assessment method. Observation sheets, which are practically one form of rubrics, may be clumsy and time-consuming to fill in amidst teacher-driven lessons, but pupil

presentations and plays, for instance, provide opportunities for organised, systematic observation. Additionally, recorded interviews, reading aloud, singing, group discussions, reports or radio plays (either self- or pre-written) provide material for observation. The recording does not necessarily restrict the teacher; pupils may be far more skilled in using technology than the teacher and are able to handle the recordings on their own using a laptop, voice recorder, a tablet or pupils' own smart phones. Yet, the storage of the recordings is worth considering as well, regarding their further use in assessment: there is no point collecting an assessment data bank if it is not used for feedback or reporting.

Bilingual testing

Bilingual tests or test sections were the second most used assessment method; some teachers deployed monolingual tests in the TL. The language of instruction appears to be the decisive factor: issues taught in English are tested in English. The quality of TL test items remained relatively unclear, but it is likely that less demanding content is assessed through the TL. The age of pupils and the level of exposure naturally affect the choice and design of test tasks. The most appropriate tests elicit stretches of language above individual word level, make use of all four language skills, and also challenge higher order thinking skills and integrate various dimensions of knowledge (see Bloom's taxonomy). Principles of task-based performance assessment as well as technology-based language assessment can be incorporated in bilingual testing, not to mention collaborative testing (see 4.3).

An untypical and unique method of group tests was discovered and reported in this research: such tests represent collaborative assessment and are substantively suitable for CLIL assessment for a number of reasons. First, collaborative tests encourage knowledge construction in social interaction according to the spirit of current socio-constructivist views of learning. Secondly, the test type may include several phases, which reduces the 'feeling' of being assessed particularly if alternative methods are employed as recommended with young language learners. The end result does not conform to paper tests or any specific testing form as the recount in section 6.1 exemplifies – group tests could thus be perceived as a form of multiphase group work. Thirdly, when an individual test is included, more specific information on individual performances and knowledge can also be gathered.

Furthermore, the scoring and assessment may be light or heavy, including or excluding self-assessment, peer assessment, teacher assessment, or criterion-referenced inferencing with rubrics (see e.g. Gottlieb 2006). Also, the language emphasis may vary in different test phases or sections according to the extent of CLIL exposure and objectives. The ultimate advantage of group or collaborative testing is, besides the social construction of knowledge, that it reduces test anxiety and occurs in the zone of proximal development (see 2.3) of some pupils, while the more capable peers benefit from passing on their knowledge and skills. I consider group testing a very promising and worthy assessment method in CLIL that has much to offer due to its flexibility.

Criterion-referenced inferencing

Criterion-referenced inferencing is encouraged by the NCC (2004), which is why it was somewhat surprising that only a few participating CLIL teachers mentioned rubrics as supporting devices. It may be that the long tradition of norm-referenced assessment has not yet receded. Rubrics or assessment grids for any types of tasks are also recommended by several CLIL scholars (see 4.4) because they are helpful in generalising from isolated observations and keeping the focus of assessment on where it is intended to be. Rubrics underline the essentials in each assessment task and make the assessment process more transparent for everyone involved; there are no hidden issues or factors in the score or grade.

Rubrics are one means to enhance learner-centricity and initiative, because pupils are capable of self-assessment using rubrics, and the criteria help pupils in planning their own goals – even more so if they have been included in the design phase. Criterion-referenced inferencing, particularly suitable for student projects, also cultivates learners' language awareness and sense of responsibility. When the learners are aware of what is necessary for a given score, grade or assessment clause, then they are better able to pursue the level they have set for themselves and wish to achieve. Furthermore, pupils are less likely to face disappointments when the criteria for assessment are defined in advance; the score will not be delivered as a surprise.

Common European Framework of Reference for Languages

The Common European Framework of References for Languages (CEFR 2001), an established set of various criteria related to language use, was rarely mentioned in the data. It was only raised once in the CLIL assessment questionnaire section inquiring about developmental directions for CLIL. This either implies that it is not very familiar to CLIL class teachers or it is not used as an assessment tool or reference. The official role of the CEFR in Finland, however, is prominent: the NCC (2004) has explicitly adopted the framework as a reference document. The CEFR is introduced to subject language teachers during their academic studies, but I doubt whether it is introduced in class teacher studies. Furthermore, since the CEFR is more than ten years old, it may have evaded the teachers who acquired their language qualifications in CLIL before the 2000s.

The CEFR, besides rubrics and portfolios, is very recommendable, because it shifts the focus from pure linguistic forms to the individual language user (the 'can do' statements) and "suggests that users compare what they can do with what they want to be able to do in the new language" (Larsen-Freeman 2008, 159–160). There are many aspects that advocate for the use of CEFR scales and descriptions as a reference. First, the framework is widely used, although it does not specifically take subject-specificity into account. For that reason, CLIL-related developmental assessment projects drawing on the CEFR have been initiated (see 4.4). Secondly, at least at certain level, the CEFR is also perceived as suitable for the assessment of young language learners (e.g. Hasselgreen 2005). The framework does not bind its users to any specific assessment method. Instead, it is to be regarded as an aid towards reliable, uniform

assessment which does not exclude creativity and adjustment. For these reasons, I recommend the integration of the CEFR as an assessment tool in CLIL contexts, with some restrictions.

These restrictions are associated the primary children's general proficiency levels and the age-appropriateness of the descriptors discussed in Hasselgreen, Kaledaitė, Maldonado-Martin and Pizorn (2011). They argue that at primary level in EFL instruction, pupils rarely exceed level B1. In CLIL, there may be more exceptions to this, as parents' estimations of the CEFR level of their children indicate – an interesting topic to study in the future. It is also reasonable to state that the original CEFR scales are written for adults and are based on the adult experience (ibid.,11). To create clear, age-appropriate and group-specific language descriptors that can be shared with pupils and their parents, we have to return to rubrics or assessment grids, although – now with coding labels from A1 to 'B1 and above', which is the highest descriptor level outlined by Hasselgreen and colleagues (2011). Some primary CLIL pupils, however, may reach B2 and C levels, as was noticed in the parents' estimations of their children's proficiency level.

It is a useful skill for a CLIL teacher to be able to draft various rubrics with the help of the original CEFR document, which provides a well-established practical starting point for linguistic assessment. The desired language levels that should be defined in the CLIL curricula are also appropriate when they conform to the CEFR scales. Additionally, there is no reason not to implement the CEFR in summative CLIL assessment as well. Since the CEFR is such an influential and wide-spread document, it might be a good idea to introduce the scales to pupils who will most likely grow up in Europe, which still relies on the CEFR system.

Self-assessment

Pupils' self-assessment was the fourth most used assessment method by teachers. It appears that self-assessment is, in addition to tests and test sections, a method that is more systematically and regularly used for explicit assessment purposes than teacher observation and dialogic interaction. The data in this study implies that self-assessments are mostly executed in connection with tests. Parents, in general, did not consider the self-assessments of their children as very valuable sources of information – neither did the children themselves. One reason for this may be, as the study of Kärkkäinen (2011) concluded, that children tend to assess their skills more positively than their teachers and parents. Kärkkäinen also noticed that younger pupils tend to be more positive about their skills than older ones.

Self-assessments were more appreciated when they also contained the teacher's estimation of the child's performance. How explicitly these self-assessments were connected with language remained largely ambiguous. Although self-assessments were not at the top of pupils' and their parents' preferred means of feedback, teachers emphasised their utility in documenting pupils' gained language awareness and self-knowledge, with which I agree. Combined with teachers' gauging, pupils' self-assessments should be maintained and increased in number especially since the assessment method is legitimised by the Finnish Basic Education Act (1988/628) and further fostered by the NCC (2004) and the NCC draft (2014).

What could be done differently, then, to increase the value of self-assessments in pupils' and parents' eyes? Firstly, the self-assessments should be addressed and discussed so that they do not just remain detached of the educational objectives and syllabi. They have to have a meaning that connects them to the everyday school work. Secondly, self-assessments should have true relevance to that school work; they need to be attached to the actual learning objectives and issues that have been studied and then discussed in more detail. Rubrics are an adequate tool in this regard. When the aims and criteria are known in advance, it is easier for pupils to reflect their own performance (either linguistic, content or both) in comparison to the pre-negotiated standards and criteria reflecting those standards. If that is achieved in co-operation, everyone could claim ownership and be the agents of their own actions and learning. Thirdly, self-assessments should not be too general in nature. They have to be more specific than just a few questions in the beginning or at the end of the test paper asking 'How did you manage in this test?' or 'Did you prepare well for this test?' in order to encourage deeper reflection, which is one of the general goals in the NCC draft (2014). Finally, reflection can be practised and skills in self-assessment developed to focus on performance and personal development.

Portfolios

Portfolios were the least used CLIL assessment methods, which I found surprising. Many of the interviewed teachers had at some point in their career experimented with portfolios but then abandoned them for some reason. In Europe, portfolios are rather widely used and popular even in early stages of learning, but the portfolio models or blueprints, which have only recently become available in Finland, are targeted for EFL study, which presents a slight problem when the focus is on CLIL-type acquisition and CLIL English. This denotes that the existing European Language Portfolio model has to be adjusted, or the teachers needs to draft a model of their own. Learning diaries, or learning logs, and portfolios are highly recommendable in CLIL environments, as the experimentations both in EFL and CLIL contexts show (Wewer forthcoming). A proponent for portfolio work is the NCC draft (2014) which explicitly mentions this in relation to assessment in CLIL.

School reports and official reporting

Assessment is traditionally conducted for diverse purposes, chiefly formative and summative. The Finnish NCC (2004) does not sharply contrast these two forms of assessment during basic education. The decisions and judgements made on the basis of multifaceted assessments are always intended to enhance and support learning, and a final assessment is only given at the very end of the basic education (grade 9), and it is based on the pupil achievement during the last two grades. This denotes that assessment throughout grades 1–7 is formative only. The Finnish assessment system in basic education is very holistic, and formative assessments can (and will) be used for summative purposes especially in the school year reports, as the NCC draft (2014)

for the revised NCC 2016 observes. This may appear slightly confusing, and the differentiation of assessment purposes would clarify assessment through grades 1–7.

The school year report is an overall evaluation of how well the given pupil has reached the objectives within the whole year and declares whether the pupil will advance to the next level or repeat the study of that year. However, due to their quality of being an average description, report grades do not disclose anything other than the general level. In order to convey a multifaceted account of a pupil's skills and knowledge, numerical assessments need to be complemented with verbal, genuinely descriptive assessments in order to actualise the spirit of NCC which sees formative, continuous assessment as the guide line in basic education.

Slightly over half of the pupils and their parents stated that they have received assessment information about the success of CLIL study through English in the school report. This is peculiar because neither of the research schools issues a separate CLIL assessment attachment and, in the actual report, CLIL language is not assessed. The participants were probably referring to the formal English grade which is given in the upper primary grades – this provides further evidence for how EFL instruction is associated with CLIL study. Furthermore, teachers have very deviating understandings on how to assess CLIL English and where the proficiency assessment should be included, if anywhere. School year reports were desired by slightly over half of the pupils as a means of receiving feedback on their language proficiency in CLIL, whereas 22% of parents mentioned reports as the preferred method. If we give weight to pupils' wishes – as I think we should – CLIL reports are worth embarking on. In some schools, as reported by the participating teachers, self-assessment forms were part of the official CLIL assessment scheme; the form was delivered as a school report attachment to pupils' homes, and it included more than just strictly language-related aspects.

I am in favour of such attachment practices because it ultimately increases the amount of assessment information reaching homes and issuing report attachments also ensures that assessment takes place. When the self-assessments are combined with teachers' assessments, the amount is doubled, and when the assessments are aligned with learning objectives, assessment is even more adequate. We have to remember that the guidelines of the NCC (2004) should be transferred to practical levels: CLIL assessment should give adequate information on the child's language proficiency and its development in relation to the premeditated objectives. This means that actual assessments and objectives should communicate with each other at all times. I propound that specific CLIL attachments should be added to normal school reports if a discrete CLIL report paper which differentiates between CLIL English and EFL is not created.

My viewpoint of assessment at primary level is, however, that it is not sensible to give overly detailed linguistic descriptions of pupils' language proficiency in CLIL. I thus advocate for designing assessment schemes that disclose *what pupils are able to do with the language in CLIL subjects* rather than pointing out what they are lacking or still cannot accomplish. Assessment should point to pupils' strengths rather than weaknesses and show how well they can actually work with the language in different subjects. The intention is not to compare pupils

with each other (normative assessment) or to rank them, but to give individual feedback in relation to the objectives. For instance, some key objectives of core CLIL subjects (e.g. mathematics, science subjects) could be highlighted, and the child's functional language proficiency analysed more closely, for example, in the following way:

The child can add and/or subtract in English in the range of 0-100.

The child masters times tables 6-9 in English.

The child is able to describe/narrate/infer habitats of arctic animals in English.

If considered relevant, adverbs of degree can also be added to the clauses. Other parts of the school year assessment could consist of lexical abilities, affective factors, self-assessment and whatever is perceived as appropriate for the given CLIL context. The extent of the language exposure is one decisive factor in defining which approach to official reporting (a form of its own right, an attachment or none) is sensible. As a rule of thumb, the extent of both formative and summative assessments should be proportionate to the extent of TL exposure. For instance, with 25% TL instruction, approximately one quarter of the school year report could be centred on studying through the TL as well as the progress in the TL. After all, such a proportion of instruction in a foreign language is a marked feature in a child's education.

It may be justified to abstain from issuing linguistic report cards to very young learners. As Katz (1997, para 6) argues, school reports with grades and achievement scores are only appropriate from the 3rd grade onwards, for, by then, "children's abilities and aptitudes are likely to have stabilized". For the young beginners, other feedback methods that concretely visualize the amount and quality of learned language may be most appropriate (e.g. portfolios).

Some teachers expressed their concern over pupils' low levels of language awareness. These results give subtle indications that more language-sensitive, focus on form CLIL instruction is essential. It might be justifiable to reinforce the language identity or language self-perception of girls since they were more in need of linguistic feedback than boys. One possibility is to address CLIL English and its progression in development discussions, because especially younger learners seem to desire feedback on their linguistic skills. It is rather a norm than an exception in Finland that class teachers organise development discussions at least once a year in replacement of semester reports or in addition to formal reports. Sometimes the pupil is invited; sometimes the discussion is free flowing; sometimes the conversation is based on a form.

It is not clear to what extent CLIL teachers touch upon linguistic issues in these discussions, but in this study, discussions were the third most used method of providing feedback to pupils. Because both parents and pupils expressed their wish to generate and maintain a more personal contact with the teacher, time and effort should be invested in these events. Development discussions are one prominent channel for bringing language progression and development to the fore. These discussions are excellent occasions to play samples of the pupil's possible audio recordings or to let the child (who should absolutely be included in discussions concerning his/her own issues) to choose and present assorted samples of portfolio work, for instance. Such practices would make the language-related discussion more evidence-based.

Indirect feedback: interpreting children's proficiency self-reliantly

One very alarming finding was that occasionally the assessment information received from school is self-reliant only. In such cases, it is expected that pupils and their parents will carry out independent investigations into the school work and test papers and then draw inferences about the language proficiency or coping in CLIL based on the material provided by the teacher. In other words, the encoding of language-related matters and the interpretation of them has been entrusted to parents and pupils, who seldom have such expertise. Indeed, they have to rely on their own assumptions, which may be inaccurate, too idealistic or pessimistic. This attitude some teachers displayed may stem from any one of the challenges mentioned in this chapter or, regarding the CLIL assessment survey, in chapter 6. The CLIL teacher should be the language expert who, based on observations and gathered evidence, together with pupils celebrates the progress made, offers suggestions for further improvement and points out how much pupils have achieved so far in their language growth.

The majority of parents (76%) in the two research schools, in line with their children, wished to receive more assessment information on language proficiency and development; 68% of parents desired more detailed information. Such a volume may well be a response to indirect practices. It is also possible that supply increases demand; that is, without this research such a need would perhaps never have come up so explicitly – after all, education provision is slowly approaching customer culture. It is obvious that indirect feedback methods involving self-reliant interpretation cannot be recommended by any means. Parents also lack the wide-angle vision teachers have on objectives, general characteristics of the class, instruction given etc. Additionally, some parents may for various reasons not speak or understand English at all. Parents also rarely have linguistic training or the language proficiency that CLIL teachers should have.

8.3. Affordances of computer simulations

One solution to the need of all-encompassing assessment information is the LangPerform computer simulation concept that was included as an empirical experimentation in this research (see 5.4.2) in order to modernise language assessment by incorporating technology and to see the affordances of such a method in CLIL and for young language learners. The results, although partly polarised, were very supportive of computer simulations as an alternative assessment method in CLIL. Technical failures excluded, as a whole, both pupils and parents were unanimous in seeing PROFICOM (2013) computer simulations as suitable for the assessment of listening comprehension, general language skills, pronunciation and fluency, and the majority of them perceived computer simulations as appropriate for assessment of content-based English language proficiency. To evaluate the simulations and the experiment more closely, I turn back to

Chapelle's (2001) six criteria for appropriate computer-assisted language learning which I rewrote to coincide with the assessment perspective (see p. 98). The criteria are: language assessment potential, test taker fit, meaning and form focus, authenticity, positive impact and practicality.

The first criterion, *language assessment potential*, refers to the capacity of simulations to provide evidence based on which the inferences of pupils' language proficiency will be made. According to the study, simulations are highly potential as alternative, future assessment method in eliciting various language samples for closer scrutiny. The assets of computer simulations include interaction with people that pupils are not likely to encounter in their normal lives and the possibility to hear different kinds of English. Pupils often referred to the possibility to speak English as a benefit. In the two simulations experimented within this study, the emphasis was more on listening comprehension and speaking than on writing or reading, but both simulations contained tasks eliciting language in all four basic skills. The size of the computer screen, however, sets some limitations to how rational it is to incorporate longer texts, but it should be more flexible to produce, type and edit text using a keyboard and a word processor programme than paper and a pencil. Using computer applications should be a basic skill for today's learners.

The second criterion, *test taker fit*, denotes the amount of opportunities to use content-specific language, i.e. academic English. The nature of the tasks defines how subject-specific and content-related the language is in the simulations. This is an issue that can be predefined in the script phase. Approximately half of the experimented simulation tasks were planned to elicit subject-specific language. The test tasks were reviewed in a polarised way by pupils. For example, in the first simulation, every single task (CLIL and EFL) received a rating as being both difficult and easy in someone's opinion. This most likely stems from individual learner differences and variation in instructional emphases; CLIL does not occur in a uniform manner in neighbouring classrooms especially if the CLIL curriculum does not provide any guidelines. This deviation in task evaluations also suggests that in testing it is important to use multiple methods as well a variety of tasks representing diverse difficulty levels which enable as many pupils as possible to demonstrate their language proficiency.

The pupils, in general, also estimated content-related tasks as the most difficult, although one class particularly stood out in this respect. This specific class did not see mathematical tasks in simulation 1 to be nearly as troublesome as did their peers. This implies that subject-specific mathematical English is not so prominently present and used in the CLIL classrooms that pupils would feel confident enough to express their content knowledge in English. They, according to my observations in the simulation sessions, seem to compensate for the lack of CLIL English by using casual English, which is an efficient strategy, but leads to inaccurate content manifestation. Another reason for this might be that the language use in the CLIL classrooms promotes conversational English, which is why the academic competencies do not reach a functional level. A third reason may be that the used TL materials are not modified to suit the learners' needs (e.g. age, proficiency level), or the learning through the TL is not scaffolded. The correlation between

experienced task difficulty and linguistic task performance in general is a topic that needs further examination.

The third criterion, *meaning and form focus*, refers to the balance of content and language in the simulation test and the extent to which the testees' attention is guided towards either one. In CLIL, the primary focus is on content, but in CLIL language assessment the focus is on the manner that content mastery is conveyed through English. As a result, the simulation contained no explicit focus on form or forms. However, pupils were scaffolded content-wise by supporting pictures and sub-topic lists. The simulations were totally in English. Translations of frequently occurring instructions that were vital for understanding (e.g. the verbs 'response', 'reply' and 'continue') were given in the introduction phase prior to the actual simulation performance. The simulation questionnaires elicited a number of ideas of how to further improve and adjust the simulations. Pupils and their parents suggested various scaffolding functions such as allowing a question to be replayed when needed.

Authenticity is the fourth criterion for the appropriateness of computer-assisted language assessment, and it signifies the degree to which the simulation corresponds to the real-world. Organising cross-contextual assessments mixing both academic (decontextualised) and practical (contextualised) environments enhances "making school activities more authentic to real-world activities and using authentic educational assessments as a vehicle for achieving this" (Messick 1994, 19). This is what we have striven to do with these simulations. The LangPerform simulation, however, must be described as a semi-authentic assessment method, because it is based on a script and the actors do not interact with the test takers in real-time; thus, they do not reply to questions – a fact that irritated a few pupils who may be used to playing online games with real-time interaction with their co-players. In that sense, the simulation is not equivalent to real-life situations, but so far it represents one of the best vehicles to enable encountering strangers safely in classroom situations.

Furthermore, native speakers would most likely use a wide variety of expressions – the dialogue was intentionally somewhat facilitated. Some pupils commented on the facility of simulation English, but some were struggling regardless of the facilitated language. This might also be the case in real encounters with English-speakers. The simulations contain a fairly large proportion of close-ups, which some pupils found slightly intimidating; they felt that their interlocutors stared at them (while acting as if they were listening to the pupils' speech). In real life, the participant could decide a comfortable distance.

If the dialogue and communication were not always perceived as totally authentic, pupils seemed to assess the filming locations as authentic. The Michigan video was filmed in Finland and the international school was not really international but an ordinary, although just renovated, Finnish suburban school. The impression of being elsewhere was created through music, images and fluent English-speaking actors, who were in fact not actors but lay persons. Some pupils stated how great it was to visit America and how nice it was to spend a day with the international school student Amy.

