

# Zeitschrift für Interkulturellen Fremdsprachenunterricht

Didaktik und Methodik im Bereich Deutsch als Fremdsprache

ISSN 1205-6545 Jahrgang 18, Nummer 2 (Oktober 2013)

---

## English language assessment in bilingual CLIL instruction at the primary level in Finland: Search for updated and valid assessment methods

Taina Wewer

Teacher Training School of Turku University  
Annikanpolku 9  
20610 Turku  
Finland

Email: [taina.wewer@utu.fi](mailto:taina.wewer@utu.fi)

**Abstract:** Bilingual instruction combining the teaching and learning of an additional, foreign language and subject content is better known as CLIL (Content and Language Integrated Learning), and it is constantly growing in popularity. The dual focus of simultaneous language and content learning is achieved by using the foreign language as the medium of instruction; the foreign language therefore is both the target and the medium of learning. The dual instructional focus presents a challenge to assessment in CLIL. Assessment is more demanding because the language aspect should be considered also; the measurement of language mastery – especially the use of academic, subject-specific language – ideally aligns with the measurement of content mastery.

In recent years, assessment has slowly drawn increasing attention in the field of CLIL research which has so far mainly dealt with issues such as language acquisition, affective factors and classroom discourse. The purpose of the research described in this article is to rise to the challenge of CLIL assessment, to survey the assessment practices currently used by CLIL class teachers in Finnish basic education and to experiment with computer simulation as a new, innovative assessment tool in CLIL contexts. First, I will give a very brief overview of CLIL in Finland. Then, in order to provide background information on my research, I will outline some of the challenges that complicate the administration of assessment in the Finnish CLIL environment. Subsequently I will present a short research and literature review of the topic and introduce the methods employed by the research which aims to increase an understanding of the complexity of CLIL assessment and seeks to broaden the range of assessment methods in CLIL. Finally, some preliminary results of the on-going research will be detailed.

Bilingualer Unterricht, der das Lehren und Lernen einer zusätzlichen Fremdsprache und eines Sachfaches vereint, ist besser bekannt als CLIL (Content and Language Integrated Learning; bilingualer Sachfachunterricht; integriertes Fremdsprachen- und Sachfachlernen). Die Doppelbetonung des gleichzeitigen Erlernens einer Sprache und eines Sachfaches wird dadurch ermöglicht, dass die Fremdsprache das Unterrichtsmedium ist; die Fremdsprache ist also Lernziel wie auch Lernmedium. Diese zweifache Zielausrichtung stellt eine Herausforderung für die Bewertung des Lernerfolgs im CLIL dar. Der Bewertungsprozess ist deshalb anspruchsvoller, weil die Sprachkomponente ebenfalls berücksichtigt werden soll. Die Evaluierung der Sprachbeherrschung – besonders der Verwendung der akademischen, sachfachspezifischen Sprache – steht im idealen Fall mit der Bewertung der Sachfachkenntnis im Einklang.

In den letzten Jahren hat der Evaluierungsprozess zunehmend an Aufmerksamkeit in der CLIL-Forschung gewonnen, die sich bis jetzt hauptsächlich mit Aspekten des Spracherwerbs, affektiver Faktoren und des Unterrichtsdiskurses beschäftigt hat. Die Absicht der in diesem Aufsatz dargestellten Forschungsarbeit ist es, sich der Herausforderung der Evaluierung im CLIL zu stellen, eine Bestandsaufnahme der gegenwärtig von CLIL-Lehrern im finnischen Grundschulunterricht verwendeten Bewertungspraktiken vorzunehmen und mit Computersimulationen als neuem, innovativem Bewertungswerkzeug im Rahmen des CLIL zu experimentieren. Zuerst werde ich einen sehr kurzen Überblick über CLIL in Finnland geben. Um Hintergrundinformationen zu meiner Forschung zu bringen, werde ich sodann einige Herausforderungen umreißen, die die Durchführung der Evaluierung im finnischen CLIL erschweren. In der Folge werde ich einen kurzen Überblick über die Forschung und Fachliteratur zum Thema liefern und die in meiner Forschung angewendeten Methoden vorstellen, welche auf ein besseres Verständnis der Komplexität der CLIL-Evaluierung abzielt und den Umfang der denkbaren Bewertungsmittel im CLIL zu erweitern sucht. Abschließend werden einige vorläufige Ergebnisse der noch laufenden Forschungsarbeit dargestellt.

**Key words:** Language assessment, bilingual education, content and language integrated learning, CLIL, language proficiency, primary education, Finland.

## 1. CLIL in Finland

Finland is a Nordic country with 5.5 million inhabitants and two official languages, Finnish and Swedish. Because these languages are rare, there is a need for Finns to acquire other languages. Finland was among the first European countries to implement CLIL and it is therefore sometimes referred to as the birth country of CLIL (e.g., Graddol 2006: 86). In 1991, a change in the Basic Education Act allowed the use of a foreign language other than Finnish or Swedish as the language of instruction. Membership in the European Union (EU) has also influenced Finland's language policy; 'the 1+2 standard' initiated by the European Commission in the mid-90s suggested that every EU citizen should master at least two supplementary languages in addition to their mother tongue (see European Commission 1995). For these reasons, CLIL instruction in basic education (classes 1-9, ages 7-15) and at upper secondary level increased steeply in the 1990s. However, the popularity of CLIL, according to follow-up studies in 1996 and 2005 (Nikula & Marsh 1996; Lehti, Järvinen & Suomela-Salmi 2006), has fluctuated since.

