

Test Validity Concerns in CLIL: Lessons from the Netherlands

Alaa Al-Musalli, Ph.D.¹

Abstract

This paper sheds light on a controversial issue in CLIL, namely, testing the content-specific knowledge gained through CLIL programmes in a language other than the language of instruction (or the target language). The fact that this example of testing exists might sound preposterous for many educators around the world. The context within which the discussion of this example of testing is set is the Netherlands, for such testing is a matter of fact for many Dutch students. The sole benefit of testing content in Dutch after receiving instruction in the target language, i.e. English, is to unify exit tests across schools/colleges in the country. However, while the risks of not using a standard school/college test are unquestionable, the negative implications of testing content through Dutch rather than the target language on test validity seem to outweigh such risks. This paper reports on teachers' perspectives regarding the current approach of testing content-specific subjects in the CLIL programme in the Netherlands; it also calls for an examination of the validity of these tests.

Key Words: CLIL, CLIL Testing, CLIL in the Netherlands, Exit Tests and CLIL, Test Validity in CLIL, and Validity Concerns.

I. Introduction

The benefits of CLIL on developing language competence have been explored by many. However, most of the available research in the field uses tests that measure language competence rather than content-specific knowledge in the target language, which is the core focus of CLIL theory. For example, Várkuti (2010) found that using English as the language of instruction for content subjects provides a more efficient method for achieving functional language proficiency than the traditional foreign language learning tools. She compares the English language achievement of CLIL secondary school students with non-CLIL intensive foreign language learners in Hungary through testing their conversational and academic language skills. CLIL students in her study manifested better lexical, grammatical, textual, and sociolinguistic knowledge skills in conversational situations (see also Escobar and Sánchez, 2009). The fact that these findings focus on conversational academic language skills, rather than content-specific knowledge achievement in the target language, shows that the focus of the CLIL programme has been reduced to a language-focus proficiency test. Therefore, instead of testing content-specific knowledge in the target language, which is what CLIL helps achieve, improvement in language skills is tested independently of the content through which it is improved. This is, evidently, one good result, but can we ignore the other facet of CLIL, i.e. content?

The CLIL programme in the Netherlands is an example that manifests an incongruity between content and language goals. On the one hand, CLIL is pictured as a successful language programme that helps develop target language skills through the study of content in the target language. On the other hand, final tests in these content-specific subjects are administered in Dutch, not the target language. To understand why this has been allowed, we need to look at the diverse secondary education programmes in the country.

¹ McGill University, Canada.

The four main types of secondary education in the country, namely, junior general secondary education (MAVO), senior general secondary education (HAVO), pre-university education (VWO), and prevocational education (VBO) provide different courses for students above the age of 12. Students select the courses they wish to study depending on the kind of education they choose, and after completing their education, students prepare for exit examinations, which include both internal and national examinations; the former is given by the schools, while the latter is administered under government supervision (Education Encyclopaedia, 2015) (see also Expatica, 2014). Nevertheless, not all schools in the Netherlands provide bilingual education, known in Dutch as *tweetalig onderwijs* (or TTO), and those which choose to follow the programme decide for themselves 'how to implement the required teaching time in English' (European Platform, 2013, p. 2-3). This clearly means that students who sit for their national exit tests in the Netherlands are the product of programmes with varying degrees of target language exposure as compared with content instruction exposure in the national language.

The diversity of the level of exposure to the target language suggests that students who come from different programmes are treated equally and are expected to perform relatively the same on national exit tests. This discrepancy in the amount of target language exposure between TTO schools themselves as compared with non-TTO schools is undoubtedly one of the main shortcomings of this programme. Target language exposure in TTO programmes has been roughly estimated. One of these estimations maintains that 'at least half the teaching time' in HAVO and VWO classes is in English (European Platform, 2013, p. 6). Also, most of these schools provide extra English classes, bringing the amount of exposure to English that TTO students get to six times higher than those studying through the monolingual, non-TTO programme.

From day one, the main instruction language in the TTO subjects is English and students are expected to speak English whenever they can. This can be quite daunting for many students, especially those who were taught the minimum of English at primary school, but experience has shown that the great majority rise to the challenge. (ibid.).