The fifth criterion is *the positive impact* that the simulation as an alternative assessment method would have on the testees. The simulations, as previously mentioned, were received very positively and acceptingly. One reason for this must be the novelty and innovativeness of the assessment method, which clearly intrigued the participants; the arrival of educational technology in schools has been rather slow regardless of visionary plans. The simulations may correspond to the digital life style the primary pupils have grown up with better than traditional paper tests; 58% of the pupils piloting simulation 2, surprisingly girls more often than boys, stated that they would choose a computer simulation test over a paper-and-pencil test. The gradual shift of Finnish baccalaureate exams to digital, computer-assisted test batteries is a firm sign of the change that will slowly but surely also be extended to primary schools, and LangPerform simulations or equivalent testing concepts are likely to proliferate in the future.

Another reason for the positive review may be that simulations appear to create meaningful language use situations that differ in quality and style to those created in the classroom. The teacher in the classroom is often bilingual and probably uses code switching, but in the simulation performance, there is no other option than to use English. Once the simulation concept and technology is seized and adopted, the method as such is not that exciting, and pupils can concentrate more on their own performance and the tasks integrated in the frame story. Signs of this emerged in the second piloting round, when the 5th graders, testing the simulation for the second time, reported less test anxiety.

The simulation concept also provides an evidence-based pool for assessments given to and by all stakeholders: pupils, their parents and teachers. Parents rarely get an opportunity to listen to their children express themselves in CLIL contexts, to 'talk mathematics' for instance. That opportunity was appreciated, although this assessment method requires quite an amount of activity from parents in addition to a computer and an Internet connection. Nowadays, almost everyone has access to computers or has one or more at their disposal; the personal computer thus serves as a digital channel between home and school (see Figure 16). Pupils' production, be it spoken or written, and choices (e.g. clicked or removed items) can be retrieved multiple times for monitoring after the actual performance. The simulations can also be repeated, which provides evidence of progress over time.

Some pupils considered the simulation experience to be unpleasant. My understanding is that the pupils who, for one reason or another, failed to produce an (to their own standards) adequate language performance were the most critical ones towards the simulation. It is a humane trait to like something one is successful with. Some pupils appeared to be more sensitive to interferences; neither of the piloting areas in the two schools were equipped or furnished to function as a language studio which would provide at least some privacy for the testees. Pupils who distinguished themselves with effortless language production may have discouraged their less linguistically able peers – some pupils did not say anything for a long time, but monitored and listened to others. It thus takes time to adopt a method like this and to be emboldened in subject-specific English expression.

Furthermore, the functional uncertainty of the simulations, partly caused by coding problems and partly by device faults, contributed to pupils' disappointment. Since the simulation concept is still a fairly recent innovation, there is work to be done to stabilise and improve it. Especially the language laboratory section (assessment area) was not adapted for young language learners. It was not visually attractive, either. The concept employs the CEFR which was probably not very familiar to pupils, and the CEFR codes may not have been clear for pupils without separate clarification. There was no time in the piloting sessions for such clarification. In addition, the language of the language biography section and the language laboratory contained demanding concepts and was, at first, a mixture of English and German – the task-related texts were given in Finnish. This was confusing for pupils, for their emerging learner language was not so evolved that they could operate with such specialised language (e.g. the concept of accuracy).

The sixth and last criterion for the appropriateness of computer simulations as an assessment method in CLIL is *practicality*, i.e. the extent of various resources devoted to the production and maintenance of the simulation environment. This is the weakest point of the LangPerform simulation. The pre-production, production and post-production take hundreds of hours of work and involve a number of people. The production structure is heavy in comparison to traditional classroom assessment. The implementation of the simulation requires authorised access; the simulations, therefore, are not accessible effortlessly and openly. The maintenance is strongly centralised in the University of Tampere which also administers the simulation environments. Thus, simulations are expensive and so far exclusive.

Once a simulation is completed, however, it is possible to implement the same simulation in many contexts. The distribution and wider experimentation of the PROFICOM simulations lack funding, so their further use for the benefit of the CLIL community is at this point uncertain. Another problem is that the CLIL contexts vary enormously in the extent of language exposure, content and implementation – in the absence of CLIL curricula there is no guarantee that the CLIL pupils in one classroom have studied the same subjects and contents through English as their peers in the next classroom. Therefore, I as a screenwriter tried to anticipate this by selecting more general rather than specialised topics that, in my opinion, would most likely be covered in the CLIL syllabi.

LangPerform computer simulations, then, have their advantages and weaknesses, but considering from the large perspective, they seem to have captured pupils' and parents' interest. The simulation thus seems to mediate a more comprehensive and diverse overview of CLIL pupils' language proficiency and specifically the mastery of contents through English. They appear to particularly facilitate parents in forming a more precise understanding of the oral language proficiency of their children. They provide motivating and modern ways to assess language proficiency comprehensively, and they represent technological advancements that are perceived as a key element in future education (e.g. Bierenbaum & al. 2006; Marsh 2013). These computer simulations coincide with the CLIL Cascade Network Foresight Think Tank report 'Talking the Future 2010-2020' (Asikainen & al. 2010, 4) which, in order to re-shape current

language education systems – including CLIL – encourages to “further develop and implement language assessment and evaluation systems to better measure what is valued in language competences for the information age, and not just value what is measured”.

8.4 Validity and significance

This study represents a mixed methods (MM) research (see 5.1) in which the advantages of both quantitative and qualitative research are merged, their weaknesses are compensated and the results are more versatile than in a monomethodological study (Teddlie and Tashakkori 2003, 14-17). Mm study and triangulation are desired aspects of CLIL research (Pérez Cañado 2012). Although MM research in mixing data collection and analysis methods seeks to provide more consolidated results and increase validity through the multimethodological approach, validity issues still need to be considered. Validity in general, regardless of research approach, refers to the quality of the research, how solid the design and the research phases as well as the obtained results are.

Different research approaches require different terminology in reference to validity, and distinguish various kinds of validity. Whereas quantitative research discusses internal validity and external validity, qualitative research deploys the terms trustworthiness and transferability. The equivalent terms for mixed methods research, as proposed by Teddlie & Tashakkori (2003), are inference quality and inference transferability. Inference quality denotes the quality of conclusions drawn from the mixed data, and inference transferability is “the degree to which the conclusions from an MM study may be applied to other settings, people, time periods, contexts, and so on” (Teddlie & Tashakkori 2009, 27). Onwuegbuzie and Johnson (2006, 52), in respect of MM studies, deploy the validity terms representation (capturing the participants’ experiences in multiple ways), legitimation (“credible, trustworthy, dependable, transferable, and/or confirmable” findings and inferences) and integration (connection of the two types of data), which may all “plague” the validity of a research. Although the alternative terms presented here are extremely accessible, and they aid in grasping the essence of validity, I will persist with the concept ‘validity’ because of its universality in research.

Validity is always a matter of degree and truth is relative. As Bergman (2008, 47) notes, “there is no single valid description of a situation because validity is a matter of perspective”, implying that several valid descriptions of the same phenomenon may coexist. To reach the most valid, comprehensive account of the phenomenon under scrutiny necessitates dealing with potential validity threats in different phases of the research. Therefore, Creswell and Plano Clark (2011, 239) define validity in MM research as “employing strategies that address potential issues in data collection, data analysis and the interpretations that might compromise the merging or

connecting the quantitative and qualitative strands of the study and the conclusions drawn from the combination”.

A proficient researcher attempts to ensure the quality of the data, results and their interpretation. Triangulation in which the data is gathered from different sources is one such validity approach, also used in this research (see 5.1). Bergman (2008, 32) highlights, however, that using triangulations to increase validity and provide complementary information should be perceived as “investigative strategies that offer evidence to inform judgements” rather than “techniques that provide guaranteed truth or completeness”. Triangulation as qualitative research in general is not free of invalidity issues. For instance, Seliger and Shohamy (1989) note: when the data are not collected in natural environments, or when data are not representative enough, there are validity issues to consider.

In the following, I will discuss some validity threats proposed by Creswell and Plano Clark (2011) and Onwuegbuzie and Burke Johnson (2006) in issues regarding data collection, data analysis and interpretation in this study.

Data collection

In order to draw meta-inferences (see Onwuegbuzie & Burke Johnson 2006, 56) between the quantitative and qualitative data sets and to avoid inappropriate sample sizes and inadequate participants, large sample sizes for the quantitative and smaller samples for the qualitative inquiries should be used in addition to choosing individuals for the qualitative follow-ups from the larger quantitative sample (Creswell & Plano Clark 2011, 242). This is exactly what was done in this study with all participants: teachers, pupils and parents. The sample size for the semi-structured CLIL assessment survey questionnaire was fairly large (n=42–109), and the following theme interviews were implemented with a much smaller number of volunteers from the original participant group (n=7-20).

The computer simulation data were combined from two different piloting rounds (n=146 for pupils and n=39 for parents) for reasons of convenience, although the number of participating pupils was adequate even in separate rounds (see Table 21). The number of parents' questionnaires was a disappointment and caused by coding problems that were outside my sphere of influence (see 5.4.2.2). Despite fairly large sample sizes, generalisability and transferability are validity challenges in this research.

Any tentative generalisations should be made with caution because this study consists of cases that only give information about certain circumstances. It has to be remembered that the parents and pupils in the initial CLIL assessment survey and following theme interviews only represent the two participant schools and the obtained results are not transferrable as such to other CLIL circumstances. As Mustaparta and Tella (1999, 43) point out in their call for more CLIL research, the transferability of CLIL research findings in Finland is challenged due to the enormous diversity of CLIL implementation models – a study conducted in one CLIL school may not necessarily apply to another.

Regardless of this obvious, yet foreseen and unavoidable, shortcoming there are lessons to be learned and reformations to be considered for each CLIL-providing educational establishment. The participating teachers, however, are more representative of Finnish CLIL classroom teachers, although the sample size of teachers could have been larger. It is commonly known that teachers are often very busy with their daily routines and therefore sometimes reluctant to participate in extra activities. Therefore, I am very grateful to every participating, and especially interviewed teacher. The teacher sample was predominantly random, because most CLIL teachers in the sample were reached through an Internet questionnaire which should increase validity. A proportion of them represented the two research schools (town school and city school), the choice of which was not random. This could be seen as a validity threat, but also as an advantage, because they were able to provide an insight into their school-specific CLIL assessment practices.

As to the questionnaire design from a theoretical viewpoint, I did not focus on either of the traditional assessment purposes, formative or summative. This could be seen as a deficit of the study. My aim was to collect any possible instances of language assessment in CLIL without giving any preference for either one. Another reason for this is the comprehensive assessment approach reflected in the NCC (2004), which is a normative document. Furthermore, the level of TL exposure was not queried or differentiated either in the teacher questionnaires, which would have given more specified information of the studied phenomenon. Nor were more specific information of the uses and characteristics of the assessment methods asked which is notable in the analysis in expressions such as ‘the use of recordings remained relatively unclear’. Such specifications could be the subject of further research, while this study served as a general opening of assessment research in primary CLIL.

My role as a researcher was not entirely unambiguous. I am known in and have connections to both research schools – a fact that may or may not have affected the results positively or negatively. No marked signs pointing to either direction were detected. This validity threat was addressed by administering anonymous questionnaires. The interviews of pupils and parents were conducted in a pleasant, relaxed atmosphere due to the fact that I was not a complete stranger. However, I kept research integrity strictly in mind. Interviews generally may also pose a validity threat, because the interviewed people might have the tendency to give “socially desirable rather than honest responses” to the researcher (Bergman 2008, 33). All the interviews were based on voluntariness; the interviewees had thus announced their willingness to participate themselves. As a result, it is assumable that they had a mission to make a difference in the CLIL assessment practices of their school, or opinions to express, none of which were judged or criticised.

The fairly lengthy interview sessions of adults elicited a massive data bank, whereas the much shorter sessions with pupils were compensated by a larger number of them. To increase the validity of pupils’ interviews, I made comprehension checks to ensure that, for example, they understood the key concept “assessment”, but there were also factors that may have influenced

pupils' willingness to give profound replies. For example, the class teacher of CS 5th graders was absent on the day of pupil interviews, and the substitute teacher had unexpectedly fallen ill, so the class was watching a film instead of normal studying. The interviewed pupils had to come to the interview in the middle of the film and were anxious to get back to the classroom which may have had a negative impact on the interviews.

The questionnaires in the first research phase, the CLIL assessment survey, were pre-tested with pupils who were younger than the actual research group; the adult questionnaire was reviewed by an outsider, which increases validity. The anonymity of all questionnaires most likely had a positive impact on the credibility of the results. A weakness related to computer simulation questionnaires was that it was not possible or sensible to pre-test the simulation questionnaires, because the topic would not have been familiar to the pre-testers without the actual simulation performance. The initial questionnaire was, however, peer-reviewed.

The questionnaires in the two simulation phases were almost, but not totally, identical; some of the tasks (e.g. the mind map) were omitted or changed. The wording was slightly altered for the second piloting round on the basis of data analysis in the first round which indicated that pupils may not have fully understood all the terms in the questionnaire. These terms were explained and exemplified in the second round, but that did not make a noticeable difference in the results. The terms were probably too abstract for most children. Furthermore, it was necessary to create a slightly different questionnaire for the 5th graders who piloted the simulation for the second time in order to collect information on how their simulation experience changed when compared to the first one.

There were also minor differences in conducting the simulation video interviews, due to the pupils not simultaneously completing the simulation. The fact that the simulation video interviews were partly conducted by external assistants and partly by me is both a negative and positive issue: it might have been easier for pupils to express their true opinions to an outsider, but when several people conduct interviews, the questions tend to fluctuate regardless of a prewritten outline. The weight of the video interviews in the piloting of simulation 1, however, was confirmatory, because similar types of questions were asked in the actual questionnaire. The video interviews, then, served as indicators of instant sentiments and allowed a more fluent means of self-expression than in writing.

Pupils taking part in the simulation experiments were aware of the pioneering nature of the project. This may have caused the Hawthorne effect which denotes that the participants become more motivated due to the research factor and as a result, they may alter their behaviour which influences the results drawn from the study (Seliger & Shohamy 1989). This is a validity threat which cannot be eliminated at this stage of piloting. Only after the assessment method is established, can more trustworthy results of its appropriateness be obtained. On the other hand, some participants may have been more critical towards the simulation concept due to the piloting – they knew that the concept was being developed and piloted and therefore not in its final stage.

Data analysis

The quantitative methods preceded the qualitative ones in the data collection, and the obtained data as a whole are extensive. The data analysis occurred in a similar order. While analysing the data, I attempted, whenever possible, to clarify inconsistencies and obscurities in the data. I wrote by email additional clarification requests on, for instance, curriculum-related matters to two interviewees and received illuminations for those issues. I also discussed the extraordinary mathematics-related finding with the given CLIL class teacher attempting to find out reasons for the protruding results; that is, why that particular class in comparison to other classes did not find mathematical tasks that difficult. This discussion provided some possible explanations to this mismatching finding. Furthermore, I kept track of the developments in CLIL group tests by maintaining occasional contact with the given teacher.

I offered the interviewed adults the possibility to read their transcripts, adjust their replies and to check whether my transcription was correct. Only three took that opportunity, and two changed the original interview slightly after reconsidering, for instance, teacher qualification issues. Pupils were not offered this opportunity although in hindsight, there was no specific reason for not doing so. I was able to identify both stronger and weaker themes in the data. The strong themes per se are intensive, extensive and frequent enough to fill in the prevailing research gap, to answer the research questions and to comprise a solid research argument. Consequently, I considered follow-ups to gain further elucidation of weaker themes as unnecessary for the validity increase of this research as suggested in Creswell & Plano Clark (2011).

Interpretation of data

Researcher bias is one form of validity threat. As a member of the research group, I possess an emic view, i.e. an insider's view to the research topic which is inherent in qualitative research (see Onwuegbuzie & Burke Johnson 2006, 58). In addition to that, in order to draw quality inferences, also an outsider's etic view inherent in quantitative data should be adopted when combining both types of data and drawing meta-inferences. Peer reviewing is one method to operate with inside–outside views and to increase validity. This report was reviewed by both insiders and outsiders in the field of CLIL or language assessment.

As a researcher, I have pursued the most objective, accurate and descriptive account of the phenomenon under scrutiny, assessment in CLIL, simultaneously acknowledging that my own background as a CLIL class teacher and EFL teacher may cause unintended bias towards language dimension in CLIL. For that reason, to avoid any bias when reporting the research results, I preserved strict objectivity in reporting, and included both positive and negative issues and stakeholder perspectives when available.

Significance of the research

The ultimate aim of scientific research is to produce new, valid information that expands, deepens, enlightens or even changes the prevailing knowledge of the topic. This research is an attempt to start filling in the obvious research gap in CLIL assessment. It has succeeded in doing that in so far as it has produced a significant amount of new information pertaining to the assessment practices current at the time of the research (2012) in bilingual primary CLIL education in Finland from the perspectives of three stakeholders (teachers, pupils and parents). The study has also reported the pioneering experimentations of LangPerform computer simulations at primary level in CLIL which upgrades the results to a genuinely developmental level and increases their practical value.

As discussed above, the obtained results are highly context-specific. Despite that, a very significant general finding is that CLIL instruction and assessment is not based on CLIL curriculum because there is none available. The findings may also have relevance in schools that have identical or resembling profiles as the two research schools (see Table 17); the results of this study indicate the caveats in CLIL assessment and show the most crucial points of development. In particular, the findings of the CLIL assessment survey concerning the comments, ideas, wishes and observations of pupils and parents should evoke thoughts and plans for developing feedback and reporting methods in near future. Furthermore, promoting alternative assessments, shifting towards educational technology and taking full advantage of its affordances as well as engaging students and their parents in mutually beneficial converse with teachers are issues worth considering in every branch of education.

In general, and at a discipline-specific level, the results of this research contribute significantly to the field of very scarce CLIL assessment research both theoretically and empirically as well as shed light on the roles of language in CLIL. This study is a trailblazer in primary CLIL assessment research. These contributions and issues related to them are further discussed in the final section concluding this report.

8.5 Conclusion and implications to CLIL

This research was the first to examine the assessment practices of pupils' English language proficiency in Finnish primary CLIL education. The obtained results, unfortunately, are not very encouraging. It appears that we still have a considerably long journey before we are able to reach the peak of the Finnish CLIL trajectory. Solid and grounded assessments are intrinsic to quality education, but this study suggests that the assessment of primary CLIL learners' English proficiency is often infrequent, incidental, implicit and based on impressions rather than evidence or the curriculum. Importance is attached to assessment in CLIL contexts, but it seems that this stance is not reflected in the CLIL field as necessitated by the National Core Curriculum for Basic Education (NCC 2004), which sets a few unequivocal norms and standards for Finnish CLIL

provision. The majority of both pupils and parents in the two research schools expressed an unambiguous wish to receive more assessment information and feedback on the level of and progression in the English proficiency of children in CLIL subjects.

However, the outcome of this research rather presents itself as an opportunity for advancement than as a rebuke; it is time to take action in CLIL assessment and develop further what has already been achieved. This research hopefully provides inspiration for that development work at municipal, school and individual CLIL teacher level. Assessment issues should be re-evaluated from its fundamentals, and sustainable, informative, resource-efficient and flexible assessment schemes need to be contemplated. In this section, I will present my suggestions for how these advancements in CLIL assessment could be yielded.

A number of challenges and obstacles for the assessment in CLIL were identified in this study, but also several good practices that can serve as a reference for future development of CLIL assessment. Because this is a developmental study, the aim of which is to contribute to the advancement of CLIL assessment, I will offer my suggestions for how to divert this state of affairs towards a more solid and grounded assessment practice. Returning to the educational tripod metaphor I used in the introduction, these implications reinforce the legs of assessment and objectives in the tripod, but also touch upon instruction indirectly. The implications are divided into three facets: a) the fundamentals of CLIL assessment, b) principles of adequate CLIL assessment and c) recommendable CLIL assessment methods. The three facets are presented as concisely as possible in the form of a 'points worth noting' list.

A. FUNDAMENTALS OF CLIL ASSESSMENT

The fundamentals of CLIL assessment refer to the most basic and constitutive aspects of CLIL assessment that compose the solid foundation and absolutely necessary basis for assessment in CLIL. The first point, CLIL curriculum, is entirely based on the current NCC (2004) and therefore a norm. The curriculum reform in 2016, however, may alter these requirements, as I have shown in this report.

1) CLIL curriculum which minimally defines

- the extent of English language exposure in subjects other than language
- the subjects which follow the CLIL curriculum
- the contents (topics) instructed through the foreign language
- the desired level of English in
 - listening comprehension
 - speaking
 - reading comprehension
 - writing
 - cultural skills

The NCC does not specify at which point the desired language level in CLIL should be defined – it merely states that the objective is to reach a higher language proficiency than is normally reached through ordinary language instruction. My suggestion is that in basic education, the desired level is mentioned at the end of 3rd, 6th and 9th grade. The CEFR codes are appropriate for this purpose.

2) The CLIL assessment scheme which states how multifaceted, evidence-based assessment information on pupils' language proficiency is gathered and conveyed to the stakeholders and how the objectives and assessment communicate. It includes the following but leaves room for individual practices:

- mutually agreed principles of assessment (formative and summative)
- the roles of subject-specific and casual-type EFL English in the assessment scheme
 - how the special characteristics of CLIL English are distinguished from EFL assessment
 - the person responsible of CLIL teaching is responsible of CLIL assessment
- collective methods of conveying assessment information to pupils and their parents, e.g.:
 - school year reports
 - development discussion procedures
 - consulting hours

I recommend that 'can do' statements are adopted for reporting rather than, or in addition to, numerical assessments and pupils' self-assessments. CLIL report could be an attachment to the conventional report.