In 1996, 11.7% of schools at primary, lower and upper secondary levels provided CLIL instruction, but in 2005 the provision of CLIL had decreased to 5.7%. These surveys were conducted at the school level. The most recent survey from 2011 at the municipality level reveals that slightly over half of the municipalities currently offering CLIL instruction estimate that the demand for CLIL will grow again in the future (see Kangasvieri, Miettinen, Palviainen, Saarinen & Ala-Vähälä 2012). English is so far the most common language in CLIL, but other languages, such as Swedish, Finnish, German, French and Russian are also available, especially in larger municipalities (see Marsh, Järvinen & Haataja 2007: 70). In teaching the subject matter, CLIL instruction follows the guidelines, objectives and contents defined in the National Core Curriculum for Basic Education (NCC 2004), but the guidelines do not fully address CLIL language objectives and assessment issues. Furthermore, the guidelines are very general, which complicates their translation into the municipal or school CLIL curriculum.

### 1.1. First challenge: Diversity of CLIL provision

Each municipality and/or educational establishment should draw upon the National Core Curriculum (NCC) and compose a local level curriculum in which the pedagogical emphases such as the provision of bilingual instruction are determined. The NCC (2004: 270) states that the education provider is to decide what name will be used for bilingual instruction. For example, Kangasvieri et al. (2012) found 12 different labels for CLIL instruction in Finland, among which the terms "foreign language medium instruction" was the most frequent and "CLIL" the second. Not only the terms, but the implementation also vary greatly (see Kangasvieri et al. 2012). The 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", i.e. the extent of the instruction given in the foreign language has to be determined in advance. According to Järvinen (2005: 438), the proportion of foreign language instruction in Finnish CLIL classes is normally at least 20% of all instruction, which indicates that most of the instruction is given in the main language of schooling. Regular English as a subject is typically not included in that percentage.

It is noteworthy that the NCC does not prescribe which content should be taught through which language; the NCC leaves that decision to be met at the local level curriculum: "The [local] curriculum specifies what subject areas that support the study of different subjects will be taught in the foreign language" (NCC 2004: 271). Furthermore, and most importantly, it provides that the objectives of foreign language acquisition in the given CLIL context have to be predetermined: "As a minimum, the objectives specify what sort of level is to be reached in the course of basic education in listening and reading comprehension skills, speaking, writing and cultural skills" (ibid.: 270). As a consequence of these considerations, each CLIL-providing municipality or school should develop a CLIL curriculum containing the specifications of, at least, the proportion to which the foreign language should be used, the linguistic aims as well as the subjects and content of CLIL instruction. Unfortunately, this desideratum is not always realised. Curricula are open-access documents, often available online, and a short web survey unveils that not all curricula – if one even exists – contain properly elaborated language objectives as required by the NCC. If the aim

---

Taina Wewer (2013), 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. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-18-2/beitrag/Wewer.pdf>.

of CLIL implementation is, as it is expressed by one school, “to study in the foreign language”, the objective set is too broad and too vague. As Dalton-Puffer (2007: 295) so aptly puts it: “But why should we be doing CLIL at all if there are no language goals present?”

Due to the considerable pedagogical freedom invested in local authorities and individual schools, the diversity of CLIL in Finland is vast: CLIL in adjacent municipalities or schools may be implemented in a totally different manner, even contravening the descriptions for its provision (e.g. Pihko 2010: 125). This applies to the classrooms in the CLIL-offering schools as well. The absence of a CLIL curriculum, or the insufficiently detailed design of it, creates an obstacle for teachers to fulfil the curricular prerequisites of the NCC and to plan their instruction aiming at achieving the linguistic CLIL objectives. As a result, they choose and teach the content they see fit using concepts and language that best suits their own language proficiency level. This leads to varying CLIL quality and inconsistent educational CLIL models. Furthermore, such situations contribute to an inequality of outcomes among the students. Hence, the diversity in CLIL provision caused by inconsistent CLIL models and missing language objectives creates the first challenge to the CLIL assessment.

There are signs that this challenge may be considered and tackled in the curriculum renewal currently occurring in Finland. The reformulated National Core Curriculum for basic education will take effect at the beginning of the autumn term 2016. The National Core Curriculum will be updated and redesigned in a process that is transparent, participatory and democratic. Instead of being prepared solely by the Finnish National Board of Education and handed out to municipalities without any consultation, the new core curriculum will have undergone several public draft stages before being issued as the final document. Pupils, their parents, teachers, administrators, scholars, stakeholders and anyone interested in curriculum development will have had their say on a web page specifically established for the purpose. Furthermore, schools were approached in this matter and asked to collect, for instance, opinions about the skills and knowledge needed in the future. Special expert panels have been invited to brainstorm and sketch the needs of future basic education and core issues in specific curricular areas – also those in CLIL. The future in this respect thus seems to be brighter than before.

## 1.2. Second challenge: CLIL teachers’ language knowledge and proficiency

The second challenge to CLIL assessment is the lack of qualified CLIL teachers. In the Finnish primary education (classes 1-6), the class teacher is generally responsible for teaching most, if not all subjects to the same group of students, and subject teachers take over the teaching no later than from the 7<sup>th</sup> grade onwards. Finnish teachers are highly educated. The minimum requirement for teacher qualification in Finland, at any level, is a master’s degree including pedagogical studies and practical training, but double qualifications are rare. Class teachers and subject teachers normally do not possess language teacher’s qualifications unless they have acquired it simultaneously with or subsequent to the master’s degree. Language teachers rarely have a diploma in subject studies such as mathematics or science.