However, the first problem evident from the above quote is that there is an assumption that on the day of the test, students must forget that instruction was done in English from day one and are capable of reproducing the content in Dutch because the test is administered only in Dutch. The second concern is clearly the fact that the programme is not only challenging but also unquestionably successful. The third problem that we see is in the mere assumption that students instructed through different levels of exposure to the target language in the CLIL programme and those who are not taking part in the bilingual programme are compatible and can sit the same content-specific exit test in Dutch. There is a big divide between the CLIL and Non-CLIL groups on the one hand and the amount of target language exposures among the CLIL group itself on the other. This leads us to believe that content-specific exit tests in the Netherlands are more suitable for students in Non-CLIL programmes as they have not had to study content in a language other than the language of the test. Also, this test is not a valid measure of students' content knowledge in CLIL programmes as the test is not administered in the language of instruction.

Gracia (2009) is one of the first to raise concerns regarding bilingual tests in the Netherlands, arguing that the inconsistency between the language of instruction and that of assessment is a major impediment to the development of bilingual education in that country. She maintains that authorities that insist on using monolingual, national tests for bilingual programmes are damaging the fundamental principle of such programmes. Gracia finds fault in the Dutch authorities' support of presenting Dutch in the final exams of foreign language content-matter instruction, comparing their model to two more successful attempts in both Britain and Germany, where tests have recognized the dual aspect of the programme, i.e. content plus two languages (the national and the target). Unfortunately, Garcia's concerns have not encouraged any changes to the way testing is done in Dutch bilingual education; hence, the questions remain; how valid are their monolingual, national tests and what effects do such tests have on students' learning experiences?

Marsh (2002) stresses the need for content specific tests that accommodate language competence, arguing,

There is, as yet, no widely applicable test of subject-specific target language competence available. There do exist, often in higher education institutions, numerous examples of language tests which are subject-specific but these are not generally widely available. The developments of such tests, which take into account both general language and subject-specific proficiency, could be made by those teacher education institutions which offer CLIL/EMILE through initial teacher education programmes. (p. 198).

This is still the case for most national tests across Europe. Each country in the European Commission has its own set of national subject-specific tests which have various degrees of target language components; some tests the target language through these content-specific tests, while others depend solely on the national language. Marsh (ibid: 201) stresses, 'Testing and assessment apparatus need to be introduced which allow learners to show the breadth of their knowledge and skills in relation to both content and language'.

It is safe to argue that when the target language is used to teach content-specific subjects, such as geography, physics, maths, or chemistry, tests for such content in the mother tongue undoubtedly yield questionable results. An accurate measure of students' knowledge in a given discipline is hard to test, so adding the target language element to the equation should not be treated lightly. In the context of exit tests in the Netherlands, students go into the CLIL programme with the unnerving knowledge of the fact that Dutch rather than the language of instruction, in their care English, will be used in testing their knowledge of the content. While there is no question that their language improves, the effects of testing in a language other than that of the language of instruction on their tests scores of the content-specific subject cannot be ignored. There is a clear shortcoming in such a module, and we will discuss below what teachers in this programme feel about it.

II. Students' Experiences and Test Validity Concerns

The advantages of CLIL cannot be denied. When it was made a priority in the Action Plan for Language Learning and Linguistic Diversity 2004-2006, teachers began receiving special training in CLIL. Some of the most important advantages of providing more contact with the target language through this programme are: improving oral communication and language competence, developing multilingual attitudes and interests, developing new classroom methods, increasing students' motivation and confidence, and, most importantly, complementing content subjects without the need for extra teaching hours (European Commission, 2011). Hence, CLIL provides students with more than just extra exposure to the target language; it trains them to process, i.e. analyze and synthesize content as a target language user, which is an ideal context to enhance study skills. The main problem we have with the Dutch interpretation of CLIL, though, is that students at different schools are expected to study and analyze some or whole parts of the new content in the target language but provide their understanding of this content in the national language through a unified exit test that treats all students equally regardless of their different instructional backgrounds. This approach triggers two main concerns: (a) content processing is in the target language, i.e. English, and (b) content testing is in the national language, i.e. Dutch. Below is a deeper look into these two concerns.

Language Exposure and Comprehension

The first concern we have with the current CLIL system in the Netherlands is that students are constantly processing and comprehending content in the target language in class, knowing that it will be tested in the national language, which adds unnecessary pressure and undermines the purpose of studying in the target language.