3) Elucidation of school-specific everyday CLIL implementation to stakeholders

- the curriculum, objectives and basic principles of CLIL instruction
- e.g. information events, days of open doors, thematic occasions, leaflets, access to the CLIL curriculum

I recommend introducing the primary CLIL objectives of the given school year in parental evenings and also in written form (a leaflet, Wilma-message, email).

4) Adequate language and CLIL competences of CLIL class teachers

- knowledge of principles of CLIL instruction
- knowledge of second language learning
- knowledge of second language assessment

It is highly recommendable that CLIL class teachers are fluent in the TL and possess language degrees. I also urge teacher recruiters to familiarise themselves with the obligating ordinance

25/011/2005 by the Ministry of Education (2005) on the language qualifications of CLIL teachers, which must be complied with.

B. PRINCIPLES OF ADEQUATE CLIL ASSESSMENT

These are the primary aspects that contribute to quality assessments in CLIL at any level of education.

1) Dual focus

- Assessment is focussed on both content and language.
- The language focus is on language use instead of language knowledge.
- Assessment is centred both on product (language) and process (working skills, affective factors).

2) Multifaceted assessment methods

- Several age-appropriate assessment methods are used in addition to observation pertaining to all four language skills and cultural skills

3) Evidence-based inferences of English proficiency

- Evidence gathering provides data which can be examined by stakeholders.
- The inferences of the level of English proficiency and its progress over time are based on the gathered evidence.
- The linguistic inferences are compared to the linguistic objectives in the CLIL curriculum.

4) Criterion-referenced inferencing

- The inferences are substantiated by pre-defined criteria.
- Everyone involved in the assessment process is aware of the criteria and the prerequisites needed for achieving a certain score or level.
- The criteria are preferably co-constructed together with the learners.
- Rubrics serve as visuals for the criteria.

5) Frequency and sufficiency of assessment information and feedback

- Provision of assessment information is regular and frequent enough to suffice for the stakeholders and to convey information that is recent enough to scaffold and reinforce learning.
- The younger the learners are, the more they should be given immediate, positive, personal and reassuring feedback on their emerging learner language.

C. RECOMMENDED CLIL ASSESSMENT METHODS

There are a number of recommendable assessment methods, traditional and innovative, that are suitable for CLIL. These examples, which are not listed in any order of preference, represent only part of them, but are particularly recommendable. I will justify these choices briefly.

1) Collaborative testing

Collaborative testing (group tests) are model examples of knowledge construction in social interaction. Pupils can train their group work skills, use the strengths of each group member and learn from each other while producing content knowledge and using the target language for a meaningful purpose. When the pressure of good performance is shared, and the work occurs in the zone of proximal development of pupils, test anxiety is more likely to be reduced. Combined with individual tests, collaborative testing serves as an additional springboard. A collective grade is recommended.

2) Technology-based language testing

Technology is an essential part of modern life, the potential of which has not yet been fully acknowledged and exploited. The computer simulations experimented with in this study proved to be very promising, although they still need adjustment to be applicable to young language learners. Currently, these simulations are not publicly available, but laptop computers, mobile device (phones and tablets) and even digital voice recorders or video cameras are within reach. These devices can be utilised for documenting various projects or performances and recording speech or songs. Digital data can easily be stored on a memory stick, CD, DVD or a hard drive file. A pupil's personal memory stick, a file in the classroom or the school computer may work as a digital portfolio. Such data are transferrable and it is possible to duplicate the data to provide the parents with a copy as evidence of language growth or as a record.

3) Task-based language testing

The simulations contain tasks integrating content and language. Tasks are more meaningful for pupils than exercises or copying texts from a PowerPoint presentation or blackboard because they allow for much more than just replicating knowledge. Tasks encourage the use of language for problem solving, creativity, application, analysis and evaluation. Therefore, Bloom's revised taxonomy provides a helpful tool for task design. Task characteristics and conditions define their difficulty and applicability to CLIL, but successful task completion requires linguistic scaffolding. One method combining technology-based and task-based language testing is the WebQuest¹⁸ concept, which denotes information retrieval from the Internet and performance of various tasks defined

¹⁸ See e.g. <http://webquest.org/> for general information and <http://prezi.com/mtpqbfh2gx/the-use-of-web-quests-in-clil-settings/> for CLIL-specific information on WebQuests.

by the teacher – or even learners. WebQuests are very suitable for primary-aged pupils, and they can be performed as group tests. WebQuests, naturally, need to be designed and composed in advance. On the Internet, there are also shared, ready-for-use WebQuests available.

4) Portfolios

Portfolios were the least used assessment methods in CLIL, yet I perceive them to be one of the most flexible and relevant means of gathering long-term evidence of what pupils can do with their language proficiency – each within his/her own level. It appears that teachers need to find their own models for portfolio work. Portfolios may range from digital portfolios to shoe boxes or notebooks. The essential element in portfolios is that, similar to computer simulations, they represent experiential learning which stresses the application of learned knowledge in new situations (knowing how) and the reflection of knowledge and experiences which in turn lead to genuine learning. Language portfolios may consist of several parts which accentuate the language history of the learner, self-assessment of current proficiency and representative samples of the language use. In CLIL, these samples would then pertain to work that integrates both content and language. Portfolios are concrete, flexible (because the owner is allowed to decide which work and information is included) and age-appropriate.

5) Projects

Projects as such are not a means of assessment but they provide plenty of material for assessment and may incorporate the previously mentioned aspects: criterion-referenced inferencing in form of rubrics, collaborative work, technology and tasks. The end product – depending on its nature – can be stored in a digital portfolio, a folder or equivalent. Another benefit of project work is that one can also monitor and assess the process (working skills) in addition to the product.

Implications for further research

The outcome of this research was more than anticipated in terms of obtained results and their significance for the field of CLIL research. Prospective implications for further research, especially linguistic research, were also generated and mentioned along this report whenever relevant. Academic language in CLIL and its development is definitely an issue that deserves closer scrutiny. There are some research and pragmatic projects already addressing this issue, but how to combine curriculum development and the enhancement of academic language in CLIL is a new territory worth exploration. This one of my next projects on which I will concentrate in collaboration with the University of Denver, Colorado. I would also encourage looking at the extent of language exposure in CLIL in relation to the achieved outcomes. Regarding research

into assessment, this was the first concerning young language learners, and hopefully there are many more studies to follow. I will continue working on the CLIL and EFL portfolio experiment data I have gathered, but questions of how to increase children's linguistic awareness through assessment and how to use ICT in CLIL assessment are interesting ideas to tackle.

Epilogue

This research was conducted in the hope of positively influencing CLIL assessment and elucidating the nature of it in Finnish primary CLIL. I hope to have portrayed an overall picture of the assessment of pupils' English language proficiency in bilingual content instruction CLIL, identified challenges and obstacles undermining the implementation of assessment in CLIL and also, by discussing the implications of this study for CLIL, led the way for more adequate, satisfying and systematic assessment practices.

The key approach to assessment in CLIL is to organise occasions for the pupils to actually use language meaningfully. Furthermore, it is important to make a distinction between EFL and subject-specific English. In order for pupils to cope with English as a tool or a medium in content study and express themselves and their content mastery in assessment situations, they need more specialised language than casual, everyday English which is the object of study in EFL.

One quintessential purpose of formative assessment is to enhance learning. Assessment is not intended to stress or indicate pupils' deficiencies or inabilities, or to punish them for not studying or meeting someone's expectations. Assessment exemplifies the abilities and skills of pupils to use the language in CLIL; it primarily reveals what they can do with the language and secondarily indicates how they can improve their language skills. The emphasis is thus on strengths and improvement instead of weaknesses or flaws. In the reports at the end of the school year (summative assessment), traits of continuous assessment can also be attached. Descriptions instead of grades could encourage pupils that do not reach the criteria for good grades but excel in relation to their own skills.

A new, fairly 'revolutionary' perspective in assessment is that teachers could also be assessed by pupils - an issue that was proposed in the NCC draft (2014, 39): "Pupils are encouraged to give constructive feedback to each other and to the teacher. With the following code of practice I would like, with the most sincere intentions, to encapsulate the essence of assessment in primary CLIL.

***Recommendation for primary CLIL class teacher's
code of assessment practice***

- Familiarise yourself with the core curriculum and local CLIL curriculum.
- Assess both content and language.
- Assess the performance and language use, not the child.
- Let the child know what s/he *can* do with the language in CLIL subjects rather than what s/he cannot.
- Convey explicit assessment information to parents as well.
- Favour personal contact but also provide written information.
- Do not settle for observation only but gather versatile evidence of the language use in CLIL study.
- Rely on evidence as the basis for your (and other stakeholders) inferences about language proficiency.
- Remember all four basic language skills, cultural knowledge as well as affective factors.
- Set the bar high for language use – go beyond word and single sentence level.
- Organise plenty of opportunities for pupils to use the language - it is easier to assess active than passive bilingual performance.
- Use the CLIL curriculum and the CEFR as a reference.
- Approach assessment from a positive perspective; there are multiple ways to praise a child.

References

- AECLIL 2013. *Assessment and Evaluation in CLIL*. Retrieved July 16, 2013 from www.aecilil.net.
- Alderson Charles J. 2005. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Continuum.
- Allen Patrick, Swain Merrill, Harley Birgit & Cummins Jim 1990. Aspects of classroom treatment: toward a more comprehensive view of second language education. In B. Harley, J. Cummins, M. Swain & P. Allen (eds). *The Development of Second Language Proficiency* (pp. 57-81). Cambridge: Cambridge University Press.
- Anderson Lorin W., Kraftwohl David R. & al. (eds) 2001. *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Andersson Theodore & Boyer Mildred 1978. *Bilingual Schooling in the United States*, 2nd ed. Austin, TX: National Educational Laboratory Publishers.
- Asikainen Tuula, Bertaux Pat, Frigols Martin María Jesús, Hughes Sue, Marsh David & Mehisto Peeter (eds) 2010. *Talking the Future 2010-2020. CNN Foresight Think Tank Report*. University of Jyväskylä: CLIL Cascade Network.
- Bachman Lyle F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman Lyle F. 2004. *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Bachman Lyle F. 2007. What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner & C. Doe (eds). *Language Testing Reconsidered* (pp. 41-72). Ottawa: University of Ottawa Press.
- Bachman Lyle F. & Palmer Adrian 1996. *Language Testing in Practice*. Oxford: Oxford University Press.
- Bachman Lyle F. & Palmer Adrian 2010. *Language Testing in Practice*. Oxford: Oxford University Press.
- Baker Colin 2011. *Foundations of Bilingual Education and Bilingualism*, 5th ed. Bristol: Multilingual Matters.
- Barbero Teresina 2007. CLIL in scientific fields: from teaching language to learning activities. In D. Marsh & D. Wolff (eds). *Diverse Contexts – Converging Goals. CLIL in Europe* (pp. 287-298). Frankfurt am Main: Peter Lang.
- Barbero Teresina & Järvinen Heini-Marja 2009. Assessment in CLIL. In H-M. Järvinen (ed.) *Handbook Language in Content Instruction* (pp. 106-111). LIC Project Publication.
- Basic Education Act 1988/628. Retrieved May 24, 2014 from <http://www.finlex.fi/en/laki/kaannokset/1998/en19980628.pdf>.
- Bentley Kay 2010. *The TKT Teaching Knowledge Test Course CLIL Module*. Cambridge: Cambridge University Press.
- Bergman Manfred Max 2008. *Advances in Mixed Methods Research*. Los Angeles, CA: Sage.
- Berry John & Nyman Melvyn A. 2002. Small-group assessment methods in mathematics. *International Journal of Mathematical Education in Science and Technology* 33 (5): 641-649.
- Bialystok Ellen 1978. A theoretical model of second language learning. *Language Learning* 28 (1): 69-83.
- Bialystok Ellen 2001. *Bilingualism in Development. Language, Literacy and Cognition*. Cambridge: Cambridge University Press.
- Binkley Marilyn, Erstad Ola & Herman Joan 2012. Defining twenty-first century skills. In P. Griffin, B. McGaw & E. Care (eds). *Assessment and Teaching of 21st Century Skills* (pp. 17-66). Boston, MA: Springer.
- Birenbaum M., Breuer K., Cascallar E., Dochy F., Dori Y., Ridgway J., Wiesemes, R. & Nickmans, G. 2006. A learning integrated assessment system. *Educational Research Review* 1: 61-67.

- Björklund Siv 1997. Immersion in Finland in the 1990s. In R. K. Johnson & M. Swain (eds). *Immersion Education: International Perspectives* (pp. 85-102). Cambridge: Cambridge University Press.
- Black Paul & Wiliam Dylan 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice* 5 (1): 7-73.
- Black Paul & Jones Jane 2006. Formative assessment and the learning and teaching of MFL: sharing the language learning road map with the learners. *Language Learning Journal* 34: 4-9.
- Block David 2003. *The Social Turn in Second Language Acquisition*. Edinburgh: Edinburgh University Press.
- Bonnet Andreas 2012. Towards an evidence base for CLIL: how to integrate qualitative and quantitative as well as process, product and participant perspectives in CLIL research. *International CLIL Research Journal* 1 (4): 66-78.
- Bot de Kees 2007. Language teaching in the changing world. *The Modern Language Journal* 91 (2): 274-276.
- Boud David 2000. Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education* 22 (2): 151-167.
- Boud David & Molloy Elizabeth 2013. *Feedback in Higher and Professional Education: Understanding and Doing It Well*. New York, NY: Routledge.
- Bovellan Eveliina (forthcoming). *Teachers' Beliefs about learning and language as reflected in their views of teaching materials for content and language integrated learning (CLIL)*. Jyväskylä: University of Jyväskylä.
- Braun Virginia & Clarke Victoria 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3: 77-101.
- Brindley Geoff 1992. Assessing achievement in a learner-centered curriculum. In C. Alderson & B. North (eds). *Language Testing in the 1990s: The Communicative Legacy* (pp. 153-166). London: Macmillan.
- Brisk María Estela 2010. Bilingual Education. In E. Hinkel (ed.). *Handbook of Research in Second Language Teaching and Learning*. New York: Routledge.
- Broadfoot Patricia & Black Paul 2004. Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education* 11 (1): 7-27.
- Brookhart Susan M. 2011. Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice* 30 (1): 3-12.
- Brown James D. & Hudson Thom 1998. The alternatives in language assessment. *TESOL Quarterly* 32 (4): 653-675.
- Brown James D., Hudson Thom, Norris John & Bonk William J. 2002. *An Investigation of Second-Language Task-Based Performance Assessments*. Honolulu, HI: University of Hawai'i Press.
- Busch Brigitte 2011. Trends and innovative practices in multilingual education in Europe: an overview. *International Review of Education* 57 (5-6): 541-549.
- Butler Ruth & Nisan Mordecai 1986. Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology* 78 (3): 210-216.
- Byrnes Heidi 2008. Assessing content and language. In E. Shohamy & N. H. Hornberger (eds). *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment* 2nd ed. (pp. 37-52). Boston, MA: Springer.
- Canale Michael & Swain Merrill 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1 (1): 1-47.
- Canale Michael 1983. From communicative competence to communicative language pedagogy. In J.C. Richards & R.W. Schmidt (eds). *Language and Communication* (pp. 2-28). Harlow: Longman.
- Carroll John Bissell 1968. The psychology of language testing. In A. Davies (ed.). *Language Testing Symposium. A Psycholinguistic Approach* (pp. 46-69). London: Oxford University Press.

- CEFR 2001. *Common European Framework of Reference for Languages: Learning teaching, assessment*, 3rd ed. Cambridge: Cambridge University Press.
- Celce-Murcia Marianne, Dörnyei Zoltán & Thurrell Sarah 1995. Communicative competence: a pedagogically motivated model with content specifications. *Issues in Applied Linguistics* 6 (2): 5-35.
- Chamot Anna Uhl 2007. Accelerating academic achievement of English language learners: a synthesis of five evaluations of the CALLA model. In J. Cummins & C. Davison (eds). *International Handbook of English Language Teaching* (pp. 317-331). Boston, MA: Springer
- Chamot Anna Uhl & O'Malley Michael J. 1987. The Cognitive Academic Language Learning Approach: a bridge to the mainstream. *TESOL Quarterly* 21 (2): 28-50.
- Chapell Mark S., Blanding Z. Benjamin, Silverstein Michael E., Takahashi Masami, Newman Brian, Gubi Aaron & McCann Nicole 2005. Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology* 97 (2): 268-274.
- Chapelle Carol A. 2001. *Computer Applications in Second Language Acquisition. Foundations for teaching, testing and research*. Cambridge: Cambridge University Press.
- Chapelle Carol A. 2004. Technology and second language learning: expanding methods and agendas. *System* 32: 593-601.
- Chapelle Carol A. & Douglas Dan 2006. *Assessing Language through Computer Technology*. Cambridge: Cambridge University Press.
- Chin Jeffrey, Dukes Richard & Gamson William 2009. Assessment in Simulation and Gaming: a review of the last 40 years. *Simulation & Gaming* 40 (4): 553-568.
- Chomsky Noam 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: M.I.T. Press.
- Clapham Caroline 2000. Assessment for academic purposes: where next? *System* (28): 511-521.
- CLILA 2013. *The Research Project CLIL Learner Assessment*. Retrieved 18 April 2013 from ph-weingarten.de/clila-forschungsprojekt/en/.
- Cohen Andrew D. 1994. *Assessing Language Ability in the Classroom*, 2nd ed. Boston: Heinle & Heinle Publishers.
- ConCLIL 2013. *Language and Content Integration: Towards a Conceptual Framework*. Retrieved December 12, 2013 from conclil.jyu.fi/.
- Cook Guy 1997. Language play, language learning. *ELT Journal* 51 (3): 224-231.
- Cook Vivian 1993. *Linguistics and Second Language Acquisition*. Basingstoke: Macmillan.
- Cook Vivian 1999. Going beyond the native speaker in language teaching. *TESOL Quarterly* 33 (2): 185-209.
- Cook Vivian 2013. *Krashen's Comprehension Hypothesis Model of L2 Learning. Key Issues in SLA*. Retrieved August 29, 2013 from homepage.ntlworld.com/vivian.c/SLA/Krashe.n.htm.
- Cook Vivian & Wei Li 2009. Introduction: applied linguistics and language teaching in the twenty-first century. In V. Cook & L. Wei (eds). *Contemporary Applied Linguistics. Language Teaching and Learning Volume 1* (pp. 1-9). London: Continuum.
- Cormier Marianne & Turnbull Miles 2009. Une approche littératiée: apprendre les sciences et la langue en immersion tardive (A literacy approach: learning science and language in late immersion). *The Modern Canadian Language Review* 65 (5): 817-840.
- Coyle, Do, Hood, Philip & Marsh, David 2010. *CLIL Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Creswell John W. & Plano Clark Vicki L. 2011. *Designing and Conducting Mixed Methods Research*, 2nd ed. Thousand Oaks, CA: Sage.
- Crookall David 2002. Editorial: simulation in language learning. *SIMULATION & GAMING* 33 (3): 273-274.