Fulfilling the official language requirements for CLIL teaching is an elaborate undertaking. The Ministry of Education (2005) issued a decree setting CLIL teacher language qualification standards which are fulfilled when the teacher has either a) Level Five in the National Certificate of Language Proficiency (NCLP, see [www.jyu.fi/hum/laitokset/solki/yki/english](http://www.jyu.fi/hum/laitokset/solki/yki/english)), b) at least 80 credits of university level studies in the language of CLIL instruction or c) acquired a teacher qualification in a country where the CLIL target language is an official language and has received acceptance of that degree from the Finnish Ministry of Education. The language requirements are quite high since Level Five in the NCLP is equivalent to Level C1 in the Common European Framework of References for Languages (see [www.coe.int/t/dg4/linguistic/Cadre1\\_en.asp](http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp)), and 80 credits of academic studies correspond to studies at the advanced level. According to the survey by Kangasvieri et al. (2012), it appears that the most important criterion for the recruitment of CLIL class teachers is their basic education, i.e. a master’s degree in education and class teacher qualification with its two secondary subjects. As a consequence, genuinely qualified CLIL teachers can rarely be recruited, and the language aspect becomes secondary in the CLIL classroom. The weight placed on content and language is a matter of judgment, but in CLIL, examined from any perspective, content comes first.

---

Taina Wewer (2013), 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. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-18-2/beitrag/Wewer.pdf>.

It follows therefore that when teachers do not have sufficient knowledge of the language learning processes in second language acquisition (SLA), good proficiency in the additional CLIL target language (TL) or any language studies at all, it is then even more challenging to combine content and language teaching. One could question whether such teachers can teach CLIL at all.

Hartiala (2000) has studied the acquisition of teaching expertise in CLIL; she distinguishes two dimensions in the process of becoming a CLIL teacher: earlier professional competence and new expertise of CLIL. She claims that CLIL teachers are perceived as experts from the beginning of their careers due to the special role of CLIL in the field of SLA, but the true basis of their expertise lies in the special skills that optimise the CLIL expertise. These skills are, according to her, related to the knowledge of and proficiency in the TL, the relationships between the TL and the learners' mother tongue, cultural factors, contents and materials, teaching methods and learner scaffolding as well as interpersonal and organisational relationships. Language skills are on top of her list, which is highly appropriate because without sound knowledge of the TL, the components of language proficiency and theories of SLA the documenting and profiling of CLIL learners' language attainment may be pseudo-language-teaching and quasi-assessment.

### 1.3. Third challenge: Lack of common assessment practices and tools

Advocacy for formative classroom assessment instead of or in addition to standardised, high-stakes tests seems to be a universal trend in the assessment literature. Assessment experts are demanding the rethinking of assessment and strive for assessment *for* learning (formative assessment) instead of summative assessment *of* learning (e.g., Birnbaum, Breuer, Cascallar, Dochy, Dori, Ridgeway, Wiesemes & Nickmans 2006; see also Inbar-Lourie 2008), because it takes the individual learners and the specific assessment contexts into better consideration and prevents teaching for assessment only (the washback effect, i.e. preparing students for tests only). This, however, is not an issue for countries not relying on standardised high-stakes tests such as Finland. The National Board of Education (2013) describes Finnish assessment policy as follows: "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". Formative assessment for learning thus is intrinsic to Finnish teachers. Nevertheless, it is still unclear which assessment methods are appropriate in assessing content mastery in the target language in CLIL contexts.

Assessing the CLIL learners' language proficiency in relation to their content knowledge is not an easy task because language has a dual function in CLIL, being both the target and medium of instruction. The National Core Curriculum of Basic Education (NCC 2004: 273) mandates that assessment in CLIL environments "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." As a result, there is no room for other interpretations: target language assessment must be carried out in the CLIL context, and the language outcomes need to be related to the predetermined goals. If assessment is not carried out at all, it is not in the spirit of the NCC. Assessment in CLIL can be carried out in a number of ways, but if there are no language goals or no proper CLIL curriculum available, it is very challenging for teachers to undertake the assessment task because it remains unclear what to assess. This raises several questions such as which means or methods are appropriate for assessment; how the gathered information should be conveyed to the parties involved and how often; how much information is sufficient; and what qualities that information should have in order to be adequate. In this respect, research is needed. The present study attempts to shed light on the current CLIL assessment practices and look at the frequency, sufficiency and adequacy of the practices.

When considering assessment practices in CLIL environments, the following notions need to be acknowledged. In the CLIL context, the TL is ubiquitous, "polymorphic"; it is manifested both in academic, discipline-related language and in casual, every-day language both in context-embedded as in context-reduced situations (see Cummins 1982). Since regular, conventional language study mostly represents conversational English, it is not appropriate to assess the academic CLIL English in the same manner and to give a combined grade for formal English and CLIL

---

Taina Wewer (2013), 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. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-18-2/beitrag/Wewer.pdf>.

English in the report. The two must be separated, the underlying criteria for CLIL language assessment must be made transparent, and assessment tools addressing functional language use in content-related situations must be developed. The search for answers to these issues also lies at the heart of the research reported in this article.

## 2. Brief research and literature review of assessment in CLIL

Although other aspects of CLIL (e.g., language development, effectiveness and affective factors) have been covered fairly extensively, research on the assessment of language proficiency in CLIL environments, especially at the primary level, is very scarce. In recent years, assessment aspects have gained more attention in the CLIL literature. For example, there is no separate chapter on assessment in the CLIL handbook by Mehisto, Marsh and Frigols (2008), although they touch on assessment through student reflection and portfolios, whereas in the CLIL publication by Coyle, Hood and Marsh (2010) a whole chapter is dedicated to assessment issues. Coyle, Hood and Marsh (2010: 112-113) discuss essential assessment principles such as objective setting, multifaceted assessment in real, meaningful language use contexts and engaging the learners in the assessment process. They suggest language portfolios, self and peer assessment as practical assessment tools (*ibid.*). Nonetheless, there are not many studies that have actually investigated the effects of various assessment methods in practice.