In most CLIL classes, students are expected to deal with and document the content in a way that would help them study for and pass their tests. Whether they take notes in their mother tongue or in the target language, students need to make sure they write down the essentials and the points their teachers stress. There are, unfortunately, no attempts to study what these students are writing down in class and whether or not they are taking notes in their mother tongue or the target language.

The act of taking notes facilitates learning since note taking aids comprehension by helping students pay attention and identify the core ideas, the basic organization and purpose of the material. It involves great effort as it requires students to make sense of the words and formulate the ideas in a way that would make sense to them later (Heaton, 1975, p. 108; Chambers and Northedge, 1997, p. 60; Lewis and Reinders, 2003, pp. 75-76). Therefore, students are processing information constantly, but they also know that they need to separate the language from the content at one point to prepare for the final test. It is safe to say that with this in mind, we can agree that testing content in a language other than that used for instruction and through which the students processed the content is very detrimental to the students' learning experience. The problem here is not simply with the different terminology associated with teaching content. CLIL promises to develop students' competence in many areas, yet to develop these set of skills in the target language and test content in the national language is not serving the purpose of the programme.

Wolff (2007) argues that while terminology was a focal point in the original CLIL theory, in modern CLIL the target language and terminology are not the main focus of the lessons; rather, the focus now is on equipping students with a set of speech acts, such as describing, explaining, evaluating, and drawing conclusions, which would help them function independently. Wolff is not exaggerating the effect of correct CLIL practices; however, one of the problems with the Dutch interpretation of this programme is the absence of the means to show language progress in the final exit content tests. This is a less devastating problem than the effect of testing content in a language other than the language of instruction on students' final scores in these content-specific subjects, but it is a problem that deserves attention, nonetheless.

Even in the most ideal contexts where CLIL tests are administered in the target language, there should be a clear divide between language and comprehension considerations depending on students' levels. Mehisto, Marsh, and Frigols (2008) stress the importance of focusing on testing comprehension rather than language skills in CLIL tests arguing that spelling and grammar should not affect the final mark of content classes in the target language as long as the information provided is understood and the learners continue to develop. This way, achievement of content as well as language goals can be tested, which reflects two primary objectives of CLIL courses. The Dutch system is not considering this factor in their testing, nor is it concerned about the challenges that students studying through CLIL are facing.

Test Validity

The second concern we have with the current testing system is that since the language of instruction used to process the content in class is different from the language of testing used to measure how much of the content the students have retained, test validity is undermined. While our first concern, i.e. language exposure and comprehension, is mostly theoretical and can largely be examined qualitatively, test validity can be easily measured and assessed. Test validity can show us the weight or worth of the test result that the current exit Dutch tests have.

Validity can be tested logically though measuring its face value (face validity) and its reflective value (construct and content validity). The former can be assessed by test experts and test-takers experiences, while the latter can be assessed by subject matter experts. In addition, validity can also be tested empirically though statistical correlations between the test results and other benchmarks (concurrent validity).

One could argue that a chemistry exit tests presented in Dutch of content taught in English can still measure what it claims to measure and is, therefore, valid; however, test validity is not only tested on face value. In the context of the CLIL programme, by arguing that a chemistry test in Dutch is deemed valid simply because it is testing knowledge in this field, we are ignoring the fact that the communication of the knowledge in question was done English. There is a clear problem here with both construct and content validity since subject matter is presented in a language other than the language with which it is tested.

Construct validity relates to the question of whether the test matches or relates to the theory behind it. Since tests are part of the programme they are testing, they should share the same philosophy and assumptions as the programme they are part of. If a programme aims to develop competence in a particular area, the test procedure should follow the same approach. On the other hand, content validity relates to the question of whether the kind of language generated in a test can reflect the syllabus (Underhill, 1987, p. 106). Therefore, we can suggest that testing through Dutch poses a problem in construct as well as content validity as it ignores English competence which is one of the fundamental objectives of the CLIL programme through which competence in the subject matter of the test is developed. The question here remains; can a study test the concurrent validity of content tests in Dutch and prove that they do indeed test the students' competence in subject matter without losing any values of acquiring such competence due to the interference of the language of instruction, i.e. English? In other words, are the experience and tests results of students who study Chemistry in English through the CLIL programme, for example, similar and as valid as those of students who study Chemistry in Dutch and do the current exit tests in the Netherlands reflect that?