- Cumming Alister 2008. Assessing oral and literate abilities. In E. Shohamy & N.H. Hornberger (eds). *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment*, 2nd ed. (pp. 3-17). Boston, MA: Springer.
- Cummins Jim 1980a. The cross-lingual dimensions of language proficiency: implications for bilingual education and the optimal age issue. *TESOL Quarterly* 14 (2): 175-187.
- Cummins James 1980b. The construct of language proficiency in bilingual education. In J.E. Alatis (ed.). *Current Issues in Bilingual Education* (pp. 81-103). Washington, D.C.: Georgetown University Press.
- Cummins Jim 1982. Tests, achievement, and bilingual students. *FOCUS* 9: 2-9.
- Cummins Jim 1984. Wanted: a theoretical framework for relating language proficiency to academic achievement among bilingual students. In C. Rivera (ed.). *Language Proficiency and Academic Achievement* (pp. 2-19). Clevedon: Multilingual Matters.
- Cummins Jim 2008. BICS and CALP: empirical and theoretical status of the distinction. In E. Shohamy & N. H. Hornberger (eds). *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment*, 2nd ed. (pp. 71-83). Boston, MA: Springer.
- Cummins Jim & Man Yee-Fun Evelyn 2007. Academic language: what is it and how do we acquire it? In J. Cummins & C. Davison (eds). *International Handbook of English Language Teaching* (pp. 797-810). Boston, MA: Springer
- Dakowska Maria 2013. *Why does CLIL work? An immodest attempt at addressing the question*. A plenary talk paper at the CLIL Conference on Modernizing Educational Practice, Perspectives in Content and Language Integrated Learning, Ustrón, Poland, April 4-5, 2013.
- Dalton-Puffer Christiane 2007a. *Discourse in Content and Language Integrated (CLIL) Classrooms*. Amsterdam: John Benjamins.
- Dalton-Puffer Christiane 2007b. Academic language functions in CLIL environment. In D. Marsh & D. Wolff (eds). *Diverse Contexts – Converging Goals. CLIL in Europe* (pp. 201-209). Frankfurt am Main: Peter Lang.
- Dalton-Puffer Christiane 2008. Outcomes and processes in content and language integrated learning (CLIL): current research from Europe. In W. Delanoy and L. Volkman, (eds.). *Future Perspectives for English Language Teaching* (pp. 139-157). Heidelberg: Carl Winter.
- Dalton-Puffer Christiane 2011. Content and language integrated learning: from practice to principles? *Annual Review of Applied Linguistics* 31: 182-204.
- Dalton-Puffer Christiane & Nikula Tarja 2006. Pragmatics of content-based instruction: teacher and student directives in Finnish and Austrian classrooms. *Applied Linguistics* 27 (2): 241-267.
- Dalton-Puffer Christiane, Nikula Tarja & Smit Ute 2010a. Charting policies, premises and research on content and language integrated learning. In C. Dalton-Puffer, T. Nikula & U. Smit (eds). *Language Use and Language Learning in CLIL Classrooms* (pp. 1-19). Amsterdam: John Benjamins.
- Dalton-Puffer Christiane, Nikula Tarja & Smit Ute 2010b. Language use and language learning in CLIL. Current findings and contentious issues. In C. Dalton-Puffer, T. Nikula & U. Smit (eds). *Language Use and Language Learning in CLIL Classrooms* (pp. 279-291). Amsterdam: John Benjamins.
- Dalton-Puffer Christiane & Smit Ute 2007. Introduction. In C. Dalton-Puffer & U. Smit (eds). *Empirical Perspectives on CLIL Classroom Discourse* (pp. 7-23). Frankfurt: Peter Lang.
- Darling-Hammond Linda 1994. Performance-based assessment and educational equity. *Harvard Educational Review* 64 (1): 5-30.
- Díaz Cobo Angel 2009. Assessment instruments for CLIL written production tasks. In D. Marsh, P. Mehisto, D. Wolff, R. Aliaga, T. Asikainen, M. J. Frigols-Martin, S. Hughes & G. Langé (eds). *CLIL Practice: Perspectives from the Field*. CLIL Cascade Network. Jyväskylä: University of Jyväskylä.
- DeKeyser Robert 1998. Beyond focus on form. In C. Doughty & J. Williams (eds). *Focus on Form in Classroom Second*

- Language Acquisition* (pp. 42-63). Cambridge: Cambridge University Press.
- DeKeyser Robert 2003. Implicit and explicit learning. In C. J. Doughty & M. H. Long (eds). *The Handbook of Second Language Acquisition* (pp. 313-348). Malden, MA: Blackwell.
- Dochy Filip 2001. A new assessment era: different needs, new challenges. *Research Dialogue in Learning and Instruction* 2: 11-20.
- Dörnyei Zoltán 2009. *The Psychology of Second Language Acquisition*. Oxford: Oxford University Press.
- Douglas Dan 2000. *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.
- Duff Patricia A. 2007. Qualitative Approaches to classroom research with English language learners. In J. Cummins & C. Davidson (eds). *International Handbook of English Language Teaching* (pp. 973-986). Boston, MA: Springer.
- eCALLMS 2013. *e-Learning Communities for Academic Language Learning in Mathematics and Science*. A project at Colorado University Denver. Personal communication with Dr. Nancy Commins, October 17, 2013.
- ECML 2013. *Language descriptors for migrant and minority learners' success in compulsory education*. Retrieved 26 December, 2013 from www.ecml.at/F6/tabid/696/language/en-GB/Default.aspx.
- Edenlobos Peter & Kubanek-German Angelika 2004. Teacher assessment: the concept of 'diagnostic competence'. *Language Testing* 21 (3): 259-283.
- Ellis Nick 1994. *Implicit and Explicit Learning of Languages*. London: Academic Press.
- Ellis Rod 2003. *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.
- Ellis Rod R. 2012. *Language Teaching Research and Language Pedagogy*. Chichester: Wiley-Blackwell.
- Ellis Rod, Loewen Shawn & Erlam Rosemary 2006. Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition* 28: 339-368.
- Ernst-Slavit Gisela & Mason Michele R. 2011. "Words that hold us up": teacher talk and academic language in five upper elementary classrooms. *Linguistics and Education* 22: 430-440.
- Eurydice 2006. *Content and Language Integrated Learning (CLIL) at School in Europe*. Brussels: Eurydice.
- Eurydice 2012. *Key Data on Teaching Languages at School in Europe 2012*. Brussels: Education, Audiovisual and Culture Executive Agency.
- Ewald Jennifer D. 2005. Language-related episodes in an assessment context: a 'small-group quiz'. *The Canadian Modern Language Review* 61 (4): 565-586.
- Feinstein Andrew Hale, Mann Stuart & Corsun David L. 2002. Charting the experiential territory. Clarifying definitions and uses of computer simulation, games, and role play. *The Journal of Management Development* 21 (10): 732-744.
- Fortanet-Gómez Inmaulada & Ruiz-Garrido Miguel F. 2009. Sharing CLIL in Europe. In M. L. Carrió-Pastor (ed.). *Content and Language Integrated Learning: Cultural Diversity* (pp. 47-75). Bern: Peter Lang.
- Fox Janna 2008. Alternative assessment. In E. Shohamy & N. H. Hornberger (eds). *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment*, 2nd ed. (pp. 97-109). Boston, MA: Springer.
- Fulcher Glenn & Davidson Fred 2007. *Language Testing and Assessment. An Advanced Resource book*. London: Routledge.
- Gallardo del Puerto Francisco, Gómez Lacabex Esther & García Lecumberri María Luisa 2009. Testing the effectiveness of content and language integrated Learning in foreign language contexts: the assessment of English pronunciation. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (eds). *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 63-80). Bristol: Multilingual Matters.

- Gattullo Francesca 2000. Formative assessment in ELT primary (elementary) classroom: an Italian case study. *Language Testing* 17 (2): 278-288.
- García Ofelia 1996. Bilingual Education. In F. Coulmas (ed.). *The Handbook of Sociolinguistics* (pp. 405-420). Oxford: Blackwell Publishers.
- García Ofelia 2009a. *Bilingual Education in the 21st Century. A Global Perspective*. Chichester: Wiley-Blackwell.
- García Ofelia 2009b. Education, multilingualism and translanguaging in the 21st century. In A. Mohanty, M. Panda, R. Phillipson & T. Skutnabb-Kangas (eds): *Multilingual Education for Social Justice: Globalising the local* (pp. 128-145). New Delhi: Orient Blackswan.
- Gebauer Sandra Kristina, Zaunbauer Anna C. M. & Möller Jens 2012. Erstsprachliche Leistungsentwicklung im Immersionsunterricht: Vorteile trotz Unterrichts in einer Fremdsprache? (Achievement development in the first language: benefits despite instruction in a foreign language?) *Zeitschrift für Pädagogische Psychologie* 26 (3): 183-196.
- Genesee Fred 1987. *Learning through Two Languages. Studies of Immersion and Bilingual Education*. Cambridge, MA: Newbury.
- Gottlieb Margo 2006. *Assessing English Language Learners. Bridges from Language Proficiency to Academic Achievement*. Thousand Oaks, CA: Corwin Press.
- Graddol David 2006. *English Next. Why global English may mean the end of 'English as a Foreign Language'*. British Council. Retrieved April 12, 2013 from britishcouncil.org/learning-research-english-next.pdf.
- Graaff de Rick & Housen Alex 2009. Investigating the effects of L2 Instruction. In M. H. Long & C. J. Doughty (eds). *The Handbook of Language Teaching*. Chisester: Wiley-Blackwell: 726-755.
- Graaff de Rick, Koopman Gerrit Jan, Anikina Yulia & Westhoff Gerard 2007. An observation tool for effective L2 pedagogy in content and language integrated learning (CLIL). *The International Journal of Bilingual Education and Bilingualism* 10 (5): 603-624.
- Gray David E. 2009. *Doing Research in the Real World*, 2nd ed. Los Angeles, CA: Sage.
- Griffin Patrick 2009. Teachers' use of assessment data. In C. Wyatt-Smith & J.J. Cunnning (eds). *Educational Assessment in the 21st Century* (pp.183-208). Boston, MA: Springer.
- Haag Nicole, Heppt Birgit, Stanat Petra, Kuhl Poldi & Pant Hans Anand 2013. Second language learners' performance in mathematics: disentangling the effects of academic language. *Learning and Instruction* 28: 24-34.
- Haataja Kim 2010. Das Konzept LangPerform. Entwicklung und Einsatz von Simulationsinstrumenten zur computermedialen Dokumentierung von (fremd-)sprachlichen Kompetenzen – innovativ und integrativ (The LangPerform concept. Development and implementation of simulation instruments for computer-assisted documenting of (foreign) language competences - innovative and integrative). *Jahrbuch Deutsch as Fremdsprache* 36: 183–199.
- Haataja Kim & Wewer Taina 2013. Elokuvapohjaisia tietokonesimulaatioita kokonaisvaltaisen kieli- ja kulttuurikasvatuksen tueksi – mutta miksi? (Film-based computer simulations in support of comprehensive linguistic and cultural education - but why?). In L. Pirkkalainen & P. Lounaskorpi (eds). *Löytöretkillä toisessa maailmassa Vol 1. Konneveden lukion verkkojulkaisu*. Retrieved May 31, 2013 from konnevedenlukio.onedu.fi/verkkojulkaisut/zine/33/article-365.
- Halliday M.A.K. 1979. Towards a sociological semantics. In C.J. Brumfit & K. Johnson (eds). *The Communicative Approach to Language Teaching* (pp. 27-45). Oxford: Oxford University Press.
- Halliday M.A.K. 1993. Towards a language-based theory of learning. *Linguistics and Education* 5: 93-116.
- Hamp-Lyons Liz 2007. The Impact of Testing Practices on Teaching: Ideologies and Alternatives. In J. Cummins & C. Davison (eds). *International Handbook of*

- English Language Teaching* (pp. 487-504). Boston, MA: Springer.
- Harjanne Pirjo & Tella Seppo 2008. Strong signals in foreign language education, with a view to future visions. In S. Tella (ed.). *From Brawn to Brain: Strong Signals in Foreign Language Education* (pp. 55-84). Helsinki: Yliopistopaino.
- Harlen Wynne 2005. Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal* 16 (2): 207-223.
- Hasselgreen Angela 2005. Assessing the language of young learners. *Language Testing* 22 (3): 337-354.
- Hasselgreen Angela, Kaledaité Violeta, Maldonado-Martin Natàlia & Pizorn Karmen 2011. *Assessment of young learner literacy linked to the Common European Framework of Reference for Languages*. Graz: European Centre of Modern Languages/Council of Europe.
- Hattie John 2003. *Teachers Make a Difference: What is the Research Evidence? Distinguishing Expert Teachers from Novice and Experienced Teachers*. Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne. Retrieved January 19, 2014 from www.acer.edu.au/documents/Hattie_TeachersMakeADifference.pdf.
- Hattie John & Timperley Helen 2007. The power of feedback. *Review of Educational Research* 77 (1): 81-112.
- Havnes Anton, Smith Kari, Dysthe Olga & Ludvigsen Kristine 2012. Formative assessment and feedback: making learning visible. *Studies in Educational Evaluation* 38: 21-27.
- Havranek Gertraud 2002. When is corrective feedback most likely to succeed? *International Journal of Educational Research* 37: 255-270.
- Heer Rex 2011. *A Model of Learning Objectives*. Retrieved August 19, 2013 from www.celt.iastate.edu/pdfs-docs/teaching/RevisedBloomsHandout.pdf.
- Hill Kathryn & McNamara Tim 2011. Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing* 29 (3): 395-420.
- Hirsjärvi Sirkka & Hurme Helena 2010. *Tutkimushaastattelu: teemahaastattelun teoria ja käytäntö* (Research interview: theory and practice of theme interview). Helsinki: Gaudeamus.
- Hofmanová Marie, Novotná Jarmila & Pipalová Renata 2008. Assessment Approaches to Teaching Mathematics in English as a Foreign Language (Czech Experience). *International CLIL Research Journal* 1 (1): 20-35.
- Hönig Ingrid 2010. *Assessment in CLIL: Theoretical and Empirical Research*. Saarbrücken: VDM Verlag.
- Housen Alex & Pierrard Michel 2005. *Studies on Language Acquisition: Investigations in Instructed Second Language Acquisition*. Berlin: Walter de Gruyter.
- Hulstijn, Jan H. 2003. Incidental and Intentional Learning. In C. J. Doughty & M. H. Long (eds). *The Handbook of Second Language Acquisition* (pp. 349-381). Malden, MA: Blackwell.
- Hunt Marilyn 2009. Progression and assessment in foreign languages at Key Stage 2. *Language Learning Journal* 37 (2): 205-217.
- Hüttner Julia, Dalton-Puffer Christiane & Smit Ute 2013. The power of beliefs: lay theories and their influence on the implementation of CLIL programmes. *International Journal of Bilingual Education and Bilingualism* 16 (3): 267-284.
- Hymes Dell H. 1967. Models of the interaction of language and social setting. *Journal of Social Issues* 23 (2): 8-28.
- Hymes Dell H. 1972. On communicative competence. In J.B. Pride & J. Holmes (eds). *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin.
- Ioanniu-Georgiu Sophie & Pavlou Pavlos 2003. *Assessing Young Learners*. Oxford: Oxford University Press.

- Jäppinen Aini-Kristiina 2004. Opettajien ammattitaito vieraskielisessä opetuksessa (Teacher professionalism in foreign language instruction). In K. Sajavaara & S. Takala (eds). *Kielikoulutus tiehaarassa* (pp. 193-206). Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus.
- Järvinen Heini-Marja 1999. *Acquisition of English in Content and Language Integrated Learning at Elementary Level in the Finnish Comprehensive School*. Turku: University of Turku.
- Järvinen, Heini-Marja 2004. Kielitaidon kehittyminen peruskoulun vieraskielisessä opetuksessa (Development of language proficiency in instruction in a foreign language in basic education). In S. Takala & K. Sajavaara (eds). *Kielikoulutus tienhaarassa*. (pp. 237–244). Jyväskylän yliopisto: Soveltavan kielentutkimuksen keskus
- Järvinen Heini-Marja 2007. Language in Content and Language Integrated Learning (CLIL). In Marsh D. & Wolff D. (eds). *Diverse Contexts – Converging Goals. CLIL in Europe* (pp. 255-260). Frankfurt am Main: Peter Lang.
- Järvinen Heini-Marja 2012a. Vieraiden kielten didaktiikkaa (Didactics of foreign languages). In A. Kallioniemi & A. Virta (eds): *Ainedidaktiikka tutkimuskohteena ja tiedonalana* (pp. 217-249). Jyväskylä: Jyväskylän yliopistopaino.
- Järvinen Heini-Marja 2012b. What does a CLIL teacher need to know about language – the linguistic competence of a CLIL teacher. In A. Koskensalo, J. Smeds, R. de Cillia & Á. Huguet (eds). *Language: Competence – Change – Contact* (pp. 47-54). Berlin: LIT.
- Jensen Murray, Johnson David W. & Johnson Roger T. 2002. Impact of positive interdependence during electronic quizzes on discourse and achievement. *The Journal of Educational Research* 95 (3): 161-166.
- Jewitt Carey 2003. Re-thinking assessment: multimodality, literacy and computer-mediated learning. *Assessment in Education* 10 (1): 83-102.
- Johnson Robert Keith & Swain Merrill 1997. *Immersion Education: International Perspectives*. Cambridge: CUP.
- Johnstone Richard 2000. Context-sensitive assessment of modern languages in primary (elementary) and early secondary education: Scotland and the European experience. *Language Testing* 17 (2): 123-143.
- Jones Glyn 1986. Computer simulations in language teaching – the Kingdom experiment. *System* 14 (2): 179-186.
- Julien Heidi 2008. Content analysis. *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 121-123). Thousand Oaks, CA: SAGE.
- Kangasvieri Teija, Miettinen Elisa, Palviainen Hannele, Saarinen Taina & Ala-Vähälä Timo 2012. Selvitys kotimaisten kielten kielikylpyopetuksen ja vieraskielisen opetuksen tilanteesta Suomessa. Kuntatason tarkastelu (Survey of immersion of national languages and instruction in a foreign language in Finland at municipal level). Retrieved March 16, 2013 from jyu.fi/hum/laitokset/solki/tutkimus/julkaisut/pdf-julkaisut/selvitys-kotimaisten-kielten-kielikylpyopetuksen-ja-vieraskielisen-opetuksen-tilanteesta-suomessa.
- Kapitanoff Susan H. 2009. Collaborative testing. Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education* 10 (1): 56-70.
- Kärkkäinen Riitta 2011. *Doing better? Children's and their parents' and teachers' perceptions of the malleability of the child's academic competencies*. Joensuu: University of Eastern Finland.
- Katz Lilian G. 1997. A developmental approach to assessment of young children. *ERIC Digest* April Issue. Retrieved July 20, 2013 from ecap.crc.illinois.edu/eecearchive/digests/1997/katz97.pdf.
- Kern Richard 2006. Perspectives on technology in learning and teaching languages. *TESOL Quarterly* 40 (1): 183-210.
- Kolb David A. 1984. *Experiential Learning. Experience as the Source of Learning and Development*. New Jersey: Prentice-Hall.
- Krashen Stephen & Brown Clara Lee 2007. *What is Academic Language Proficiency?*

- STETS Language and Communication Review* 6 (1), 1–4.
- Krashen Stephen D. 1985. *The Input Hypothesis: Issues and Implications*. New York: Longman.
- Krashen Stephen D. & Terrell Tracy D. 1983. *The Natural Approach. Language Acquisition in the Classroom*. Oxford: Pergamon.
- Lado Robert 1961. *Language Testing. The Construction and Use of Foreign Language Tests. A Teacher's Book*. London: Longmans.
- Lado Robert 1964. *Language Teaching. A Scientific Approach*. New York: McGraw-Hill.
- Langé Gisela 2007. Postscript to CLIL 2006 and future action. In D. Marsh & D. Wolff (eds). *Diverse Contexts – Converging Goals. CLIL in Europe* (pp. 351-354). Frankfurt am Main: Peter Lang.
- Lantolf James P. 2000. Introducing Sociocultural Theory. In J.P. Lantolf (ed.). *Sociocultural Theory and Second Language Learning* (pp. 1-26). Oxford: Oxford University Press.
- Lantolf James P. & Poehner Matthew E. 2004. Dynamic assessment of L2 development: bringing the past into the future. *Journal of Applied Linguistics* 1 (1): 49-72).
- Lantolf James P. & Poehner Matthew E. 2008. Dynamic Assessment. In E. Shohamy & N. H. Hornberger (eds). *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment* 2nd ed. (pp. 273–284). Boston, MA: Springer.
- Larsen-Freeman Diane & Freeman Donald 2008. Language moves: the place of “foreign” languages in classroom teaching and learning. *Review of Research in Education* 32: 147-186.
- Lasagabaster David 2008. Foreign language competence in content and language integrated courses. *The Open Applied Linguistic Journal* 1: 31-42.
- Lasagabaster David & Sierra Juan Manuel 2010. Immersion and CLIL in English: more differences than similarities. *ELT Journal* 64 (4): 367-375.
- Laurillard Diana 1992. Learning through collaborative computer simulations. *British Journal of Educational Technology* 23 (3): 164-171.
- Lazaruk Wally 2007. Linguistic, academic, and cognitive benefits of French immersion. *The Canadian Modern Language Review* 63 (5): 605-628.
- Lee Doris 1993. Learning English as the second language: the case for computer simulation. *British Journal of Educational Technology* 24 (3): 221-222.
- Lehti Lotta, Järvinen Heini-Marja & Suomela-Salmi Eija 2006. Kartoitus vieraskielisen opetuksen tarjonnasta peruskouluissa ja lukioissa (Survey of provision of instruction in a foreign language in basic education and upper secondary schools). In P. Pietilä, P. Lintunen & H-M. Järvinen (eds). *Kielenoppija tänään – Language Learners of Today. AFinLan vuosikirja 2006* (pp. 293-313). Jyväskylä: Suomen Soveltavan Kielitieteen Yhdistyksen AFinLan julkaisuja 64.
- Lightbown Patsy M. & Spada Nina 1994. An innovative program for primary ESL in Quebec. *TESOL Quarterly* 28 (3): 563-579.
- Lightbown Patsy M. & Spada Nina 2008. Form-focused instruction: isolated or integrated? *TESOL Quarterly* 42 (2): 181-207.
- Linturi Hannu & Rubin Anita 2011. *Toinen koulu, toinen maailma. Oppimisen tulevaisuus 2030* (Another school, another world. The future of learning 2030). TUTU-julkaisuja 1/2011. Turun yliopisto: Tulevaisuuden tutkimiskeskus.
- Little David 2007. The Common European Framework of Reference for Languages: perspectives on the making of supranational language education policy. *The Modern Language Journal* 9 (4): 645-655.
- Llinares Ana, Morton Tom & Whittaker Rachel 2012. *The Roles of Language in CLIL*. Cambridge: Cambridge University Press.
- Llosa Lorena 2011. Standards-based classroom assessments of English proficiency: a review of issues, current developments and future directions for research. *Language Testing* 28 (3): 367-382.