One such inquiry was conducted by Hönig (2010); her case study was carried out at the upper secondary level (students aged 15-18) in Austria on the subject of oral history exams. She interviewed four CLIL teachers, observed and videotaped 11 lessons of two teachers as well as gathered one teacher's observations about the students' transcribed oral history exams in the TL. Hönig noticed that although teachers firmly denied that the target language played a role in the grading of the mastery of history, it still appeared to have an effect on their assessment: an eloquent TL performance during the test resulted in better grading. In other words, linguistic accuracy and fluency conveyed the (sometimes false) impression of good content mastery, and language was unintentionally co-assessed with content. Based on insight, Hönig concluded that separating content and language in oral tests is very challenging, if not impossible. She called for other than oral assessment instruments which facilitate a clear distinction between content and language, for the assessment of both dimensions according to separate criteria and for a raised awareness of language competency among the teachers. At the same time it must be remembered that language is the medium and the objective of learning in CLIL, and they cannot be treated totally separately. Criterion-referenced assessment is one solution to the dilemma; content mastery and language proficiency – or rather language *use* – should have distinct criteria against which they will be measured separately although they co-occur in elicited performance samples.

Some theoretical and practical initiatives and attempts have been made to find pragmatic solutions to the CLIL assessment problems outlined above. For instance, Mohan, Leung and Slater (2010) have launched the concept of Integrated Assessment of Language and Content (IALC), which refers to simultaneous assessment of language and content in texts by using a systemic functional linguistic framework. They anchor the IALC in Halliday's view of language as a demonstration of learning: "If language is the primary evidence for learning, then assessment is primarily assessment of text or discourse, and how wording constructs meaning in text" (221). They give examples of IALC meaning-wording analyses both in written texts and in formative classroom discourses as in recasts (responses to and rephrasings of utterances). Text analyses might be, with reservations, suitable for upper primary classes, but in that case teachers should be competent in both the linguistics of the TL and the subject content. Employing such analyses requires that classroom interaction is actually conducted in the TL (and not in the principal language of classroom); the teacher should be very fluent in speaking the TL and be knowledgeable about linguistic forms. The IALC approach may be more applicable to upper levels of CLIL education where the learners exhibit higher language proficiency and more sophisticated discourse skills.

Mohan, Leung and Slater (*ibid.*) recommend formative classroom interaction assessment practises, as do also Llinares, Morton and Whittaker (2012) who argue strongly for formative, continuous classroom assessment in CLIL. Although admitting that assessment should primarily focus on content, they provide examples of criterion-based content/language scales to facilitate formative classroom assessment. More practice-oriented approaches to assessment in CLIL have been taken in various projects. The EU-funded, pan-European project Assessment and Evalua-

---

Taina Wewer (2013), 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. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-18-2/beitrag/Wewer.pdf>.

tion in CLIL (AECLIL 2013), and the Swiss-German research project CLIL Learner Assessment (CLILA 2013), have tackled the dilemma of discovering appropriate assessment tools. The AECLIL partners, consisting of a wide range of educational levels from primary schools to higher education and teacher training, present band sheet grids and rubrics as well as self-evaluation sheets for assessment of the learners' linguistic and disciplinary skills (see Quartapelle 2012) and adhere to the principles of criterion-referenced assessment. The rubrics are based on a conceptual framework minted within the project combining aspects of content, thinking skills and academic language (see Barbero 2012).

The research project CLILA, still on-going, aims at creating an assessment tool combining descriptors both in language (the Common European Framework of Reference for languages (CEFR 2001) and *Lingualevel*, a Swiss collection of assessment tools) and content (based on the curriculum and disciplinary skills). The project will be introduced in more detail in a forthcoming publication by Massler and Stotz (in prep.). These assessment initiatives meet the needs of assessment in CLIL as well as answer the call for formative, primarily learner-centred practices. However, this may still not be sufficient, for Birenbaum et al. (2006: 61) call for the modernisation of assessment also regarding the media:

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.

One of the forerunners of assessment modernisation will be the Finnish Matriculation Examination in which traditional paper-and-pencil tests will be gradually abandoned and replaced by electric, computer-mediated examination by 2019 (read more in <http://tuus.fi/news/article/ville.php>). In the CLIL environment, the technology project Profiling Learning Progression in CLIL Environments through Computer Simulations (PROFICOM 2013), funded by the Finnish National Board of Education, also includes information technology and modern educational technology in assessment by experimenting with film-based, narrative computer simulations with various content and language tasks embedded in a frame story. Within the project, three simulations are produced. They are specifically designed to profile primary CLIL pupils' content knowledge in the TL and to reveal components of the test takers' functional language competence in both receptive and productive language skills. PROFICOM simulations are based on the LangPerform concept innovated by Haataja (2010).