III. Methodology

Students studying through the current bilingual stream in the Netherlands are experiencing more challenges than those studying through a monolingual stream with the foreign language taught as a separate subject. Some of these students are graduating to become teachers in the same bilingual programme they studied through; this, undoubtedly, helps them relate to the challenges of studying content-subjects in a foreign language while knowing they would be tested in their national language all throughout the learning process.

These teachers also have their own perceptions regarding how valid they think such tests are. Therefore, understanding their perspectives could shed light on what needs to be implemented to improve students' learning experiences.

To understand teachers' viewpoints, a number of points were raised in two workshops on testing through CLIL presented to two groups of college teachers in the Netherlands. The workshops were presented by the author in 2012. The data for this study has been collected through questionnaires followed by a short discussion. In order to guarantee the participants' understanding of the concept of test validity, three main types of test validity, face, content, and concurrent validity, were provided on a slide. The two groups of participants, a total of 15 teachers, were asked to tick one of five boxes, Strongly Agree, Agree, Not Sure, Disagree, or Strongly Disagree, underneath each of the statements below to indicate their views regarding the use of Dutch in exit tests administered after CLIL courses. They were also encouraged to comment on the bottom of the questionnaire page or present their concerns orally after filling in the questionnaire. The questionnaire statements are:

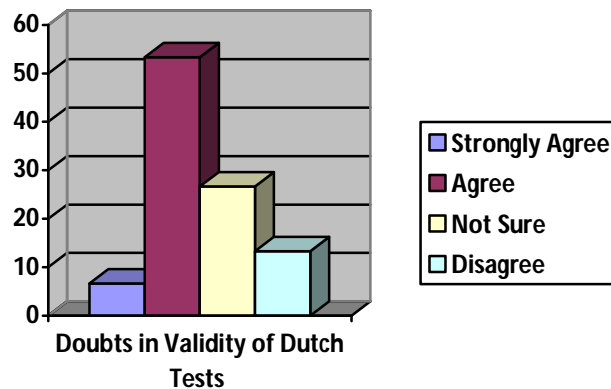
1. I have doubts in the validity of using Dutch in final tests for bilingual classes.
2. I believe that pupils in bilingual education find moving from the language of instruction to the language of testing easy.
3. Final tests for bilingual education in the Netherlands should use the language of instruction to test accurate knowledge of the content classes without focusing on language mistakes.
4. Using the language of instruction without focusing on language mistakes in bilingual tests is less stressful for both pupils and teachers than testing through Dutch.

IV. Findings and Discussion

The participants' responses to the above questionnaire statements are as follows:

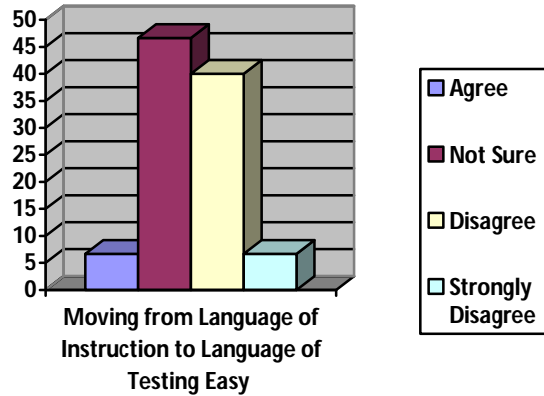
60% of the participants reported that they have doubts in the validity of using Dutch in final (exit) tests in their CLIL courses, 53.33% of whom agreed with the statement, while 6.66% strongly agreed. On the other hand, 26.66% of the participants stated that they are not sure regarding their standpoint, while 13.33% disagreed, arguing that they have no doubts in the validity of the exit tests in Dutch (see also Figure 1 below).

Figure 1 – Doubts in the Validity of Dutch Exit Tests

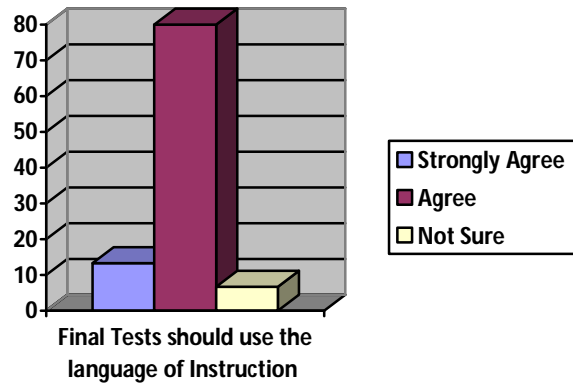


As for the second statement, 46.66% of the participants believe that pupils in bilingual education do not find moving from the language of instruction to the language of testing easy. 40% of these participants disagree with the statement given in the questionnaire as compared with 6.66% who strongly disagree with it. On the other hand, another 46.66% of participants reported that are not sure about the students' experiences, while 6.66% agree that indeed students do find it easy to move from one language to the other (see also Figure 2 below).