- Lo Yen Yi & Murphy Victoria A. 2010. Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education* 24 (3): 215-238.
- Loewen Shawn 2005. Incidental focus on form and second language learning. *Studies in Second Language Acquisition* 27 (3), 361-386.
- Long Michael H. 1996. The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (eds). *Handbook of Second Language Acquisition* (pp. 413-468). San Diego, CA: Academic Press.
- Long, Michael & Robinson, Peter 1998. Focus on form: theory, research and practice. In C. Doughty & J. Williams (eds). *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.
- Lorenzo Francisco 2007. The interface of language and content in CLIL courses – implications for curriculum organization in a European context. In D. Marsh & D. Wolff (eds). *Diverse Contexts – Converging Goals. CLIL in Europe* (p. 261-271). Frankfurt am Main: Peter Lang.
- Lorenzo Francisco 2013. Genre-based curricula: multilingual academic literacy in content and language integrated learning. *International Journal of Bilingual Education and Bilingualism* 16 (3): 375-388.
- Lorenzo Francisco, Casal Sonia & Moore Pat 2010. The effects of content and language integrated learning in European education: key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics* 31 (3): 418-442.
- Lorenzo Francisco & Moore Pat 2010. On the natural emergence of language structures in CLIL: towards a theory of European educational bilingualism. In C. Dalton-Puffer, T. Nikula & U. Smit (eds). *Language Use and Language Learning in CLIL classrooms* (pp. 23-38). Amsterdam: John Benjamins Publishing Company.
- Luukka Minna-Riitta, Pöyhönen Sari, Huhta Ari, Taalas Peppi, Tarnanen Mirja & Keränen Anna 2008. Maailma muuttuu – Mitä tekee koulu? Äidinkielen ja vieraiden kielten tekstikäytänteet koulussa ja vapaa-ajalla (World changes - what does the school? The textual practices of mother tongue and foreign languages in school and leisure time). Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus.
- Luxemburg Presidency Position on Plurilingual Education 2005. Council of the European Union. Retrieved July 19, 2013 from register.consilium.europa.eu/pdf/en/05/st08/st08269.en05.pdf.
- Lynch Brian K. 2001. Rethinking assessment from a critical perspective. *Language Testing* 18 (4): 351-372.
- Lynch Brian & Shaw Peter 2005. Portfolios, power, and ethics. *TESOL Quarterly* 39 (2): 21.
- Lyster Roy 2004. Research on form-focused instruction in immersion classrooms: implications for theory and practice. *Journal of French Language Studies* 14 (3): 321-341.
- Lyster Roy 2011. Content-based second language teaching. In E. Hinkel (ed.). *Handbook of Research in Second Language Teaching and Learning Volume II* (pp. 611-630). New York: Routledge.
- Mäensivu Kirsti 1999. *Opettaja määrittelijänä - oppilas määriteltävänä: sanallisen oppilaan arvioinnin sisällön analyysi* (Teacher as the definer - pupil being defined: content analysis of verbal pupil assessment). Jyväskylä: Jyväskylän yliopisto.
- Maljers Anne, Marsh David & Wolff Dieter 2007. *Windows on CLIL: Content and Language Integrated Learning in the European Spotlight*. Alkmaar: Europees Platform.
- Marsh David 2009. Foreword. In Y. Ruiz de Zarobe & R.M. Jiménez Catalán (eds). *Content and Language Integrated Learning. Evidence from Research in Europe*. Bristol: Multilingual Matters.
- Marsh David 2013. *The CLIL Trajectory: Educational Innovation for the 21st Century iGeneration*. Córdoba: Servicio de Publicaciones, Universidad de Córdoba.
- Marsh David, Järvinen Heini-Marja & Haataja Kim 2007. Finland. In A. Maljers, D. Marsh & D. Wolff (eds). *Windows on CLIL:*

- Content and Language integrated Learning in the European Spotlight* (pp. 63-80). Alkmaar: Europees Platform.
- Marsh David & Wolff Dieter (eds) 2007. *Diverse Contexts - Converging Goals*. Frankfurt am Main: Peter Lang.
- Martin-Beltrán Melinda 2010. Positioning proficiency: How students and teachers (de)construct language proficiency at school. *Linguistics and Education* 21: 257-281.
- Massler Ute & Stotz Daniel (in preparation). *Handbuch zur Leistungserhebung und Leistungsbewertung im fremdsprachlichen Sachfachunterricht (CLIL) der Primarstufe* (Handbook of enhancement and assessment of achievement in content instruction in a foreign language CLIL at primary level). Trier: Wissenschaftlicher Verlag Trier.
- May Stephen 2008. Bilingual/immersion education: what the research tells us. In J. Cummins & N. H. Hornberger (eds). *Encyclopedia of Language and Education*, 2nd edition. Volume 5: *Bilingual Education* (pp. 19-34). Boston, MA: Springer.
- McKay Penny 2006. *Assessing Young Language Learners*. Cambridge: Cambridge University Press.
- Macmillan 2007. *Macmillan English Dictionary for Advanced Learners*. CD-ROM 2nd Edition Version 2.3. Macmillan Publishers Limited.
- McMillan James H. 2003. Understanding and improving teachers' classroom assessment decision making: implications for theory and practise. *Educational Measurement Issues and Practice* 22 (4): 34-43.
- McNamara Tim 1996. *Measuring Second Language Performance*. London: Longman.
- McNamara Tim 2000. *Language Testing*. Oxford: Oxford University Press.
- Mehisto Peeter & Lucietto Sandra 2010. *CLIL Essentials*. Retrieved July 21, 2011 from ccn-clil.eu/index.php?name=File&nodeIDX=5060
- Mehisto Peeter, Marsh David & Frigols Jesús María 2008. *Uncovering CLIL: Content and Language Integrated Learning in Bilingual and Multilingual Education*. Oxford: Macmillan Education.
- Merisuo-Storm Tuula 2013. Kaksikielinen opetus ja äidinkielen kehittyminen (Bilingual instruction and mother tongue development). In L. Tainio & H. Harju-Luukkainen (eds). *Kaksikielinen koulu – tulevaisuuden monikielinen Suomi* (pp. 167-191). Jyväskylä: Suomen kasvatustieteellinen seura.
- Messick Samuel 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23 (2): 13-23.
- Met Myriam 1999. *Content-based instruction: defining terms, making decisions*. NFLC Reports. Washington, DC: The National Foreign Language Center. Retrieved 28 August, 2013 from carla.umn.edu/cobaltt/modules/principles/decisions.html.
- Miettinen Elisa, Kangasvieri Teija & Saarinen Taina 2013. Vaihtelevaa toteutusta ja kasvavaa kiinnostusta: kielikylypöpetus ja vieraskielinen opetus kunnissa (Varying implementation and growing interest: immersion and instruction in a foreign language in municipalities). In L. Tainio & H. Harju-Luukkainen (eds). *Kaksikielinen koulu – tulevaisuuden monikielinen Suomi* (pp. 71-90). Jyväskylä: Jyväskylän yliopistopaino.
- Miller Megan & Hegelheimer Volker 2006. The SIMs meet ESL. Incorporating authentic computer simulation games into the language classroom. *Interactive Technology & Smart Education* 4: 311-328.
- Ministry of Education 2005. Opettajalta edellytettävä kielen hallinta muulla kuin koulun opetuskielellä annettavassa esi- ja perusopetuksessa (Requirements for teachers' language mastery in pre- and basic education given through other than school's language of instruction). Retrieved April 19, 2013 from finlex.fi/data/normit/26301-opf250112005su.pdf.
- Miramontes Ofelia B., Nadeau Adel & Commins Nancy L. 2011. *Restructuring Schools for Linguistic Diversity. Linking Decision Making to Effective Programs*, 2nd ed. New York: Teachers College Press.
- Mohan Bernard 1986. *Language and Content*. Reading, MA: Addison-Wesley.

- Mohan Bernard, Leung Constant & Slater Tammy 2010. Assessing language and content: a functional perspective. In A. Paran & L. Secru (eds). *Testing the Untestable in Language Education* (pp. 217-240). Bristol: Multilingual Matters.
- Mohan Bernard and Slater Tammy 2005. A functional perspective on the critical 'theory/practice' relation in teaching language and science. *Linguistics and Education* 16: 151-172.
- Morgan Carol 2006. Appropriate language assessment in content and language integrated learning. *Language Learning Journal* 33: 59-67.
- Mustaparta Anna-Kaisa & Tella Anneli 1999. *Vieraskielisen opetuksen järjestäminen* (Organising instruction in a foreign language). Kehittyvä koulutus 1/1999. Helsinki: Opetushallitus.
- National Board of Education 2013. *Education*. Retrieved April 19, 2013 from oph.fi/english/education.
- NCC 2004. *The National Core Curriculum for Basic Education*. National Board of Education. Retrieved July 20, 2013 from oph.fi/english/publications/2009/national_core_curricula_for_basic_education.
- NCC draft 2014. *Draft of the foundations for the new National Core Curriculum for Basic Education 2016* (in Finnish). Retrieved April 15, 2014 from www.oph.fi/download/156870_perusopetus_perusteluonnos_luvut_1_12.pdf.
- Newton Paul E. 2007. Clarifying the purposes of educational assessment. *Assessment in Education* 14 (2): 149-170.
- Nikula Tarja 2005. English as an object and tool of study in classrooms: interactional effects and pragmatic implications. *Linguistics and Education* 16: 27-58.
- Nikula Tarja 2010. Effects of CLIL on a teacher's classroom language use. In C. Dalton-Puffer, T. Nikula & U. Smit (eds). *Language Use and Language Learning in CLIL Classrooms* (pp. 105-124). Amsterdam: John Benjamins.
- Nikula Tarja 2012. On the role of peer discussions in the learning of subject-specific language use in CLIL. In E. Alcón & M. Pilar Safont (eds). *Language learning through classroom discourse* (pp. 133-153). Amsterdam: Rodopi (Utrecht Studies in Language and Communication).
- Nikula Tarja & Järvinen Heini-Marja 2013. Vieraskielinen opetus Suomessa (Instruction in a foreign language in Finland). In L. Tainio & H. Harju-Luukkainen (eds). *Kaksikielinen koulu – tulevaisuuden monikielinen Suomi* (pp. 143-166). Jyväskylä: Jyväskylän yliopistopaino.
- Nikula Tarja & Marsh David 1996. *Kartoitus vieraskielisen opetuksen tarjonnasta peruskouluissa ja lukioissa* (Survey of provision of instruction in a foreign language in basic education and upper secondary schools). Helsinki: Opetushallitus.
- Nikula Tarja & Smit Ute (eds) 2010. *Language Use and Language Learning in CLIL Classrooms*. Amsterdam: John Benjamins.
- Norrena Juho 2013. *Opettaja tulevaisuuden taitojen edistäjänä: "Jos haluat opettaa noita taitoja, sinun on ensin hallittava ne itse"* (Teacher enhancing 21st century skills: "if you want to teach those skills you must master them yourself"). Jyväskylä studies in computing 169. Jyväskylä: Jyväskylän yliopisto.
- Nunan David 1986. The ghost in the machine: an examination of the concept language proficiency. *Prospect* 2 (2): 1-14.
- Onwuegbuzie Anthony J. & Johnson Burke R. 2006. The validity issue in mixed research. *Research in the Schools* 13 (1): 48-63.
- Ortega Lourdes 2009. *Understanding Second Language Acquisition*. London: Hodder Education.
- Pandey Carol & Kapitanoff Susan 2011. The influence of anxiety and quality of interaction on collaborative test performance. *Active Learning in Higher Education* 12 (3): 163-174.
- Pérez-Cañado María Luisa 2012. CLIL research in Europe: past, present, and future. *International Journal of Bilingual Education and Bilingualism* 15 (3): 315-341.
- Pérez-Vidal Carmen 2007. The need of focus on form (FoF) in content and language

- integrated approaches: an exploratory study. *Volumen Monográfico*: 39-54.
- Pica Teresa 2002. Subject-matter content: how does it assist the interactional and linguistic needs of classroom language learners? *The Modern Language Journal* 86 (1): 1-19.
- Pierce Lorraine Valdez 2002. Performance-based assessment: promoting achievement for English language learners. *ERIC/CLL Newsbulletin*. Retrieved July 20, 2013 from www.cal.org/resources/archive/news/2002fal/CLLNewsBulletin_Fa02c.pdf.
- Pihko Marja-Kaisa 2010. *Vieras kieli kouluopiskelun välineenä: Oppilaiden kokemuksista vihjeitä CLIL-opetuksen kehittämiseen* (Foreign language as a medium of school study: implications of pupils' experiences for development of CLIL instruction). Jyväskylä: Jyväskylän yliopisto.
- Poisel Eva 2007. Assessment modes in CLIL to enhance language proficiency and interpersonal skills. *VIEWS* 16 (3): 43-46.
- Pöyhönen Sari, Tarnanen Mirja, Kyllönen Teija, Vehviläinen Eeva-Maija & Rynkänen Tatjana 2009. *Kielikoulutus maahanmuuttajien kotoutumiskoulutuksessa. Tavoitteet, toteutus ja hallinnollinen yhteistyö* (Language education in the social integration of immigrants. Aims, implementation and executive collaboration). Jyväskylän yliopisto: Soveltavan kielentutkimuksen keskus.
- PROFICOM 2013. *Profiling learner progression in CLIL environments through computer simulations*. Retrieved July 16, 2013 from rule.uta.fi/proficom/proficom-in-brief/.
- Pulfrey Caroline, Buchs Céline & Butera Fabrizio 2011. Why grades engender performance-avoidance goals: the mediating role of autonomous motivation. *Journal of Educational Psychology* 103 (3): 683-700.
- Pulfrey Caroline, Darnon Céline & Butera Fabrizio 2013. Autonomy and task performance: explaining the impact of grades on intrinsic motivation. *Journal of Educational Psychology* 105 (1): 39-57.
- Quartapelle Franca (ed.) 2012. *Assessment and Evaluation in CLIL*. Retrieved July 16, 2013 from aecilil.net.
- Rahman Helinä 2012. *Finnish Pupils' Communicative Language Use of English in Interviews in Basic Education Grades 1-6*. Research Report 340. Helsinki: University of Helsinki.
- Rea-Dickins Pauline 2000. Assessment in early years language learning contexts. Editorial. *Language Testing* 17 (2): 115-122.
- Rea-Dickins Pauline 2001. Mirror, mirror on the wall: identifying processes of classroom assessment. *Language Testing* 18 (4): 429-462.
- Rea-Dickins Pauline 2007. Classroom-based assessment: possibilities and pitfalls. In J. Cummings & C. Davidson (eds). *International Handbook of English Language Teaching* (pp. 505-520). New York: Springer.
- Rea-Dickins Pauline 2008. Classroom-based assessment. In E. Shohamy & N.H. Hornberger (eds). *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment*, 2nd ed. (pp. 257-271). Boston, MA: Springer.
- Rea-Dickins Pauline & Gardner Sheena 2000. Snares and silver bullets: disentangling the construct of formative assessment. *Language Testing* 17 (2): 215-243.
- Ricci Garotti Federica 2007. Five methodological research questions for CLIL. In D. Marsh & D. Wolff (eds). *Diverse Contexts – Converging Goals: CLIL in Europe* (pp. 131-144). Frankfurt am Main: Peter Lang.
- Robinson Peter & Gilabert Roger 2007. Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching* 45: 161-176.
- Rodgers Daryl M. 2006. Developing content and form: encouraging evidence from Italian content-based instruction. *The Modern Language Journal* 90 (3): 373-386.
- Rothbauer Paulette M. 2008. Triangulation. In L.M. Given (ed.). *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 893-895). Thousand Oaks: SAGE.

- Ruiz de Zarobe Yolanda 2008. CLIL and foreign language learning: a longitudinal study in the Basque Country. *International CLIL Research Journal* 1 (1): 60-73.
- Ruiz de Zarobe Yolanda & Jiménez Catalán Rosa María (eds.) 2009. *Content and Language Integrated Learning. Evidence from Research in Europe*. Bristol: Multilingual Matters.
- Ruiz de Zarobe Yolanda 2010. Written production and CLIL. In C. Dalton-Puffer, T. Nikula & U. Smit (eds). *Language Use and Language Learning in CLIL classrooms* (pp. 191-209). Amsterdam: John Benjamins.
- Rumlich Dominik 2013. *In search of CLIL effects: Results from a large-scale longitudinal study on CLIL and non-CLIL students' academic self-concept*. Presentation at the CLIL Conference on Modernizing Educational Practice, Perspectives in Content and Language Integrated Learning, Ustrón, Poland, April 4-5, 2013.
- Russell Carol & Shepherd John 2010. Online role-play environments for higher education. *British Journal of Educational Technology* 41 (6): 992-1002.
- Sanders William L. & Rivers June C. 1996. *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*. Research Progress Report. University of Tennessee Value-Added Research and Assessment Center. Retrieved June 20, 2013 from mccluelearning.com/wp-content/uploads/2011/09/Cumulative-and-Residual-Effects-of-Teachers.pdf.
- Satterly David 1983. *Assessment in Schools*. Oxford: Basil Blackwell.
- Scarcella Robin 2003. *Academic English: A Conceptual Framework*. Technical Reports. University of California Linguistic Minority Research Institute, UC Berkeley.
- Schleppegrell Mary J., Achugar Mariana & Oteiza Teresa 2004. The grammar of history: enhancing content-based instruction through a functional focus on language. *TESOL Quarterly* 38 (1): 67-93.
- Schleppegrell Mary J. 2006. The challenges of academic language in school subjects. In I. Lindberg & K. Sandwall (eds). *Språket och Kunskapen: att lära på sitt andraspråk i skola och högskola* (pp. 47-69). Göteborg: Göteborgs universitet institutet för svenska som andraspråk.
- Schmidt Richard 1993. Awareness and second language acquisition. *Annual Review of Applied Linguistics* 13: 206-226.
- Schmidt Richard 2010. *Attention, awareness, and individual differences in language learning*. Retrieved August 4, 2013 from nflrc.hawaii.edu/PDFs/SCHMIDT%20Attention,%20awareness,%20and%20individual%20differences.pdf.
- Seikkula-Leino Jaana 2007. CLIL learning: achievement levels and affective factors. *Language and Education* 21 (4): 328-341.
- Seliger Herbert W. & Shohamy Elana 1989. *Second Language Research Methods*. Oxford: Oxford University Press.
- Serra Cecilia 2007. Assessing CLIL at primary school: a longitudinal study. *International Journal of Bilingual Education and Bilingualism* 10 (5): 582-602.
- Serragiotto Graziano 2007. Assessment and evaluation in CLIL. In D. Marsh & D. Wolff (eds). *Diverse Contexts – Converging Goals. CLIL in Europe* (pp. 271-283). Frankfurt am Main: Peter Lang.
- Shabaan Kassim 2001. Assessment of young learners. *English Teaching Forum* 39 (4): 16 -23.
- Shepard Lorrie A. 2000. The role of assessment in a learning culture. *Educational Researcher* 29 (7): 4-14.
- Shohamy Elana 2001. *The Power of Tests. A Critical Perspective on the Uses of Language Tests*. Essex: Pearson.
- Short Deborah J. 1993. Assessing integrated language and content instruction. *TESOL Quarterly* 27 (4): 627-656.
- Siegel Jeff 2003. Social context. In C. J. Doughty & M. H. Long (eds). *The Handbook of Second Language Acquisition* (pp. 178-223). Malden, MA: Blackwell.
- Sjöberg Anne 2004. Toimiva kielitaito työelämässä (Functional language proficiency in working life). In K. Sajavaara & S. Takala (eds). *Kielikoulutus tienhaarassa*

- (pp. 137-147). Jyväskylä: Jyväskylän yliopisto.
- Skehan Peter 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Skehan Peter 2001. Tasks and language performance assessment. In M. Bygate, P. Skehan & M. Swain (eds). *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 167-185). Harlow: Pearson.
- Skehan Peter 2003. Focus on form, tasks, and technology. *Computer Assisted Language Learning* 16 (5): 391-411.
- Skehan Peter 2009. Modelling second language performance: integrating complexity, accuracy, fluency and lexis. *Applied Linguistics* 30 (4): 510-532.
- Snow Catherine E. & Uccelli Paola 2009. The challenge of academic language. In D. R. Olson & N. Torrance (eds): *The Cambridge Handbook of Literacy* (pp. 112-133). New York: Cambridge University Press.
- Snow Marguerite Ann, Met Myriam & Genesee Fred 1989. A conceptual framework for integration of language and content in second/foreign language instruction. *TESOL Quarterly* 23 (2): 201-217.
- Somekh Bridget 2008. Action research. In L.M. Given (ed.). *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 5-8). Thousand Oaks, CA: SAGE.
- Somers Thomas & Surmont Jill 2012. CLIL and immersion: how clear-cut are they? *ELT Journal* 66 (1): 113-116.
- Spada Nina & Lightbown Patsy M. 2008. Form-focused instruction: isolated or integrated? *TESOL Quarterly* 42 (2): 181-207.
- Spolsky Bernard 2008. Language assessment in historical and future perspective. In E. Shohamy & N.H. Hornberger (eds). *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment*, 2nd ed. (pp. 445-454). Boston, MA: Springer.
- Stockwell Glenn 2012. *Computer-Assisted Language Learning. Diversity in Research and Practice*. Cambridge: Cambridge University Press.
- Stoller Fredrika 2008. Content-based instruction. In N. Van Deusen-Scholl & N. H. Hornberger (eds): *Encyclopedia of Language and Education, Volume 4: Second and Foreign Language Education*, 2nd edition (pp. 59-70). Boston, MA: Springer.
- Stoyhoff Stephen 2012. Looking backward and forward at classroom-based language assessment. *ELT Journal* 66 (4): 523-532.
- Stryker Stephen B. & Leaver Betty Lou 1997a. Content-based instruction: some lessons and implications. In S.B. Stryker & B.L. Leaver (eds). *Content-Based Instruction in Foreign Language Education: Models and Methods* (pp. 285-312). Washington, D.C: Georgetown University Press.
- Stryker Stephen B. & Leaver Betty Lou 1997b. Content-based instruction: from theory to practice. In S.B. Stryker & B.L. Leaver (eds). *Content-Based Instruction in Foreign Language Education: Models and Methods* (pp. 3-28). Washington, D.C: Georgetown University Press.
- Sundman Marketta 2013. Kaksikielisiä kouluja kaksikieliseen Suomeen? (Bilingual schools in bilingual Finland?). In L. Tainio & H. Harju-Luukkainen (eds). *Kaksikielinen koulu – tulevaisuuden monikielinen Suomi* (pp. 47-67). Jyväskylä: Jyväskylän yliopistopaino.
- Suomela-Salmi Eija 2010. What is alternative assessment? In F. Dervin & E. Suomela-Salmi (eds). *New Approaches to Assessing Language and (Inter-)Cultural Competences in Higher Education* (pp. 209-234). Frankfurt am Main: Peter Lang.
- Sylvén Kerstin 2010. *Teaching in English or English Teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*. Göteborg: University of Gothenburg.
- Sylvén Kerstin 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.