While simulations such as flight or medical operation simulators replicate real life situations very truthfully, LangPerform simulations can be described as semi-authentic and semi-interactive due to the fact that the language and content use situations are predetermined, written as a manuscript, and the characters do not reply to the performer's questions. However, simulations provide opportunities for the meaningful application of language skills and content knowledge within contexts that enable pupils to travel, encounter native and non-native English speakers and act in situations beyond traditional school contexts. The simulation environment contains 1) the simulation itself, 2) a language laboratory with an evaluation section for self-assessment, teacher assessment and parental monitoring drawing on the CEFR (2001) as well as 3) a language profile section resembling the European Language Portfolio where the simulation performer can portray his/her language biography (see, for example, Haataja 2010; Haataja & Wewer 2013). The student performance, i.e. given replies or speech, written texts and choices made (e.g., drag-and-drop tasks), is saved in the LangPerform central server from where it can be retrieved either immediately or even at a long period of time after the performance for closer examination and assessment. Assessment of the language and content knowledge sample can be performed by all parties involved as shown in Figure 1 because they all have access to the performance and assessment in the central server.

---

Taina Wewer (2013), 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. Abrufbar unter <http://zif.spz.tu-darmstadt.de/jg-18-2/beitrag/Wewer.pdf>.

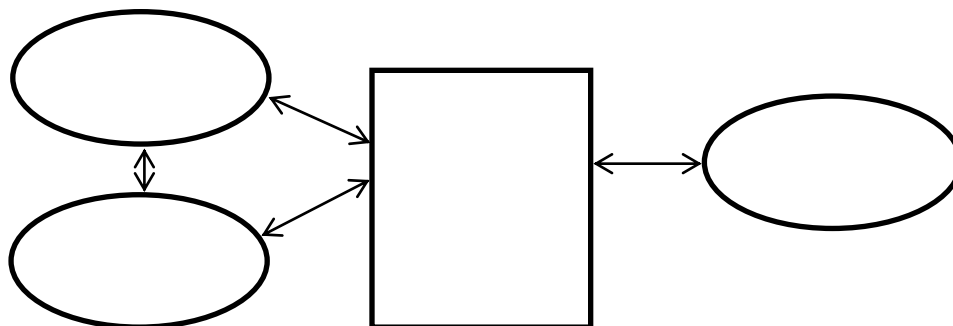


Figure 1: The assessment model applied in the project PROFICOM (based on Wewer 2013)

In addition to having the option to review the simulation again regardless of time and place, the pupils and parents can see the teacher's assessment and overall opinion of the performance; they can listen to and read their original performance, execute self-assessment and write comments or notes. These are not accessible to others, for each pupil has an individual user name and password. Furthermore, browsing the simulation environment and assessment may elicit fruitful discussions between the child, the parents and the teacher involving languages, culture and studying. The PROFICOM simulations comprise the empirical part of the CLIL assessment research introduced in the following section.

### 3. Assessment of Finnish CLIL pupils' English language proficiency in Finnish primary grades 1-6

The three challenges (diversity of CLIL provision, CLIL teachers' language knowledge and proficiency as well as the lack of common assessment practices and tools) described above and the gap in CLIL assessment research laid the foundations for this research. Since assessment issues in CLIL contexts have not been studied at all in Finland, and hardly elsewhere, this investigation of how CLIL class teachers in Finland perceive and carry out language assessment is important. Furthermore, the computer simulation as an innovative assessment method was chosen for closer scrutiny and experimentation. In this chapter, I will introduce the core elements of the study and some preliminary results. The tentative conclusions drawn and implications considered based on those results will be presented in the final chapter 4.

#### 3.1. Purpose of the study, research questions and design

The aim of this research is to establish an overview of the assessment of pupils' English language proficiency in Finnish CLIL classrooms and to gain an insight into current assessment practices, which is one of the principal purposes of assessment research (see McKay 2006: 65). The main focus is on formative assessment, although the analysis also takes an interest in practices related to summative assessment – the two assessment forms acknowledged in the National Core Curriculum for Basic Education (NCC 2004). An essential part of the study is to investigate the feasibility of LangPerform computer simulations for CLIL assessment purposes at the primary level – also a precursory undertaking and a major purpose of assessment research (McKay 2006: 65). Furthermore, the research is carried out in hopes of contributing to the establishment and expansion of a more solid theoretical base for assessment in CLIL and providing a practical orientation toward participatory, learner-centred formative assessment. The research questions are:

1. What kind of assessment methods do CLIL class teachers employ for the assessment of English language proficiency?
  - 1.1. To what extent do CLIL class teachers assess language?
  - 1.2. Do the pupils and their parents receive sufficient information on the pupils' level of English language proficiency and its development in CLIL contexts?
  - 1.3. Which methods and practices do the pupils and their parents prefer in receiving information on the pupils' English language proficiency and its development in CLIL instruction?

2. What kind of information do computer simulations yield on the pupils' English language proficiency, and are they feasible for CLIL assessment purposes?
3. What kind of thoughts and experiences are generated by the simulation experiments in the pupils and their parents?

### 3.2. Research phases and methods

The choice of research methods contributes to the validity of research results. Therefore both data and methodology triangulation are applied. Assessment is approached from the perspectives of all three parties involved: teachers, pupils and their parents. The data were collected using three methods: semi-structured questionnaires containing both fixed and open-ended response fields, audio-recorded theme interviews and video interviews related to the simulation experiments. The data collection occurred in three phases during the years 2012 and 2013: the initial survey was the first phase followed by simulation experiments 1 (second phase) and 2 (third phase).