Figure 2 – Moving from the Language of Instruction to the Language of Testing is Easy

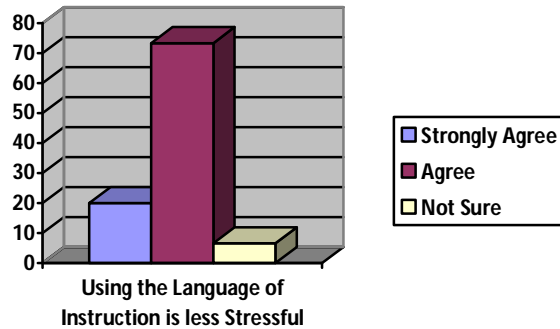


The third statement revealed a general agreement regarding the need for the use of the language of instruction in exit tests for bilingual education in the Netherlands to test accurate knowledge of content classes without focusing on language mistakes. 13.33% of the participants strongly agreed with this statement, and 80% agreed. The remaining 6.66%, which accounts for one participant, expressed uncertainty in regards to the answer (see also Figure 3 below).
 Figure 3 – Final (Exit) Tests should use the Language of Instruction



Ratings for the fourth statement also supported tests in the language of instruction as 20% of the participants stated that they strongly agree in addition to another 73.33% of the participants who stated they agree that using the language of instruction without focusing on language mistakes in bilingual tests is less stressful for both pupils and teachers than testing through Dutch; on the other hand, only 6.66% of the participants reported that they are uncertain about their viewpoint regarding this.

Figure 4 – Using the Language of Instruction is less Stressful



The fact that there was an agreement on some of responses, even though it is a very small sample, only 15 participants, indicates that teachers have more questions than answers. There is a clear need to educate teachers about the difference between bilingual education and pure CLIL language programmes and the methods by which each can be tested. The high percentage of 'Not Sure' comments given for the first two statements in the questionnaire shows a clear lack of understanding of CLIL and bilingual education and methodology behind them. Kaplan and Baldauf (1997) explain that teachers may not be comfortable with a new methodology when they do not understand the theory behind it or when it contradicts or differs from the ways in which they learned the target language (p. 134). We can only assume that this is why there is uncertainty regarding some of the comments given in the questionnaire, which emphasizes the need for teachers and colleges to develop a better understanding of their programmes.

The participants provided very few written comments on the questionnaire paper; the three comments received are:

1. "In tests, I don't think pupils should be punished for mistakes. I don't do that in my non-CLIL classes either. Test is on content, in class it's both content and language". By "Test" the participant is referring to national exit tests.
2. The students' experience "might differ" depending on subject content.
3. "Just test in English!"

The reflection discussion that followed the questionnaire indicated that teachers needed support on many levels. In summary, the group generally agreed that:

- a. They are unhappy with the fact that they are forced to teach some content areas in English, especially when students' proficiency in English is weak.
- b. They are not fully aware of when they should use Dutch versus English in class, but they tend to rely on Dutch when students' comprehension of the concepts in English breaks down. This reminds us of the rough estimate given by the European Platform (discussed above) regarding the amount of exposure to the two languages in such programmes.
- c. They are uncertain about the validity of national exit tests and would like to learn more about the effects of not using the language of instruction on their pupils' learning experience and marks, keeping in mind that they generally use both in varying degrees in class to ensure comprehension. This is a very important argument as it supports the need for a re-examination of the validity of exit tests administered in Dutch for CLIL programmes, which is the main focus of this paper.
- d. They are uncertain as to whether or not pupils find moving from English to Dutch for the exit tests easy, but a couple of teachers who have studied through the same bilingual programme that they are teaching have indicated that it took them roughly two weeks to translate concepts from English to Dutch to prepare for the national tests; they maintained that concept terms were hard to master as some are completely different in the two languages. This dependence on terminology indicates that the form of CLIL used in the Netherlands is what Wolff (2007) calls the *original CLIL* as opposed to the *modern CLIL* (discussed above).
- e. They are uncertain as to whether final tests can be tested in English since some teachers rely more on Dutch than others. They, however, agree that if English is used, language mistakes should not be penalized. They agree that using English without focusing on language mistakes would be less stressful for both pupils and teachers. This agrees with Mehisto, Marsh, and Frigols' (2008) argument that comprehension, rather than language mistakes, should be the focus of such tests (discussed above).