- Swain Merrill 1988. Manipulating and complementing content teaching to maximise second language learning. *TESL Canada Journal* 6 (1): 68-83.
- Swain Merrill 1993. The Output Hypothesis: just speaking and writing aren't enough. *The Canadian Modern Language Review* 50 (1): 158-164.
- Swain Merrill & Johnson Robert Keith 1997. Immersion education. In R. K. Johnson & M. Swain (eds). *Immersion Education: International Perspectives* (pp. 1-16). Cambridge: Cambridge University Press.
- Swain Merrill & Lapkin Sharon 1995. Problems in output and the cognitive processes they generate: a step towards second language learning. *Applied Linguistics* 16: 371-391.
- Swain Merrill & Lapkin Sharon 2005. The evolving sociopolitical context of immersion education in Canada: some implications for program development. *International Journal of Applied Linguistics* 15 (2): 169-186.
- Takala Sauli 1992. *Virikkeitä uutta kokeilevaan koulutyöhön* (Stimuli for innovative school work). Jyväskylä: Jyväskylän yliopisto, Kasvatustieteiden tutkimuslaitos.
- Takala Sauli, Marsh David & Nikula Tarja 1998. Vieraskielinen opetus Suomessa (Instruction in a foreign language in Finland). In S. Takala & K. Sajavaara (eds). *Kielikoulutus Suomessa* (pp. 139-170). Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus.
- Teddlie Charles & Tashakkori Abbas 2003. Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (eds). *Handbook of Mixed Methods in Social & Behavioral Research* (pp. 3-50). Thousand Oaks, CA: Sage.
- Teddlie Charles & Tashakkori Abbas 2009. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage.
- Tella Seppo 1999. Teaching through a foreign language revisited: from tool to empowering mediator. In S. Tella, A. Räsänen & A. Vähäpassi (eds). *Teaching Through a Foreign Language. From Tool to Empowering Mediator* (pp. 26-31). Publications of Higher Education Evaluation Council 5:1999. Helsinki: Edita.
- Tillema Harm, Leenknecht Martijn & Segers Mien 2011. Assessing assessment quality: criteria for quality assurance in design of (peer) assessment of learning - A review of research studies. *Studies in Educational Evaluation* 37: 25-34.
- Turnbull Miles, Hart Doug & Lapkin Sharon 2003. Grade 6 French immersion students' performance on large-scale reading, writing, and mathematics tests: building explanations. *The Alberta Journal of Educational Research* XLIX (1): 6-23.
- Väljjarvi Jouni 2012. Suomalainen arviointitodellisuus kansainvälisessä vertailussa (Finnish assessment reality in international comparison). In J. Kivirauma, A. Jauhiainen, P. Seppänen & T. Kaunisto (eds). *Koulutuksen yhteiskunnallinen ymmärrys. Social Perspectives on Education* (pp. 171-200). Jyväskylä: Jyväskylän yliopistopaino.
- Várkuti Anna 2010. Linguistic benefits of the CLIL approach: measuring linguistic competences. *International CLIL Research Journal* 1 (3): 67-79.
- Vollmer Helmut J. 1983. The structure of foreign language competence. In A. Hughes & D. Porter (eds). *Current Developments in Language Testing* (pp. 3-30). London: Academic Press.
- Vukovic Rose K. & Lesaux Nonie K. 2013. The language of mathematics: investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology* 115: 227-244.
- Vygotski Lev S. 1978. *Mind in Society*. Cambridge, MA: Harvard University Press.
- Warschauer Mark 2003. *Technology and Social Inclusion: Rethinking the Digital Divide*. Cambridge, MA: MIT Press.
- Webster Jonathan J. 2009 (ed.). *The Essential Halliday*. London: Continuum.
- Weigle Sara Cushing & Jensen Linda 1997. Issues in assessment for content-based instruction. In M. A. Snow & D. M. Brinton

- (eds). *The Content-Based Classroom. Perspectives on Integrating Language and Content* (pp. 201-212). New York: Longman.
- Wewer, Taina 2013a. Minuutissa Michiganiin: tietokonesimulaatiot englannin kielen arviointimenetelmänä alakoulun vieraskielisessä CLIL-opetuksessa (One minute to Michigan: computer simulations as an assessment method of English language proficiency in primary CLIL instruction). In: H. Juuso & al. (eds). *Tutkimusperustaisuus koulussa ja opettajankoulutuksessa* (pp. 100-109). Suomen harjoittelukoulujen julkaisu. Oulu: Oulun yliopisto, Oulun normaalikoulu.
- Wewer, Taina 2013b. English language assessment in bilingual CLIL instruction at primary level in Finland: Quest for updated and valid assessment methods. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 18 (2): 76-87. Retrieved November 22, 2013 from <http://zif.spz.tu-darmstadt.de/jg-18-2/beitrag/Wewer.pdf>.
- Wewer, Taina 2013c. CLIL-opetuksen tavoitteista (On objectives of CLIL instruction). In: T. Wewer (ed.). *CLIL liitton! Johdatus vieraskieliseen opetukseen Turun normaalikoulussa* (pp. 1-2). Turun normaalikoulun julkaisu 1/2013.
- Wewer Taina (forthcoming). *Portfolio as an indicator of Young Learners' English proficiency in Mainstream English and Bilingual Content Instruction*. Turku: University of Turku.
- White Paper on Education and Training* 1995. Teaching and learning. Towards the Learning Society. European Commission. Retrieved April 16, 2014 from http://europa.eu/documents/comm/white_papers/pdf/com95_590_en.pdf.
- Wigglesworth Gillian & Storch Neomy 2009. Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing* 26 (3): 445-466.
- Willis Dave & Willis Jane 2007. *Doing Task-based Teaching*. Oxford: Oxford University Press.
- Wolff, Dieter 2007. CLIL: bridging the gap between school and working life. In D. Marsh & D. Wolff (eds). *Diverse Contexts – Converging Goals* (pp. 15-25). Frankfurt am Main: Peter Lang.
- Xanthou Maria 2011. The impact of CLIL on L2 vocabulary development and content knowledge. *English Teaching: Practice and Critique* 10 (4): 116-126.
- Yin Robert K. 2009. *Case Study Research: Design and Methods*, 4th ed. Los Angeles, CA: SAGE.
- Zuengler Jane & Brinton Donna 1997. Linguistic form, pragmatic function: Relevant research from content-based instruction. In M. Snow & D. Brinton (eds). *The content-based classroom: Perspectives on integrating language and content* (pp. 263-273). White Plains, NY: Longman.
- Zydatiś Wolfgang 2012. Linguistic thresholds in the CLIL classroom? The Threshold Hypothesis revisited. *International CLIL Research Journal* 1 (4): 17-28.

Appendices

Appendix 1: Introduction letter for CLIL assessment survey and parental consent form

Appendix 2: Introduction letter for simulation 1 pilot and questionnaire, parental consent form

Appendix 3: CLIL assessment survey questionnaire for teachers

Appendix 4: CLIL assessment survey questionnaire for pupils

Appendix 5: CLIL assessment survey questionnaire for parents

Appendix 6: Theme interview question plan (teachers, pupils and parents)

Appendix 7: Simulation 1 questionnaire for pupils

Appendix 8: Simulation 2 questionnaire for pupils

Appendix 9: Simulation 2 questionnaire for pupils, different questions for 5th graders

Appendix 10: Simulation questionnaire 2 for parents

Appendix 11: Simulation 1 tasks

Appendix 12: Simulation 2 tasks

Hyvä huoltaja,

Raisiossa 10.1.2012

vieraskielisessä CLIL-opetuksessa olevan oppilaan suuren mielenkiinnon kohteena on usein oman kielitaidon kehittyminen ja sen taso. Kielitaidossa edistyminen kiinnostaa varmasti myös useimpia huoltajia. Arviointitieto kielitaidon karttumisesta on myös opettajalle tärkeä asia mm. opetuksen suunnittelua varten.

Tutkimus

Oppilaan kielitaidon profilointi ja arviointi vieraskielisessä CLIL-opetuksessa on aihe, jota Suomessa ei ole aikaisemmin tutkittu, ja Euroopassakin hyvin vähän. Olen valmistellut tutkimusta, jolla on kaksi päätavoitetta:

- 1) kerätä tietoa nykyisistä arviointimenetelmistä ja niiden riittävydestä
- 2) oppilaslähtöisten arviointimenetelmien kehittäminen, kokeileminen ja käyttäjäkokemusten kerääminen

Tutkimukseni ”CLIL-oppilaan kielitaidon profilointi ja arviointi alakoulussa” liittyy Turun yliopiston kasvatustieteiden tiedekunnan tohtorikoulutusohjelmaan, ja ohjaajanani toimii vieraiden kielten didaktiikan professori Annikki Koskensalo Turun opettajankoulutuslaitoksesta.

Kokeilut

Tutkimukseen olen pyytänyt mukaan [redacted]n koulun (3., 4. ja 5.) ja vastaavat [redacted]koulun englantiluokat kokeilemaan erilaisia arviointimenetelmiä: kolmasluokkalaiset salkkutyöskentelyä (itsearviointipainotteinen) sekä 4.- ja 5.-luokkalaiset LangPerform-tietokonesimulaatiota, jossa oppilas reagoi erilaisiin filmipohjaisiin, reaali maailman tilanteita simuloiviin kielenkäyttötilanteisiin tietokoneen avulla. LangPerform-simulaatiota ei aiemmin ole kehitetty ja kokeiltu alakoulussa. Kokeiluosuus on tarkoitus toteuttaa vuoden 2012 aikana.

Kyselyt

Kokeilun lisäksi kerään oppilailta, opettajilta ja huoltajilta tietoa kyselylomakkein nykyisistä arviointimenetelmistä ja -toiveista ennen kokeiluosuutta. Lisäksi aihetta syvennetään vapaaehtoisia haastatellen. Toinen kyselyosuus tapahtuu simulaatiokokeilujen jälkeen, ja myös siihen liittyvät syventävät vapaaehtoisten haastattelut. Nämä kaikki tapahtuvat koulupäivän aikana.

Kaikki tutkimusta varten kerätty nimetön, luottamuksellinen tieto tulee vain tutkijan itsensä käsiteltäväksi, ja tutkimustulokset julkaistaan siten, että ketään yksittäistä henkilöä ei voi tunnistaa.

Tutkimuslupa

Oppilaiden kyselyjä (ja mahdollisia haastatteluja) varten tarvitaan huoltajan allekirjoittama kirjallinen lupa, joka palautetaan omalle opettajalle. Lupapaperi on seuraavalla sivulla. Vastaan mielelläni tarkemmin kaikkiin tutkimukseen liittyviin lisäkysymyksiin.

Ystävällisesti,

Taina Wewer, luokanopettaja, tohtoriopiskelija, taina.wewer@utu.fi, 040-5242921

TUTKIMUSLUPA

(Oppilaan nimi)

(luokka)

- saa osallistua (Sopiva vaihtoehto rastitetaan.)
- ei saa osallistua

erillisellä paperilla selvitettyyn väitöstutkimukseen ”CLIL-oppilaan kielitaidon profilointi ja arviointi alakoulussa” liittyviin oppilaan kielitaidon arviointia koskevaan kartoituskyselyyn sekä arviointikokeilujen jälkeiseen kokemuksekyselyyn.

Kyselyihin voi liittyä myös syventävä haastattelu, johon oppilas voi osallistua halutessaan.

Kyselyt toteutetaan aikavälillä kevätlukukausi 2012 – syyslukukausi 2012.

Tutkimuksessa kerättävä tieto on nimetöntä ja luottamuksellista, eikä tuloksia julkaistaessa tutkimukseen osallistuneita henkilöitä pysty yksilöimään.

(Huoltajan allekirjoitus)

(Nimenselvennys)

Tämä lupapaperi palautetaan omalle opettajalle mahdollisimman pian.

TUTKIMUSLUPA SIMULAATIOARVIOINTIKYSELYYN 4. ja 5. 20.1.2013

Hyvä huoltaja,

███n koulusta monet oppilaat ja huoltajat ovat osallistuneet väitöstutkimukseni "CLIL-oppilaan kielitaidon profilointi ja arviointi alakoulussa" ensimmäiseen osaan, jossa kartoitettiin nykyisin CLIL-luokissa käytössä olevia englannin kielen arviointimenetelmiä ja -käytänteitä sekä toiveita ja näkemyksiä niiden kehittämisestä.

Toisessa osassa alkusyksyllä 2012 kokeiltiin ensimmäistä Opetushallituksen rahoituksella PROFICOM-hankkeessa (lisätietoa: <http://rule.uta.fi/fi/proficom/proficom-lyhyesti/>) toteutetuista tarinapohjaisista tietokonesimulaatioista, jossa he kohtasivat erilaisia englannin kielen käyttötilanteita ja pääsivät reagoimaan niihin esimerkiksi puhumalla tai kirjoittamalla. Olen erittäin kiitollinen kaikista tähän asti saamistani kyselyvastauksista.

Nyt toinen simulaatio on valmistunut, ja kokeilyryhmäksi on valittu sekä ███n eli ███ että ███ koulun neljäs- ja viidesluokkalaiset CLIL-oppilaat. Viime kokeilussa ███ssa oli sen verran teknisiä ongelmia, että tällä kertaa toteutamme kokeilun ███koulussa vieraskielisen opetuksen yhteistoiminta- ja tutustumispäivän yhteydessä maanantaina 4.2.2013.

Käytännössä ███n kokeiluluokat lähtevät hankerahoituksella kustannetulla bussikuljetuksella päiväretkelle ███iin, jossa he ovat kahden CLIL-luokan vieraana. Päivän aikana oppilaat työskentelevät yhdessä isäntäluokkien oppilaiden kanssa (esim. kuvataideprojekti, tutustumista). Oppilaat myös ruokailevat ███issa. Koulupäivän kesto on n. 8.00-14.45. Luokanopettajat antavat tarkempia ohjeita esim. mukana tarvittavista asioista ja tarvikkeista.

Vierailupäivänä tutkimusluvan saaneet oppilaat osallistuvat simulaatiokokeiluun. Simulaatiosuoritus ei vaikuta englannin todistusarvosanaan. Tutkimuslupaa tarvitaan siihen, että oppilas saa simulaation jälkeen täyttää kyselyn, jossa tiedustellaan mm. kokemuksia, ajatuksia ja tunteita simulaatiosta sekä pyydetään arvioimaan omaa suoriutumista ja kielitaitoa. Kyselyt ovat nimettömiä, ja tuloksia käsitellään suurina kokonaisuuksina, jolloin ketään ei voi missään vaiheessa yksilöidä.

Lisäksi vapaaehtoisia luvan saaneita oppilaita voidaan lyhyesti haastatella omista kokemuksistaan joko videolle tai ääninauhurille. Haastattelujenkaan oppilaiden henkilöllisyyttä ei voida tutkimustuloksia julkaistaessa tunnistaa. Kaikki yksilöivät tiedot jäävät vain tutkijalle itsensä tietoon.

Jokainen oppilas saa oman tunnuksen ja salasanan, joiden avulla oppilas ja huoltaja voivat kotonakin kirjautua Tampereen yliopiston kielilaboratorioon ja katsella/kuunnella oppilaan suorituksen uudelleen. Koteihin lähetän huoltajille oman simulaatiokyselyn vastattavaksi heti samana päivänä, kun oppilas on Norssissa simulaatiota kokeillut. Eri näkökulmat (oppilas, huoltaja) antavat arvokasta tietoa simulaatioiden soveltuvuudesta kielitaidon tarkasteluun ja niiden kehittämistä varten.

Litteenä oleva tutkimuslupa palautetaan kouluun omalle opettajalle viimeistään perjantaina 1.2.2013. Vastaan mielelläni kaikkiin tutkimusta tai simulaatiokokeilua koskeviin kysymyksiin.

Ystävällisin terveisin,

Taina Wewer, luokanopettaja, tohtorikoulutettava, Turun yliopisto
taina.wewer@utu.fi, 040-5242921

OPPILAAN TUTKIMUSLUPA SIMULAATIOKYSELYÄ VARTEN

Palautetaan omalle opettajalle viimeistään 1.2.2012.

Tutkimusluvassa mainittu tietokonesimulaatiokysely on tarkemmin selvitetty erillisellä lupapaperin mukana tulleella tiedotteella.

_____ (oppilaan nimi)

_____ (luokka)

Oppilas saa luvan suoritettuaan ███vierailupäivänä tietokonesimulaation osallistua väitöstutkimuskyselyyn, jossa tiedustellaan mm. mielipiteitä ja kokemuksia simulaatiosta sekä arviota omasta suoriutumisesta. Lisäksi hän saa halutessaan osallistua lyhyeen samanaiheiseen haastatteluun.

Oppilas ei saa osallistua tutkimuskyselyyn.

(Sopiva vaihtoehto rastitetaan.)

_____ (huoltajan allekirjoitus)

_____ (huoltajan nimenselvennys)

Arviointikysely kieliluokkien opettajat 2012

Tällä kyselyllä kartoitetaan CLIL-luokissa käytössä olevia erilaisia englannin kielen taitotason arviointimenetelmiä, kartoitetaan esteitä ja puutteita arvioinnin suorittamiselle sekä kysytään visioita sekä tarpeita uuden arviointikäytännön kehittämiseksi.

Kaikki vastaukset ovat erittäin tärkeitä, joten pyydän pohtimaan ja kirjaamaan ajatukset vapaisiin vastauskenttiin mahdollisimman tarkasti. Tarvittaessa voit jatkaa paperin kääntöpuolelle.

Huom! Tässä kyselyssä kaikki oppilaan englannin kielitaidon ja osaamisen arviointia koskevat kysymykset tarkoittavat muussa oppiaineessa kuin formaalienglannin opiskelussa tapahtuvaa arviointia tai osaamisen kartoitusta.

Kyselyyn voit vastata ajatellen laajemmin kuin pelkästään tätä kuluva lukuvuotta koskien.

Taustatiedot:

1. Kuinka kauan olet toiminut vieraskielisen CLIL -luokan opettajana? Rastita.

- 0-2 vuotta
- 3-5 vuotta
- 6-9 vuotta
- 10 -14 vuotta
- 15 vuotta tai enemmän

2. Mitä CLIL-luokkaa opetat?

- 1. luokka
- 2. luokka
- 3. luokka
- 4. luokka
- 5. luokka
- 6. luokka

Arviointikysely

1. Pidätkö arviointitiedon keräämistä oppilaittesi englannin kielen taidon tason ja sen kehittymisen seuraamista tärkeänä muissa oppiaineissa? Rastita.

- erittäin tärkeänä hyvin tärkeänä melko tärkeänä vähän tärkeänä en lainkaan tärkeänä

2. Millä tavoin keräät tietoa oppilaittesi englannin kielen osaamisesta oppitunneillasi? Rastita.

- en kerää tietoa systemaattisesti
- observointi
- kirjalliset kokeet, testit tai koeosiot
- suulliset kuulustelut tai haastattelut
- vuorokeskustelut, vuorovaikutustilanteet
- esitykset
- esseet tai kirjoitelmat
- simuloivat harjoitteet
- oppilaiden itsearviointi ja reflektio
- oppilaiden välinen vertaisarviointi
- muun toimijan (esim. natiiviopettaja, avustaja) havainnot
- portfolio tai kielisalkku
- muu, mikä/mitkä?

3. Millä tavoin informoit tai olet informoinut oppilaita heidän kielitaitonsa tasosta tai osaamisesta?

- en ole informoinut
- suullinen palaute
- kirjallinen palaute (esim. kokeen tai itsearvioinnin yhteydessä)
- erillinen tiedote
- kehityskeskustelu tai ei-formaali keskustelu kahden kesken
- luokkakeskustelu tai yhteinen palautekeskustelu
- lukuvuositodistus
- muu, mikä/mitkä?

4. Kuinka usein/säännöllisesti annat oppilaille palautetta heidän kielitaidostaan, sen tasosta tai edistymisestään englannin kielessä?

5. Tiedotatko oppilaiden huoltajia heidän lastensa kielitaidon tasosta, osaamisesta tai sen kehittymisestä?

kyllä en

Jos vastasit ”kyllä”, vastaa myös kysymyksiin 6-8. Jos vastasit ”en”, hyppää suoraan kysymykseen 9.

6. Kuinka usein tiedostat koteja oppilaittesi englannin kielen tasosta tai osaamisesta?

säännöllisesti silloin tällöin harvoin

7. Kerro tarkemmin aikamäärein (esim. kerran lukuvuodessa, viikoittain) , miten usein tiedostat huoltajille koteihin oppilaittesi englannin kielen tasosta tai osaamisesta.

8. Millä tavoin tiedostat huoltajille?

9. Mitkä asiat koet hankaliksi tai haastaviksi CLIL-oppilaiden kielellisen osaamisen, kielitaidon tason tai sen kehittymisen arvioinnissa?

10. Miten mielestäsi CLIL-arviointia voisi kehittää tai minkälaisia arviointityökaluja tarvittaisiin?

Haastattelu

Jos haluat vapaaehtoisena osallistua myöhemmin kevätlukukaudella toteutettavaan syventävään, nauhoitettavaan haastatteluun tästä samasta aiheesta, pyydän kirjoittamaan nimesi ja yhteystietosi alla olevaan tyhjään tilaan yhteydenottoa varten. Haastateltavan nimeä tai tunnistetietoja ei julkaista tutkimuksessa, vaan ne jäävät ainoastaan tutkijan tietoon.

Lämmin kiitos vastauksistasi ja hyvää kevätlukukauden jatkoa 2012!

Taina Wewer, KM, luokanopettaja, tohtoriopiskelija, taiwew@utu.fi, 040-5242921

Arviointikysely 1

kevätlukukausi 2012

Kieliluokkien oppilaat

██████████ koulu ja ██████n koulu, ██████

Tässä paperissa kysytään sinun mielipiteitäsi ja käsityksiäsi niistä tiedoista, joita sinä englantiluokan oppilaana saat koulusta tai opettajaltasi omasta englannin kielen taidostasi **muilla kuin englannin kielen tunneilla**.