The initial survey queried the assessment methods used and their adequacy, frequency and sufficiency as well as wishes for practices to be reconsidered and future visions. The questionnaire was sent to the headmasters of primary schools enlisted in the CLIL Network web site (see [clil-network.uta.fi/index.php?id=8](http://clil-network.uta.fi/index.php?id=8)) with a request to forward it to the CLIL class teachers, and handed to the CLIL class teachers of the two participating primary schools in the spring of 2012. Altogether 42 primary teachers returned the questionnaire, a substantial number of whom were first and second grade teachers. The 3<sup>rd</sup>, 4<sup>th</sup> and 5<sup>th</sup> graders (9-11-year-olds) of the two participating schools filled in the pupils' paper questionnaire (n= 109), a version of which was also sent to their parents. Altogether 99 parents returned the questionnaire (90.8% of the questionnaires sent out). Additionally, volunteers in each participant group (10 teachers, 20 pupils and 7 parents) were interviewed using elaborate theme interviews, the objective of which was to collect more profound and detailed information on the current assessment practices in Finland and in the research schools. The schools had quite different profiles in several variables, e.g., location, size, cultural and linguistic environment, ethnic background of pupils as well as in the training and experience of teachers.

The two simulations were produced and financed within the project PROFICOM (2013). The first simulation entails content knowledge mainly from mathematics and the environmental sciences, while the second included primarily science, arts and physical education. The simulations were piloted in the two participating research schools in autumn 2012 and the following winter 2013. The 5<sup>th</sup> and 6<sup>th</sup> graders (11-12-year-olds, n= 74) piloted the first simulation, the 4<sup>th</sup> and 5<sup>th</sup> graders (n= 72) the second. Since the piloting occurred within one school year, the 5<sup>th</sup> graders participated in two different simulations thus providing a deepened understanding of how the simulation experience evolves and triggers pupils' perceptions of themselves as TL users.

The simulations were conducted by means of a computer and a headset; each pupil was given a user name and a password to sign in which were also sent home to provide the parents the opportunity to familiarise themselves with the simulation environment and to listen and see the saved responses. The first simulation was situated in Michigan, U.S., under the pretence of attending a school visit during which the testee was accommodated in a local family, while the second setting was a visit in an international school (see, e.g., Haataja & Wewer 2013). Both settings thus were within the pupils' experience of the world and therefore less likely to create anxiety. Furthermore, the pupils and parents were informed that the underlying purpose of the simulation was to pilot the assessment method new to CLIL and the primary settings, that the actual performances would not be measured and that they would not affect grading. The pupils were to react to various communicative and content-knowledge-measuring situations activating the four basic language skills.

All piloting sessions were concluded with research questionnaires which each participating pupil filled in. They were not completely identical in the two phases. The questionnaires related to the simulations were basically subjective self-assessments as well as expressions of opinions and impressions aiming at finding out how the simulations as an assessment method in CLIL contexts at the primary level are received, perceived and accepted; the analysis of the actual simulation outcomes is a matter of further research. Furthermore, volunteers gave an interview which was videotaped in a so-called diary room. The entire simulation sessions in the other school were videotaped in the sec-



ond phase. Instructions on how the recorded and saved performances can be retrieved were sent to the pupils' homes alongside with questionnaires for the parents. The number of returned parents' questionnaires was very low ( $n=26$  in the first round,  $n=13$  in the second) compared to the parents' responses in the initial survey. This is most likely due to the technical difficulties experienced in both piloting phases, some of which caused the loss of performances in the second piloting phase. As a result, there was nothing for parents to monitor. Another factor contributing to the low response rate may have been the reluctance that some pupils indicated to let their parents know how they managed in the simulation. Such pupils may never have handed on the information and questionnaire envelope.

### 3.3. Sample of preliminary results

The results presented here represent only a part of the total findings. Preliminary results from the initial survey (questionnaires and interviews) revealed that the most common methods of assessment are observation, bilingual tests (some sections or tasks in the foreign language) and dialogic interaction with pupils; the least used methods were peer assessment, simulations of any kind and portfolios. Assessment in CLIL was perceived as challenging, agonising even, due to various reasons such as absence of the CLIL curriculum and shared practices. Of the CLIL class teachers 21% reported that they did not gather systematic assessment information. The information and feedback which CLIL teachers give to the pupils and their parents seem to be occasional, informal and oral rather than systematic and evidence-based. However, 48% of teachers perceived the assessment of the learners' language skills as either highly important or very important. 49% of the pupils emphasized the importance of receiving feedback on coping in English in CLIL subjects; the younger they were, the more highly they valued it – girls even more so than boys. A strikingly low proportion of pupils, only 8%, reported receiving sufficient feedback. Accordingly, 62% of the pupils expressed a wish to receive more feedback than they did at the time of the study. Parents were in agreement with their children: 76% of parents would have liked to know more about how their children cope in the CLIL context in English and how their language skills develop.

At this point, only the complete analysis of the first simulation piloting is available. As to the feasibility of computer simulations, the results are encouraging in terms of introducing a new, semi-interactive and technology-based assessment method. The majority, 80% of the pupils and over 90% of the parents responding stated that computer simulations are feasible for CLIL assessment, although a number of suggestions for further development were also received. The participants' self-assessments showed that no fewer than 45% of the pupils were content with their performance, boys slightly more so than girls, and yet 46% thought they coped moderately well in English. Some pupils reported that their language self-concepts were reinforced because of the simulation, while others noticed areas for further improvement in their language skills.

In general, pupils perceived "warm-up" tasks requiring conversational, everyday English easier than tasks pertaining to content. The task estimated to be the easiest by 59% of pupils was the one where they were expected to introduce themselves and to talk about themselves. The most difficult task from the viewpoint of 25% of pupils was the explanation of a calculation (subtraction in columns). Nonetheless, one class in particular regarded the subtraction task as being less difficult than other classes on average. An interesting feature was that every single task received a mention of being either the easiest or the most difficult. Pupils appreciated especially the opportunity to hear one's speech and to listen to the pronunciation.