V. Conclusion and Recommendation

The reason this study has taken a long time to be published is the lack of information fluency in the Netherlands. Many attempts have been made by the author since 2012 to contact authorities at the ministry of education to request access to their statistics and programme objectives, yet no feedback has been provided. Also, attempts were made to collect more questionnaire responses from other colleges, but most bilingual coordinators were reluctant and some refused to share their ideas and reflections. It was clear that much is missing in the CLIL programme across colleges in the Netherlands and very little has been done to make it transparent.

Although the sample in this study is small, the findings presented are indicative of a need to reassess the tests administered after bilingual or CLIL courses. This brings Marsh's (2002) call for an assessment apparatus that tests both content and language skills into the spotlight. Such tools do exist in some European countries, such as Germany and England, where exit tests reflect both content and target language skills; the question is: why have they not been adopted by other countries?

To the best of our knowledge, no studies have been conducted to support the validity of exit tests delivered in Dutch for students in the CLIL programme in the Netherlands. There is also a misunderstanding regarding the difference between bilingual and CLIL programmes.

We can only assume that information such as test scores or teachers' standpoints is not very sensitive and could be shared with academics interested in remedial research; however, we were left with the idea that access to such information could be provided only to academics affiliated with one of the major universities which provide CLIL training in the Netherlands. Therefore, we present this article with the hope that a CLIL enthusiast, preferably a Dutch-national who has studied through the bilingual stream, would be interested in addressing these concerns through empirical research; if any correlation can be found between students' scores on content subjects delivered in the target language and content subjects delivered in Dutch, we would have no concerns about the future or learning experiences of students studying through bilingual education in the Netherlands.

References

- Chambers, E. and Northedge, A. (1997). *The arts good study guide*. London: The Open University.
- Education Encyclopedia. (2015). *Netherlands - secondary education*. Retrieved July 6, 2015, from <http://education.stateuniversity.com/pages/1069/Netherlands-SECONDARY-EDUCATION.html>
- Escobar, U. C., and Sánchez, S. A. (2009). *Language learning through tasks in a content and language integrated learning (CLIL) science classroom*. Retrieved July 10, 2015, from <http://hdl.handle.net/10481/31835>
- European Commission. (2011). *Content and language integrated learning*. Retrieved August 8, 2011, from http://ec.europa.eu/languages/language-teaching/content-and-language-integrated-learning_en.htm
- European Platform. (2013). *Bilingual education in Dutch schools: A success story*. Retrieved July 8, 2015, from <http://www.europeesplatform.nl/.../2131107-Bilingual-education-in-Dutch-sch>
- Expatica. (2014). *Education in the Netherlands*. Retrieved July 7, 2015, from http://www.expatica.com/nl/education/Education-in-the-Netherlands_100816.html
- Gracia, O. (2009). *Bilingual education in the 21st Century: A global perspective*. UK: Wiley-Blackwell.
- Heaton, J.B. (1975). *Studying in English*. England: Longman Group Ltd.
- Kaplan, R.B. and Baldauf, R. B. (1997). *Language planning: From practice to theory*. UK: Multilingual Matters Ltd.
- Lewis, M. and Reinders, H. (2003). *Study skills for speakers of English as a second language*. London: Palgrave Macmillan.
- Marsh, D. (2002). *CLIL/EMILE the European dimension: Actions, trends and foresight potential*. Finland: UniCOM, Continuing Education Centre.
- Mehisto, P., Marsh, D. and Frigols, M.J. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.
- Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.
- Várkuti, A. (2010). 'Linguistic benefits of the CLIL approach: measuring linguistic competences'. *International CLIL research journal* 1/3.
- Wolff, D. (2007). 'Some educational and methodological principles of content and language integrated learning (CLIL)'. *Goethe institut: CLIL integriertes fremdsprachen – und sachfach -lernen*. Retrieved December 14, 2011, from <http://www.goethe.de/ges/spa/dos/ifs/met/en2747826.htm>