Vastaamalla tähän kyselyyn autat tutkijaa selvittämään, minkälaista tietoa kielitaidostaan oppilaille annetaan ja miten. Tutkimuksessa selvitetään myös, minkälaista tietoa oppilaat haluaisivat omasta osaamisestaan saada ja miten.

Sinun ei tarvitse kirjoittaa nimeäsi, eikä sinua voi tutkimuksessa mitenkään tunnistaa.

Lue kysymykset tarkasti. Vastaa rehellisesti siten kuin sinä ajattelet. Pyydä opettajalta tai tutkijalta apua, jos et ymmärrä jotain kohtaa.

Kiitos!

Taustatiedot (rastita)

1. Minä olen tyttö
 poika.
2. Olen kolmasluokkalainen
 neljäsluokkalainen
 viidesluokkalainen.
3. Kouluni on ██████n koulu
 ██████n koulu.

Arviointikysely

Huom! Alla tarkoitetaan englannin kielen osaamista muissa oppiaineissa (esim. matematiikka, musiikki, ympäristöoppi, biologia) kuin englannin tunneilla. Voit vastata ajatellen koko kouluaikaasi ensimmäisestä luokasta alkaen – ei vain nykyistä luokkaa koskien.

Rastita lähinnä omaa mielipidettäsi oleva vaihtoehto.

1. Onko sinulle tärkeää saada tietoa siitä, miten hyvin olet oppinut englannin kieltä muissa oppiaineissa?
 hyvin tärkeää jonkin verran tärkeää ei tärkeää

2. Millä tavoin olet saanut koulussa tietoa omasta englannin kielen taidostasi muilla kuin englannin tunneilla? Rastita.

- opettajan suullinen palaute tunnilla / opettaja kertoo
 - kokeet
 - itsearviointi
 - luokkatoverit
 - muu opettajan, harjoittelijan tai avustajan palaute
 - koulutodistus
 - kuukausi- tms. raportti
 - kehityskeskustelu tai vanhempainvartti
 - koulutehtävistä selviytyminen
 - portfolio tai kansiotyöskentely
 - muu, mikä?
-
-

3. Kuinka usein saat koulussa tietoa tai palautetta omasta englannin kielitaidostasi?

- hyvin usein silloin tällöin harvoin en koskaan

4. Millä tavoilla sinä haluaisit saada tietoa omasta kielitaidostasi ja edistymisestäsi englannin kielellä opiskelusta? Rastita.

- opettajan suullinen palaute tunnilla / opettaja kertoo
 - kokeet
 - itsearviointi
 - luokkatoverit
 - muu opettaja, harjoittelija tai avustaja antaa palautetta
 - koulutodistus
 - kuukausi- tms. raportti
 - kehityskeskustelu tai vanhempainvartti
 - muu, mikä?
-
-

5. Haluaisitko saada enemmän tietoa siitä, miten pärjää englanniksi muissa oppiaineissa?

- kyllä en osaa sanoa en

6. Jos sinä olisit kieliluokkaopettaja, minkälaisia uusia tapoja sinä kehittäisit oppilaillesi, jotta he voisivat osoittaa kielitaitoaan? Keksi ja kirjoita.

Haastattelu

Haluatko tulla haastateltavaksi myöhemmin?

Vapaaehtoisia oppilaita haetaan haastatteluun, jossa kysytään enemmän ja paremmin niistä tilanteista, joissa sinun kielitaitoasi on arvioitu. Tutkija haluaa tietää myös tarkemmin, miten ja minkälaista tietoa sinä tulevaisuudessa haluaisit saada oman kielitaitosi tasosta ja kehittämisestä.

Haastattelut nauhoitetaan, eikä sinun henkilöllisyyttäsi paljasteta missään yhteydessä.

Haastattelunauhoja käsittelee vain tutkija, ei kukaan muu.

Jos haluat olla vapaaehtoinen tutkimushaastateltava, kirjoita oma nimesi alla olevaan tyhjään tilaan, niin sinut voidaan kutsua myöhemmin jonkin toisen koulupäivän aikana haastatteluun.

Kiitos.

Arviointikysely 1 kieliluokkien oppilaiden huoltajat 2012

██████████ koulu ja ████████ n koulu, ████████

Tällä kyselyllä kartoitetaan vieraskielisten CLIL-luokkien (=englantiluokat) oppilaiden huoltajien kokemuksia lapsiensa kielitaidosta ja sen kehittymisestä saatavasta tiedosta KM Taina Wewerin väitöstutkimusta ”CLIL-oppilaan kielitaidon profilointi ja arviointi alakoulussa” varten. Kyselyyn voitte halutessanne vastata siinäkin tapauksessa, että oppilaalle ei ole annettu lupaa osallistua oppilaskyselyyn.

Täytetyn tutkimuskyselyyn voitte palauttaa 15.2.2012 mennessä tässä samassa tunnuksettomassa kirjekuoressa oppilaan mukana takaisin luokkaan, jossa on palautuslaatikko niitä varten. Luokan opettaja ei kerää kyselyjä. Kuoren voi lisäksi sulkea, jolloin kukaan muu kuin tutkija ei näe vastauksia.

Koska kysely on nimetön, ei tietoja pysty millään tavalla yksilöimään tai jäljittämään takaisin vastaajaan. Kaikki tieto on tutkimukselle luottamuksellista ja arvokasta, joten kiitän lämpimästi jokaista vastaajaa.

Taina Wewer
luokanopettaja, tohtoriopiskelija

Taustatiedot:

Miksi lapsenne opiskelee vieraskielisen opetuksen luokalla? Rastittakaa sopivat vaihtoehdot.

- paremman kielitaidon saavuttaminen
- menestyminen työelämässä
- englantilaisen kielialueen kulttuuriin tutustuminen
- sukulaiset tai ystävät ulkomailla
- matkailuun liittyvät syyt
- lapsen oma toive
- vanhempien päätös
- muu syy (Mikä? Mitkä? Tarvittaessa vastausta voi jatkaa

kääntöpuolelle.)

Arviointikysely

Huom! Seuraavat kysymykset eivät liity englannin oppiaineeseen, vaan niillä viitataan lapsenne englannin kielen omaksumiseen muiden oppiaineiden (esimerkiksi matematiikka, ympäristötieto tai biologia) tunneilla.

Kysely kattaa koko lapsen vieraskielisellä luokalla opiskelun – ei pelkästään nykyistä lukuvuotta. Voitte siis ajatella tässä opiskelua ensimmäisestä luokasta alkaen. Viivoille voi halutessaan kirjoittaa tarkentavaa lisätietoa, joka on tutkijalle erityisen arvokasta.

Saatteko mielestänne tarpeeksi tietoa lapsenne englannin kielitaidosta ja sen tasosta?

kyllä ei

Saatteko mielestänne tarpeeksi tietoa lapsenne edistymisestä englannin kielessä?

kyllä ei

Millä tavoin olette tähän mennessä saaneet tietoa lapsenne englannin kielitaidosta tai edistymisestä? Rastittakaa sopivat vaihtoehdot.

- lapsen puheen tai kirjoittamisen havainnointi englannin kielen käyttötilanteissa (esim. koulutehtävät, ulkomaanmatkat)
- lapsen oma kertomus osaamisestaan
- koulu- tai kotitehtävät
- kokeet
- oppilaan itsearviointi (esim. kaavake)
- koulutodistus

- opettajan muu tiedote (esim. kuukausitiedote)
- kehityskeskustelu tai vanhempainvartti
- muu keskustelu tai viestittely opettajan kanssa (esim. Wilma, s-posti)
- portfolio tai kielisalkku
- esiintymiset esim. koulun juhlissa tai esiintymisilloissa
- muu tapa (Mikä? Mitkä?)

Haluaisitteko saada vielä enemmän tietoa lapsenne englannin kielitaidon tasosta ja edistymisestä?

- kyllä ei

Haluaisitteko saada tarkempaa, eritellympää tietoa lapsenne englannin kielitaidon tasosta ja edistymisestä?

- kyllä ei

Missä muodossa haluaisitte enemmän tai tarkempaa tietoa lapsenne englannin kielitaidon tasosta tai edistymisestä?

- kokeet tai koeosiot
- oppilaan itsearviointi
- koulutodistus
- opettajan muu tiedote
- kehityskeskustelu tai vanhempainvartti
- muu keskustelu tai viestittely opettajan kanssa
- portfolio tai kielisalkku
- muu tapa (Mikä? Mitkä?)

Mahdollisuus osallistua syventävään tutkimushaastatteluun

Myöhemmin kevättalvella 2012 toteutetaan vapaaehtoinen ja nauhoitettava lisähaastattelu, jossa tarkastellaan syvällisemmin kieliluokan oppilaan huoltajan näkökulmaa oman lapsensa kielitaidon arvioinnista saatuun tietoon sekä toiveita siitä, kuinka paljon ja minkälaista arviointitietoa huoltaja haluaisi lapsestaan saada.

Jos haluatte osallistua vapaaehtoisena haastatteluun, pyydän kirjoittamaan alla olevaan tyhjään tilaan nimenne ja yhteystietonne, jotta haastatteluajankohdasta voidaan sopia.

Haastateltavan nimeä tai tietoja ei julkaista tutkimuksessa, vaan ne jäävät ainoastaan tutkijan tietoon.

Kiitos.

HAASTATTELURUNKO OPETTAJAT

Kerrotko aluksi, mitä ajatuksia CLIL-englannin arviointi sinussa herättää.

Arvioinnin tarpeellisuus

kenelle: eri näkökulmat (opettaja, oppilas, huoltaja)

miksi: tavoitteet

mitä pitäisi arvioida

Koulun virallinen käytäntö: OPS + arviointi

määrä

laatu

riittävyys

Itse käytetyt arviointimenetelmät

käytännön esimerkkejä

ekonomisuus

hyödyllisyys

käyttökelpoisuus

luonnollinen konteksti

esiintymistiheys

Muiden käyttämiä arviointimenetelmiä

ks ed.

Mitkä asiat vaikuttavat käytettyihin arviointimenetelmiin

opettajan kielikoulutus/pätevyys

materiaalien saatavuus/tieto

Kehityskaari näkyviin

Arvioinnin haasteet ja vaikeudet

Onko uran aikana muutosta

Mitä toivoisi tulevaisuudessa

Muuta

HAASTATTELURUNKO OPPILAAT

Miten englannin kielellä opiskelu sujuu?

Kuinka tärkeää kielen oppiminen englantiluokkaopetuksessa sinulle on?

Mitä sana arviointi mielestäsi tarkoittaa?

Onko sinun kielitaitoasi arvioitu koulussa (englannin tunteja lukuun ottamatta)?

Miten? Kerro esimerkkejä.

Voiko arviointi auttaa oppimisessa? Miten?

Vaikuttaako halukkuuteen oppia? Miten?

Minkälaisista arviointitavoista sinä pitäisit/olisivat sinulle sopivia? Miksi?

Mitä asioita haluaisit tietää omasta kielitaidostasi? Miten?

Saatko koulusta riittävästi tietoa omasta kielitaidosta? Sen kehittymisestä?

Mitä hyötyä arvioinnista sinulle? Haittaa?

Kuka olisi hyvä arvioitsija?

Onko sinulla kokemusta eri opettajien arviointitavoista?

Miten arviointitietoa oppilaiden kielitaidosta voitaisiin kerätä enemmän? Paremmiin?

Muuta?

HAASTATTELURUNKO HUOLTAJAT

odotukset lapsen kieliluokalla opiskeluun liittyen toteutuminen

Miten kieliluokkaopiskelu näkyy kotona? muualla? esimerkkejä

Onko merkkejä kielitaidon karttumisesta? esimerkkejä

Miten koulusta viestitetään oppilaan kielitaidosta? tyytyväisyys koulusta tulevaan informaatioon

määrä

laatu

tiheys

lapsen suhtautuminen – arviointistressi - motivaatio

eri opettajien/luokkien arviointi- tai tiedotuskäytänteet (muilta kuultua)

mikä vaikuttaa

yhtenäinen linja / erilaisuus

opetussuunnitelma: sisällöt ja arviointi opettajan koulutustausta

Minkälaista arvioinnin pitäisi olla?

lapsiystävällisyys

kehityskaaren näkyminen

toiveet, muutokset

Muuta?

Taustakysymykset:

- Koulusi? 2. Luokkasi? 3. Sukupuolesi?
-  5. lk tyttö
-  6. lk poika

Simulaatiokysymykset

Vastaa alla oleviin äskeistä simulaatiosuoritustasi koskeviin kysymyksiin mahdollisimman tarkasti. Kysymysten alla oleville viivoille voit tarvittaessa kirjoittaa asiaan liittyviä lisäkommentteja.

Ymmärsitkö ennen simulaatiosuoritusta annetut toimintaohjeet? *Rastita sopiva vaihtoehto.*

- kyllä en en osaa sanoa

Kuinka paljon sinua jännitti simulaation alussa? *Rastita sopiva vaihtoehto.*

- erittäin paljon paljon jonkin verran vähän ei lainkaan

Mitä jännitykselle (jos sitä oli) tapahtui simulaation aikana?

- pysyi ennallaan väheni kasvoi

Mikä seuraavista väittämistä kuvaa parhaiten omaa panostasi simulaatiossa? *Rastita sopiva vaihtoehto.*

- Yritin parhaani.
 Olisin halutessani pystynyt parempaan.
 En yrittänyt tosissani.

Miten omasta mielestäsi selvisit yleisesti ottaen tehtävistä englannin kielellä? *Rastita sopiva vaihtoehto.*

- erittäin hyvin melko hyvin kohtalaisesti melko huonosti huonosti

Mitkä tehtävistä olivat sinulle kielellisesti helppoja? *Merkitse helpoin numerolla 1, toiseksi helpoin numerolla 2 ja kolmanneksi helpoin numerolla 3. Älä merkitse enempää.*

- _____ itsestä kertominen (koulu, harrastukset jne.)
_____ ääneen lukeminen (USA/Michigan-teksti)
_____ sanastoaukkotehtävä (USA/Michigan-sanojen siirtäminen)
_____ Suomesta kertominen kuvien avulla
_____ mustakarhutehtävä (tekstin ymmärtäminen ja sanojen selitys suomeksi)
_____ laskuongelman selittäminen (puuttuva numero)
_____ allekkainkertolaskun selittäminen
_____ postikortin kirjoittaminen
_____ kellonajan kertominen

Mitkä tehtävistä olivat sinulle kielellisesti vaikeita? *Merkitse vaikein numerolla 1, toiseksi vaikein numerolla 2 ja kolmanneksi vaikein numerolla 3. Älä merkitse enempää.*

- _____ itsestä kertominen (koulu, harrastukset jne.)
_____ ääneen lukeminen (USA/Michigan-teksti)
_____ sanastoaukkotehtävä (USA/Michigan-sanojen siirtäminen)
_____ Suomesta kertominen kuvien avulla
_____ mustakarhutehtävä (tekstin ymmärtäminen ja sanojen selitys suomeksi)
_____ laskuongelman selittäminen (puuttuva numero)
_____ allekkainkertolaskun selittäminen
_____ postikortin kirjoittaminen
_____ kellonajan kertominen

Miten pärjäsit omasta mielestäsi seuraavissa kielitaidon osa-alueissa? *Numeroi kaikki kohdat järjestykseen 1-4 (parhaiten 1, toiseksi parhaiten 2 jne.).*

- _____ puheen ymmärtäminen
 _____ tekstin ymmärtäminen
 _____ puhuminen
 _____ kirjoittaminen

Nyt simulaation tehtyäsi mikä seuraavista kuvauksista sopii mielestäsi parhaiten englannin yleiseen kielitaitoosi? *Merkitse rasti sen sarakkeen eteen.*

A1	Ymmärrän ja osaan ilmaista tuttuja arkipäivän asioita, kuten "I'm hungry.". Osaan esitellä itseni ja muita. Osaan vastata kysymyksiin itsestäni (esim. nimi, asuinpaikka, omistamani tavarat) ja kysyä samaa muilta. Pystyn keskustelemaan englanniksi, jos toinen puhuu hitaasti, käyttää yksinkertaisia ilmauksia ja auttaa.
A2	Ymmärrän yksinkertaisia lauseita ja ilmauksia, jotka liittyvät esimerkiksi ostosten tekoon, omaan perheeseen ja koulunkäyntiin. Pystyn kertomaan tutuista, jokapäiväisistä asioistani (esim. harrastukset ja ympäristö) ja kuvailemaan omaa elämäni ja perhettäni.
B1	Ymmärrän pääkohdat selkeistä teksteistä (esim. tarinat, peliohjeet). Selviän useimmista tilanteista matkustaessani englantia puhuvissa maissa. Osaan kirjoittaa itseäni kiinnostavista aiheista (esim. lempimusiikki) ja selittämään mielipiteitäni. Pystyn myös kuvailemaan tapahtumia, unelmia ja kokemuksia.
B2	Ymmärrän pääajatuksia vaativampiakin aiheita (esim. ilmastonmuutos) käsittelevistä teksteistä. Pystyn puhumaan vaivattomasti ja sujuvasti englantia

		puhuvien kanssa. Pystyn kirjoittamaan hyvin erilaisista aiheista. Osaan esittää mielipiteeni jostain ajankohtaisesta asiasta (esim. uutiset) ja perustelemaan asian haitat ja edut.
	C1	Ymmärrän erityyppisiä, pitkiä tekstejä ja sanojen piilomerkityksiä. Pystyn esittämään ajatuksiani ilman vaikeuksia löytää sanoja ja ilmaisuja. Osaan käyttää kieltä kaikenlaisilla elämän osa-alueilla (esim. terveydenhoito, työ). Pystyn kirjoittamaan monimutkaisista aiheista hyvin rakentunutta, selkeää ja yksityiskohtaista tekstiä, jossa on yhdylauseita.
	C2	Ymmärrän vaikeuksitta kaikenlaista kirjoitettua ja puhuttua kieltä. Osaan yhdistää tietoja eri lähteistä ja muodostaa niistä yhtenäisen, yksityiskohtaisen esityksen. Pystyn ilmaisemaan ajatuksiani erittäin sujuvasti ja täsmällisesti. Erotan kielessä paljon eri merkitysvaihtoehtoja.

Rastita mielipidettäsi vastaava vaihtoehto kuhunkin kohtaan. Onko sinun mielestäsi tällainen simulaatio hyvä tapa arvioida oppilaan englannin kielen

yleistä kielitaitoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
ääntämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
puheen sujuvuutta?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
sanavarastoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
luetun ymmärtämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
kuullun ymmärtämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
lukutaitoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
oikeinkirjoitusta?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
rakenteiden hallintaa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
kieliopin hallintaa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
yrittämistä ja ponnistelua?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
kulttuuritietoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa

Onko sinun mielestäsi tällainen simulaatio hyvä tapa arvioida sitä, miten oppilas osaa eri oppiaineiden (esim. matematiikka tai ympäristötieto) asioita englanniksi?

kyllä ei en osaa sanoa

Jatka seuraavia lauseita.

Minusta simulaatio oli _____

Haluaisin, että simulaatioita _____

Simulaatiokokemus oli minusta _____

Kaikkein parasta simulaatiossa oli _____

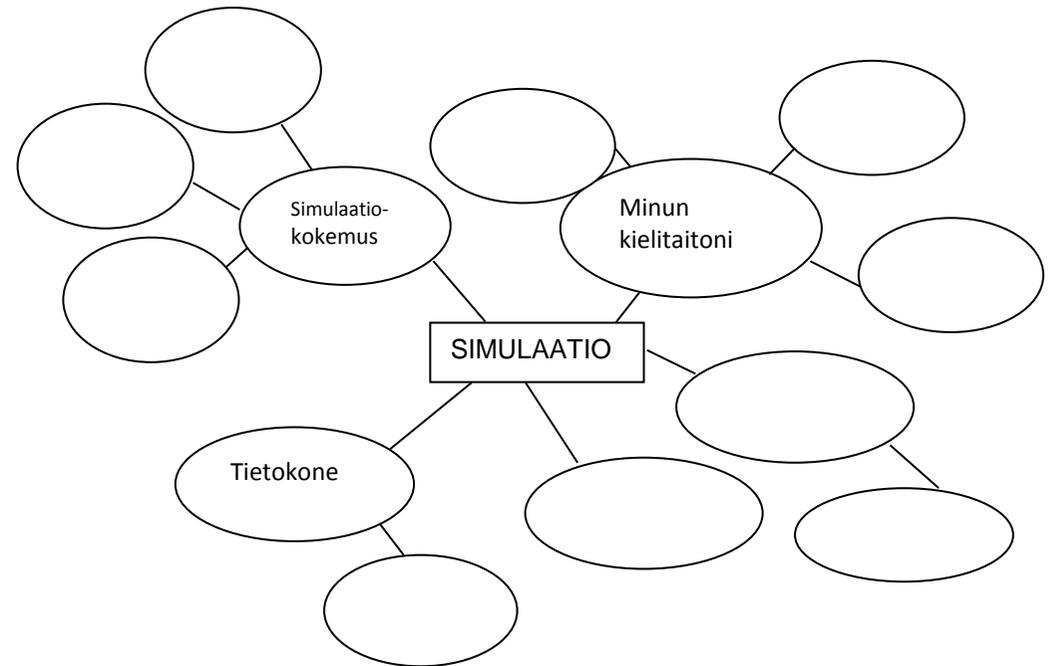
Kaikkein kamalinta simulaatiossa oli _____

Simulaation avulla huomasin, että _____

Simulaatiosta sanoisin kavereilleni, että _____

Mitä parannettavaa tai muutettavaa tässä simulaatiossa olisi? Miten siitä saisi tehtyä paremman?