## 4. Conclusions and implications

It is too early to draw specific inferences based on the data, but general tendencies salient in the data enable us to make some general-level conclusions. The conclusions concerning CLIL teachers are based on the query responses received from all over Finland, whereas the information obtained from pupils and their parents refers only to the two participant schools. The results as such are not widely generalisable, but it is fairly likely that they, at least to some extent, apply to other CLIL contexts as well. First, there are strong indications that assessment and feedback in CLIL needs to be reorganised. The basic prerequisite for CLIL assessment to succeed is the CLIL curriculum which defines the contents to be taught in the additional language and describes the desired level of language proficiency. Second, pupils and parents wish to be informed of the progress in the additional language in reference to the learn-

ing objectives. This implies that CLIL teachers should arrange more functional language use situations for pupils in which they can exhibit their language skills, and teachers should practice more systematic observation and data gathering of the progress made in language development. Third, it is very important for pupils to get constructive and direct feedback on their emerging (academic) learner language in order to encourage them to use the TL. Informal CLIL school year reports are also worth considering as one option to convey information to parents.

The LangPerform simulation also appears to be a valuable potential assessment method for a number of reasons. First, the medium of a computer seems to help the pupils to put their language skills to proper use. Second, simulations provide opportunities to step out of the traditional classroom context, travel across the ocean to another continent within minutes and encounter (native) speakers with various Englishes and accents. Third, the circumstances where content knowledge is used are embedded in the film-based stories in a credible, authentic way which makes the demonstration of that knowledge more meaningful than filling in book exercises in the classroom. On the one hand, simulations in their communicativeness are a breath of fresh air in the assessment tool kit, but on the other hand, they are not yet widely available, they are time-consuming and costly to produce and there still are technical improvements to be made. When this is achieved, computer simulations should be available in every language classroom. I see such computer simulations as essential elements in providing concrete evidence of progress on language skills when repeated, for example, twice a year. Computer simulations, when designed to include the core contents, could also serve as “final exams” at the end of each school year. Computer simulations provide a technique to update assessment to the standards of the Digital Era we are living in as required by Birenbaum et al. (2006); they activate pupils in ways that are normally not possible in classroom contexts; and they represent a form of participatory, informative assessment that is modern and fun. To conclude in a pupil’s words: the PROFICOM simulation was “thrilling but at the end totally WONDERFUL!”

## References

- AECLIL (2013), *Assessment and Evaluation in CLIL* [Available online at [www.aeclil.net/](http://www.aeclil.net/). April 18<sup>th</sup>, 2013].
- Barbero, Teresina (2012), Assessment tools and practices in CLIL. In: Quartapelle, Franca (ed.), *Assessment and Evaluation in CLIL*, 38-56 [Available online at [riviste.unimi.it/index.php/promoitals/article/viewFile/2827/3030](http://riviste.unimi.it/index.php/promoitals/article/viewFile/2827/3030). April 20<sup>th</sup>, 2013].
- Birenbaum M.; Breuer K.; Cascallar E.; Dochy F.; Dori Y.; Ridgeway J.; Wiesemes R. & Nickmans G. (2006), A learning integrated assessment system. *Educational Research Review* 1, 61-67.
- CEFR (2001), *Common European Framework of Reference for Languages: Learning, teaching, assessment*, 3rd ed. Cambridge: Cambridge University Press.
- CLILA (2013), *CLIL Learner Assessment* [Available online at [www.ph-weingarten.de/clila-forschungsprojekt/en/](http://www.ph-weingarten.de/clila-forschungsprojekt/en/). April 18<sup>th</sup>, 2013].
- Coyle, Do.; Hood, Philip & Marsh, David (2010), *CLIL – Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Cummins, Jim (1982), Tests, achievement, and bilingual students. *FOCUS* 9, 2-9.
- Dalton-Puffer, Christiane (2007), *Discourse in Content and Language Integrated (CLIL) Classrooms*. Amsterdam: John Benjamins.
- European Commission (1995), *White Paper on Education and Training. Teaching and Learning: Towards Learning Society* [Available online at [europa.eu/documents/comm/white\\_papers/pdf/com95\\_590\\_en.pdf](http://europa.eu/documents/comm/white_papers/pdf/com95_590_en.pdf). March 16<sup>th</sup>, 2013].
- Graddol, David (2006), *English Next. Why global English may mean the end of ‘English as a Foreign Language’*. British Council [Available online at [britishcouncil.org/learning-research-english-next.pdf](http://britishcouncil.org/learning-research-english-next.pdf). April 21<sup>st</sup>, 2013].

- Haataja, Kim (2010), Das Konzept LangPerform. Entwicklung und Einsatz von Simulationsinstrumenten zur computermedialen Dokumentierung von (fremd-)sprachlichen Kompetenzen – innovativ und integrativ. *Jahrbuch Deutsch as Fremdsprache* 36, 183–199.
- Haataja, Kim & Wewer, Taina (2013), Elokuvapohjaisia tietokonesimulaatioita kokonaisvaltaisen kieli- ja kulttuurikasvatuksen tueksi – mutta miksi? In: Pirkkalainen, Lauri & Lounaskorpi, Petri (eds.), *Löytöretkillä toisessa maailmassa Vol 1*. Konneveden lukion verkkojulkaisu [Available online at [konnevedenlukio.onedu.fi/verkkojulkaisut/zine/33/article-365](http://konnevedenlukio.onedu.fi/verkkojulkaisut/zine/33/article-365). May 31<sup>st</sup>, 2013].
- Hartiala, Aini-Kristiina (2000), *Acquisition of Teaching Expertise in Content and Language Integrated Learning*. Turku: Turun yliopisto.
- Hönig, Ingrid (2010), *Assessment in CLIL: Theoretical and Empirical Research*. Saarbrücken: VDM.
- Inbar-Lourie, Ofra (2008), Constructing a language assessment knowledge base: a focus on language assessment course. *Language Testing* 25: 3, 385-402.
- Järvinen, Heini-Marja (2005), Language learning in content-based instruction. In: Housen, Alex & Pierrard, Michel (eds.), *Investigations in Instructed Second Language Acquisition*. Berlin: Mouton de Gruyter, 432-456
- Kangasvieri, Teija; Miettinen, Elisa; Palviainen, Hannele; Saarinen, Taina & Ala-Vähälä, Timo (2012), *Selvitys kotimaisten kielten kielikylpyopetuksen ja vieraskielisen opetuksen tilanteesta Suomessa. Kuntatason tarkastelu* [Available online at [ju.fi/hum/laitokset/solki/tutkimus/julkaisut/pdf-julkaisut/selvitys-kotimaisten-kielten-kielikylpyopetuksen-ja-vieraskielisen-opetuksen-tilanteesta-suomessa](http://ju.fi/hum/laitokset/solki/tutkimus/julkaisut/pdf-julkaisut/selvitys-kotimaisten-kielten-kielikylpyopetuksen-ja-vieraskielisen-opetuksen-tilanteesta-suomessa). March 16<sup>th</sup>, 2013].
- Lehti, Lotta; Järvinen, Heini-Marja & Suomela-Salmi, Eija (2006), Kartoitus vieraskielisen opetuksen tarjonnasta peruskouluissa ja lukioissa. In: Pietilä, Päivi; Lintunen, Pekka & Järvinen, Heini-Marja (eds.), *Kielenoppija tänään – Language Learners of Today*. AFinLan vuosikirja 2006. Jyväskylä: Suomen Soveltavan Kielitieteen Yhdistyksen AFinLan julkaisuja 64, 293-313.
- Llinares, Ana; Morton, Tom & Whittaker, Rachel (2012), *The Roles of Language in CLIL*. Cambridge: Cambridge University Press.
- Marsh, David; Järvinen, Heini-Marja & Haataja, Kim (2007), Perspectives from Finland. In: Maljers, Anne; Marsh, David & Wolff, Dieter (eds.), *Windows on CLIL – Content and Language Integrated Learning in the European Spotlight*. Alkmaar: Europees Platform, 63-83.
- Massler, Ute & Stotz, Daniel (in prep.), *Handbuch zur Leistungserhebung und Leistungsbewertung im fremd-sprachlichen Sachfachunterricht (CLIL) der Primarstufe*. Trier: Wissenschaftlicher Verlag Trier.
- McKay, Penny (2006), *Assessing Young Language Learners*. Cambridge: Cambridge University Press.
- Mehisto, Peeter; Marsh, David & Frigols, María Jesús (2008), *Uncovering CLIL. Content and Language Integrated Learning in Bilingual and Multilingual Education*. Oxford: Macmillan Education.
- Ministry of Education (2005), *Opettajalta edellytettävä kielen hallinta muulla kuin koulun opetuskielillä annettavassa esi- ja perusopetuksessa* [Available online at [www.finlex.fi/data/normit/26301-oph250112005su.pdf](http://www.finlex.fi/data/normit/26301-oph250112005su.pdf). April 19<sup>th</sup>, 2013].
- Mohan, Bernard; Leung, Constant & Slater, Tammy (2010), Assessing language and content: a functional perspective. In: Paran, Amos & Sercu, Lies (eds.), *Testing the Untestable in Language Education*. Bristol: Multilingual Matters, 217-240.
- National Board of Education (2013), *Education* [Available online at [oph.fi/english/education](http://oph.fi/english/education). April 19<sup>th</sup>, 2013].
- NCC (2004), *National Core Curriculum for Basic Education* [Available online at [oph.fi/english/publications/2009/national\\_core\\_curricula\\_for\\_basic\\_education](http://oph.fi/english/publications/2009/national_core_curricula_for_basic_education). March 16<sup>th</sup>, 2013].
- Nikula, Tarja & Marsh, David (1996), *Kartoitus vieraskielisen opetuksen tarjonnasta peruskouluissa ja lukioissa*. Helsinki: Opetushallitus.

- Pihko, Marja-Kaisa (2010), *Vieras kieli kouluopiskelun välineenä. Oppilaiden kokemuksista viitteitä CLIL-opetuksen kehittämiseen*. Tutkimuksia 88. Jyväskylä: Jyväskylän yliopisto, Opettajankoulutuslaitos.
- PROFICOM (2013), *Profiling Learning Progression in CLIL Environments through Computer Simulations* [Available online at [rule.uta.fi/proficom](http://rule.uta.fi/proficom). April 19<sup>th</sup>, 2013].
- Quartapelle, Franca (ed.) (2012), *Assessment and Evaluation in CLIL* [Available online at [riviste.unimi.it/index.php/promoitals/article/viewFile/2827/3030](http://riviste.unimi.it/index.php/promoitals/article/viewFile/2827/3030). April 20<sup>th</sup>, 2013].
- Wewer, Taina (2013), Minuutissa Michiganiin: tietokonesimulaatiot englannin kielen arviointimenetelmänä alakoulun vieraskielisessä CLIL-opetuksessa. In: Juuso, Hannu et al. (eds.), *Tutki, kokeile ja kehitä*. Suomen harjoittelukoulujen julkaisu. Oulu: Oulun yliopisto.