Täytä alla oleva miellekartta niin hyvin kuin pystyt. Voit tarvittaessa lisätä siihen osia.



Taustakysymykset:

Koulusi?

2. Sukupuolesi?

 tyttö

 poika
Kielitaustakysymykset: kieliprofiili

Ennen simulaatiota täytit omasta kielitaidostasi perustietoja. Seuraava kysymys liittyy siihen.

Oliko sinusta mukavaa tuoda esiin oman kielitaidon osaamista kielitaito-osuudessa?

 kyllä

 ei

 en osaa sanoa

Miksi?

Simulaatiokysymykset

Vastaa alla oleviin äskeitä simulaatiosuoritustasi koskeviin kysymyksiin mahdollisimman tarkasti. Kysymysten alla oleville viivoille voit kirjoittaa asiaan liittyviä lisäkommentteja.

Ymmärsitkö simulaation aikana annetut ohjeet? *Rastita sopivin vaihtoehto.*

 kyllä

 en

 en osaa sanoa

Kuinka paljon sinua jännitti simulaation alussa? *Rastita sopivin vaihtoehto.*

 erittäin paljon

 paljon

 jonkin verran

 vähän

 ei lainkaan

Mitä jännitykselle tapahtui simulaation aikana? *Rastita sopivin vaihtoehto.*

 pysyi ennallaan

 väheni

 kasvoi

 ei sitä ollut

Mikä seuraavista väittämistä kuvaa parhaiten omaa panostasi simulaatiossa? *Rastita sopivin vaihtoehto.*

 Yritin parhaani.
 Olisin halutessani pystynyt parempaan.
 En yrittänyt tosissani.

Miten omasta mielestäsi selvisit yleisesti ottaen tehtävistä englannin kielellä? *Rastita sopivin vaihtoehto.*

 erittäin hyvin

 melko hyvin

 kohtalaisesti

 melko huonosti

 huonosti

Arvioi simulaation tehtäviä. Kuinka helppo tai vaikea sinun oli selvittää simulaation tehtävistä englanniksi? *Merkitse tehtävän alle rastilla vaikeustaso.*

oppilaitten kysymyksiin vastaaminen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

asioiden nimeäminen kuvasta

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

omista välituntipeleistä ja leikeistä kertominen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

British Bulldogs -leikin sääntöjen selittäminen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

Coca-colan katoamisen selittäminen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

Tulipalo-ohjeiden järjestäminen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

Tulipalo-ohjeiden lukeminen ääneen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

omasta lempisarjakuvasta kertominen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

sarjakuvatekstin kirjoittaminen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

Viimeisen sarjakuvaruudun tapahtumista kertominen

 hyvin vaikea

 vaikea

 ei vaikea, ei helppo

 helppo

 hyvin helppo

Outojen hahmojen läsnäolon selittäminen

hyvin vaikea vaikea ei vaikea, ei helppo helppo hyvin helppo

Outojen hahmojen tuntomerkkien kirjoittaminen

hyvin vaikea vaikea ei vaikea, ei helppo helppo hyvin helppo

oman suoriutumisen arviointi simulaatiossa

hyvin vaikea vaikea ei vaikea, ei helppo helppo hyvin helppo

Kirjoita tehtävän nimi tai käytä edellisen tehtävien kirjainkoodeja.

Mikä tehtävistä oli sinulle kaikkein helpoin?

Miksi?

Kaikkein vaikein?

Miksi?

Miten pärjäsit omasta mielestäsi seuraavissa kielitaidon osa-alueissa? *Merkitse rastilla.*

Englanninkielisen puheen ymmärtäminen oli minulle

hyvin vaikeaa vaikeaa ei vaikeaa eikä helppoa helppoa hyvin helppoa.

Englanninkielisen tekstin ymmärtäminen oli minulle

hyvin vaikeaa vaikeaa ei vaikeaa eikä helppoa helppoa hyvin helppoa.

Englannin puhuminen oli minulle

hyvin vaikeaa vaikeaa ei vaikeaa eikä helppoa helppoa hyvin helppoa.

Englannin kirjoittaminen oli minulle

hyvin vaikeaa vaikeaa ei vaikeaa eikä helppoa helppoa hyvin helppoa

Mikä seuraavista kuvauksista sopii mielestäsi parhaiten englannin yleiseen kielitaitoosi? *Merkitse rasti sarakkeen eteen. Merkitse VAIN yksi rasti.*

A1	Ymmärrän ja osaan ilmaista tuttuja arkipäivän asioita, kuten "I'm hungry.". Osaan esitellä itseni ja muita. Osaan vastata kysymyksiini itsestäni (esim. nimi, asuinpaikka, omistamani tavarat) ja kysyä samaa muilta. Pystyn keskustelemaan englanniksi, jos toinen puhuu hitaasti, käyttää yksinkertaisia ilmauksia ja auttaa.
A2	Ymmärrän yksinkertaisia lauseita ja ilmauksia, jotka liittyvät esimerkiksi ostosten tekoon, omaan perheeseen ja koulunkäyntiin. Pystyn kertomaan tutuista, jokapäiväisistä asioistani (esim. harrastukset ja ympäristö) ja kuvailemaan omaa elämääni ja perhettäni.
B1	Ymmärrän pääkohdat selkeistä teksteistä (esim. tarinat, peliohjeet). Selviän useimmista tilanteista matkustaessani englantia puhuvissa maissa. Osaan kirjoittaa itseäni kiinnostavista aiheista (esim. lempimusiikki) ja selittämään mielipiteitäni. Pystyn myös kuvailemaan tapahtumia, unelmia ja kokemuksia.
B2	Ymmärrän pääajatuksia vaativampiakin aiheita (esim. ilmastonmuutos) käsittelevistä teksteistä. Pystyn puhumaan vaivattomasti ja sujuvasti englantia puhuvien kanssa. Pystyn kirjoittamaan hyvin erilaisista aiheista. Osaan esittää mielipiteeni jostain ajankohtaisesta asiasta (esim. uutiset) ja perustelemaan asian haitat ja edut.
C1	Ymmärrän erityyppisiä, pitkiä tekstejä ja sanojen piilomerkityksiä. Pystyn esittämään ajatuksiani ilman vaikeuksia löytää sanoja ja ilmaisuja. Osaan käyttää kieltä kaikenlaisilla elämän osa-alueilla (esim. terveydenhoito, työ). Pystyn kirjoittamaan monimutkaisista aiheista hyvin rakentunutta, selkeää ja yksityiskohtaista tekstiä, jossa on yhdylauseita.
C2	Ymmärrän vaikeuksitta kaikenlaista kirjoitettua ja puhuttua kieltä. Osaan yhdistää tietoja eri lähteistä ja muodostaa niistä yhtenäisen, yksityiskohtaisen esityksen. Pystyn ilmaisemaan ajatuksiani erittäin sujuvasti ja täsmällisesti. Erotan kielessä paljon eri merkitysvivahteita.

Rastita mielipidettäsi vastaava vaihtoehto kuhunkin kohtaan. Onko sinun mielestäsi tällainen simulaatio hyvä tapa arvioida oppilaan englannin kielen

yleistä kielitaitoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
ääntämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
puheen sujuvuutta?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
sanavarastoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
luetun ymmärtämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
kuullun ymmärtämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
lukutaitoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
oikeinkirjoitusta?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
rakenteiden hallintaa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
kieliopin hallintaa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
yrittämistä ja ponnistelua?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
kulttuuritietoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa

Onko sinun mielestäsi tällainen simulaatio hyvä tapa arvioida sitä, miten oppilas osaa eri oppiaineiden (esim. ympäristötieto, liikunta) asioita englanniksi? *Rastita sopivin vaihtoehto.*

kyllä ei en osaa sanoa

Oman suorituksen kuuntelu ja katselu

Simulaation lopuksi sait kuunnella omia vastauksiasi, nähdä kirjoittamasi tekstit ja oikeat vastaukset.

Seuraavat kysymykset liittyvät niihin.

Mitä huomasit kielitaidostasi tai osaamisestasi englannin kielellä, kun kuuntelit omia vastauksiasi?

Mitä huomasit kielitaidostasi tai osaamisestasi englannin kielellä, kun näit omat vastauksesi?

Lopuksi vielä kommenttisi seuraavista asioista:

Jatka seuraavia lauseita.

Minusta simulaatio oli

Haluaisin, että simulaatioita

Simulaatiokokemus oli minusta

Kaikkein parasta simulaatiossa oli

Kaikkein kamalinta simulaatiossa oli

Simulaation avulla huomasin, että

Simulaatiosta sanoisin kavereilleni, että

Jos saisit valita, ottaisitko vaikka ympäristötiedon kokeeksi mieluummin

perinteisen paperikokeen? tietokonesimulaation? (*Rastita mieluisampi.*)

Miksi?

Onko koulussa käyty simulaatiossa esiintyneitä asioita englanniksi muilla kuin englannin tunneilla?

kyllä, monia kyllä, joitakin ei kovin monia ei oikeastaan
 en ole varma

Muita kommentteja? Kirjoita paperin kääntöpuolelle.

1. Jännittkö sinua tällä kerralla yhtä paljon kuin ensimmäisessä simulaatiokokeilussa? *Rastita sopivin vaihtoehto.*

enemmän yhtä paljon vähemmän ei jännittänyt viimeksi eikä nytäkään

2. Kummasta simulaatiosta pidit enemmän? Rastita sopivin vaihtoehto.

ensimmäisestä toisesta molemmista saman verran

Miksi? _____

3. Kumpi simulaatio oli mielestäsi englannin kielen suhteen haastavampi?

ensimmäinen toinen

Miksi?

KYSELY HUOLTAJILLE SIMULAATIO 2 27.2.2013

Lapsenne on tänään koulussa suorittanut englanninkielisen tietokonesimulaatio 2:n pilottiversion, jossa on ollut erilaisia tehtäviä. Teillä on mahdollisuus katsella ja kuunnella oman lapsenne simulaatiosuorituista ja siten saada käsitys hänen englannin kielitaidostaan. Lapsi saa mielellään olla mukana. Ohjeet ovat alla.

Huom! Simulaatio ja arviointilaboratorio ovat vasta kokeiluasteella, joten on mahdollista, että kaikki ei toimi odotetulla tavalla.

Koska kyseessä on ainutlaatuinen pilottikokeilu, toivon mahdollisimman monen lapsensa suorituksen katsoneen ja kuunnelleen huoltajan vastaavan tutkimuskyselyyn, joka palautetaan kouluun kirjekuoressa viimeistään ennen hiihtolomaa oman luokan opettajalle. Erityinen kiitos kaikille osallistujille arvokkaasta tutkimustiedosta. Toivottavasti ilo ja hyöty ovat molemminpuoleiset!

Ystävällisin terveisin,

Taina Wewer,
luokanopettaja, tohtorikoulutettava tutkija
Turun yliopiston normaalikoulu, taina.wewer@utu.fi 040 524 29 21

OHJEET SIMULAATION KATSOMISEEN KOTONA:

Avatkaa sivu <https://langperform-lab.uta.fi/lab> ja kirjautukaa tämän kyselyn mukana tulleella lapsenne omalla käyttäjätunnuksella ja salasanalla palvelimeen. Olkaa tarkkoina, että merkit tulevat oikein ja oikeassa muodossa. Etusivulla avautuu päävalikko, jossa on kolme alaotsikkoa:

- Profiilini (oppilaan kieliprofiili, jonka voitte halutessanne yhdessä lapsen kanssa täydentää)
- Suoritukset (tarkemmat ohjeet alla)
- Aloita simulaatio (simulaation voi tehdä kotona uudelleen mikrofonikuulokkeiden kanssa)

Klikatkaa kohtaa 'Suoritukset', jolloin ruudulla näkyy lapsenne simulaatiosuoritus. Sitä tuplaklikkaamalla pääsee kohtaan 'Suorituksen arviointi', missä lapsen

vastaukset kysymyksiin ja hänen kirjoittama teksti on nähtävissä sekä puhe kuunneltavissa audiopalkeissa kohta kohdalta. Valikosta 'Resources' pitäisi nähdä kulloisenkin tehtävän joko videona tai kuvana. Sivuja on useita.

Oikealla puolella ovat eri puhutun kielen osa-alueiden itsearviointiasteikot A1-C2, jotka noudattavat eurooppalaista kielten taitotasoaasteikkoa. Kenttään 'Lisätiedot' voitte halutessanne kirjoittaa huomioita ja kommentteja lapsen kielitaidosta ja suoriutumisesta. Viimeisellä sivulla, jos oppilas on arvioinut itseään, näkyy oppilaan oma kokonaisarvio koko suorituksesta. Kaikki muutokset arviointiin täytyy tallentaa painamalla kohtaa 'Save' sivun alalaidassa.

TUTKIMUSKYSELY

Taustakysymykset (Rastittakaa sopiva vaihtoehto.)

Lapsenne koulu? 2. Lapsenne luokka?

4. lk

5. lk

Simulaatiokysymykset

Kysymysten alla oleville viivoille voitte kirjoittaa asiaan liittyviä lisäkommentteja tai tarkennuksia ja lisätietoa vastauksiinne liittyen.

Onnistuiko simulaatiosuorituksen katselu/kuuntelu kotona?

kyllä ei

Jos ei onnistunut, niin mikä meni pieleen?

Oletteko saaneet/voineet aikaisemmin seurata lapsenne (joko saman tai sisaruksen) simulaatiosuoritusta?

kyllä ei, tämä oli ensimmäinen kerta ei, ensimmäinen ei onnistunut

Jos vastasitte kyllä, niin kumpi simulaatio oli mielestänne onnistuneempi?

ensimmäinen toinen (uudempi)

Miksi? _____

Katsoitteko simulaatiosuorituksen yhdessä lapsenne kanssa?

kyllä ei

Jos vastasitte kyllä, niin mistä asioista syntyi keskustelua? _____

Jos vastasitte ei, niin miksi ette katsoneet? _____

Mikä seuraavista kuvauksista sopii mielestänne parhaiten lapsenne yleiseen englannin kielitaitoon? *Merkittävä rasti sen sarakkeen eteen. Jos ette osaa arvioida, älkää rastittako mitään.*

A1	Hän ymmärtää ja osaa ilmaista tuttuja arkipäivän asioita, kuten "I'm hungry". Hän osaa esitellä itsensä ja muita. Hän osaa vastata kysymyksiin itsestään (esim. nimi, asuinpaikka, omistamansa tavarat) ja kysyä samaa muilta. Hän pystyy keskustelemaan englanniksi, jos toinen puhuu hitaasti, käyttää yksinkertaisia ilmauksia ja auttaa.
A2	Hän ymmärtää yksinkertaisia lauseita ja ilmauksia, jotka liittyvät esimerkiksi ostosten tekoon, omaan perheeseen ja

		koulunkäyntiin. Hän pystyy kertomaan tutuista, jokapäiväisistä asioistaan (esim. harrastukset ja ympäristö) ja kuvailemaan omaa elämäänsä ja perhettään.
	B1	Hän ymmärtää pääkohdat selkeistä teksteistä (esim. tarinat, peliohjeet). Hän selviää useimmista tilanteista matkustaessaan englantia puhuvissa maissa. Hän osaa kirjoittaa itseään kiinnostavista aiheista (esim. lempimusiikki) ja selittämään mielipiteitään. Hän pystyy myös kuvailemaan tapahtumia, unelmia ja kokemuksia.
	B2	Hän ymmärtää pääajatuksia vaativampiakin aiheita (esim. ilmastonmuutos) käsittelevistä teksteistä. Hän pystyy puhumaan vaivattomasti ja sujuvasti englantia puhuvien kanssa. Hän pystyy kirjoittamaan hyvin erilaisista aiheista ja osaa esittää mielipiteensä jostain ajankohtaisesta asiasta (esim. uutiset) ja perustelemaan asian haitat ja edut.
	C1	Hän ymmärtää erityyppisiä, pitkiä tekstejä ja sanojen piilomerkityksiä. Hän pystyy esittämään ajatuksiaan ilman vaikeuksia löytää sanoja ja ilmaisuja. Hän osaa käyttää kieltä kaikenlaisilla elämän osa-alueilla (esim. terveydenhoito, työ). Hän pystyy kirjoittamaan monimutkaisista aiheista hyvin rakentunutta, selkeää ja yksityiskohtaista tekstiä, jossa on yhdylauseita.
	C2	Hän ymmärtää vaikeuksitta kaikenlaista kirjoitettua ja puhuttua kieltä. Hän osaa yhdistää tietoja eri lähteistä ja muodostaa niistä yhtenäisen, yksityiskohtaisen esityksen. Hän pystyy ilmaisemaan ajatuksiaan erittäin sujuvasti ja täsmällisesti ja erottaa kielessä paljon eri merkitysvahteita.

Mitä huomioita teitte oman lapsenne suoriutumisesta ja osaamisesta englannin kielellä?

Saitteko simulaation avulla käsityksen siitä, miten lapsenne hallitsee eri oppiaineiden asioita englanniksi?

kyllä ei en osaa sanoa

Oletteko tietoinen siitä, mitä asioita tai oppiaineiden aihekokonaisuuksia koulussa on opiskeltu englannin kielellä?

kyllä ei en osaa sanoa

Onko teidän mielestänne tällainen simulaatio hyvä tapa arvioida sitä, miten oppilas osaa eri oppiaineiden (esim. ympäristötieto, liikunta) asioita englanniksi?

kyllä ei en osaa sanoa

Mitä kielitaidon mielestänne osa-alueita simulaatio parhaiten mittaa? *Númeroikaa järjestykseen 1-4 (parhaiten 1, toiseksi parhaiten 2, kolmanneksi parhaiten 3 ja heikoiten 4).*

_____ puheen ymmärtäminen
_____ puhuminen

_____ tekstin ymmärtäminen
_____ kirjoittaminen

Rastittakaa mielipidettänne vastaava vaihtoehto kuhunkin kohtaan. Onko teidän mielestänne tällainen simulaatio hyvä tapa arvioida oppilaan englannin kielen

yleistä kielitaitoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
ääntämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
puheen sujuvuutta?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
sanavarastoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
luetun ymmärtämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa

kuullun ymmärtämistä?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
lukutaitoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
oikeinkirjoitusta?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
rakenteiden hallintaa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
kieliopin hallintaa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
yrittämistä ja ponnistelua?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa
kulttuuritietoa?	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei	<input type="checkbox"/> en osaa sanoa

Jatkakaa seuraavia lauseita.

Minusta simulaatio oli _____

Haluaisin, että simulaatioita _____

Simulaation etu on _____

Simulaation haittapuoli on _____

Simulaation avulla huomasin, että _____

Muita ajatuksia tai kommentteja simulaatioon liittyen? Mitä parannettavaa tai muutettavaa tässä simulaatiossa olisi? Miten siitä saisi tehtyä paremman?

APPENDIX 11

SIMULATION TASKS embedded in simulation 1 (Accommodation in an American family due to school visit in Michigan)

Task	Purpose	BICS/CALP	Specific features for assessment (e.g.)
pushing the correct doorbell	warm-up	both	differentiating sounds
reacting to welcoming	warm-up	BICS	listening comprehension, speaking
introducing oneself (school, hobbies etc.)	testing	BICS	speaking, fluency, vocabulary
reading aloud (text about U.S. and Michigan)	testing	both	ability to read English language
gap filling exercise (USA/Michigan vocabulary)	testing	CALP	reading comprehension, subject-specific vocabulary (geography, biology)
telling about Finland with help of pictures	testing	CALP	subject-specific vocabulary (environmental and social sciences)
black bear (textual comprehension and explaining subject-specific vocabulary in Finnish)	testing	CALP	comprehension of subject-specific language, skimming, inferencing meaning from surrounding textual cues
explaining a mathematical problem (missing number)	testing	CALP	subject-specific language use and knowledge of mathematical terms
explaining a multiplication in columns	testing	CALP	subject-specific language use and knowledge of mathematical terms
writing a postcard	testing	BICS	past tense, expressions in writing interpersonal messages, orthography
telling the time	testing	BICS	mastery of conventions in telling time

SIMULATION TASKS in simulation 2 (Spending a day in an international school)

Task	Purpose	BICS/CALP	Specific features for assessment (e.g.)	Objective in Bloom's taxonomy (see p. 29)
Simple introduction of oneself	warm-up	BICS	name etc.	-
Answering the questions of pupils in the international school	testing	BICS	everyday issues (e.g. breakfast)	-
naming things	testing	CALP	subject-specific vocabulary in various subjects	remembering
telling about one's own games and activities during recesses	testing	both	speaking, fluency, vocabulary (e.g. P.E.)	remembering
explaining the rules of the game British bulldog	testing	both	listening comprehension, summarising	understanding
explaining the disappearance of the Coca-Cola	testing	CALP	using subject-specific language and vocabulary (environmental sciences/physics)	applying prior knowledge
organising instructions on what to do in case of fire	testing	CALP	understanding subject-specific language and vocabulary (environmental and social sciences)	analysing (organisation of parts, problem solving)
reading the organised text aloud	testing	CALP	ability to read English and pronunciation	analysing (letter-sound connection)
telling about one's favourite comic	orientation to the next set of tasks	BICS	listening comprehension, speaking, fluency	
writing text in the speech bubbles in a comic strip	testing	CALP	subject-specific language use (arts), orthography	analysing (drawing conclusions, synthesising integrating)
telling about the possible events in the last, empty frame	testing	both	future/present tense, expressions in writing interpersonal messages	evaluating (anticipating, reflecting)
Explaining why weird characters might have been lurking at school	testing	BICS	mastery of conventions in telling time	creating (generating)
Writing the description of the weird characters seen at school	testing	BICS	mastery of adjectives and concise expression	many objectives (e.g. remembering, applying, analysing etc.)
Assessing one's own performance	self-assessment	-	reflection of own language proficiency and performance	analysing and evaluating