12 Certifying teachers' foreign language proficiency: developing a performance test for Italian CLIL teachers

Geraldine Ludbrook
Dept of Language Sciences, University of Venice

Abstract

As the provision of Content and Language Integrated Learning (CLIL) is moving increasingly into mainstream education, the call for certified qualification of CLIL teachers is growing. A project is being developed at the University of Venice that seeks to identify the L2 weaknesses and needs of CLIL teachers in Italy, and to respond with specialised pre- or in-service training. The project aims to design a performance test to certify both the L2 competence of CLIL teachers and their knowledge of CLIL methodology. For the purposes of the pilot test, it will focus on the teaching of science through English.

As CLIL is not easily understood as a construct, making the measurement of ability complex, there are many directions for research within this context, which include examining how the interplay of general foreign language proficiency, subject-specific language, the language of classroom interaction, and code-switching contribute to the construction of CLIL science classroom discourse, in addition to what minimum L2 language proficiency is required of the CLIL teacher to effectively handle the methodology needed to implement this approach. This paper will discuss how investigation of the target language use through the qualitative analysis of data from CLIL science classroom observation can help to address some of the main issues that challenge performance test design, such as construct description and test task development.

Introduction

Content and Language Integrated Learning (CLIL) is an educational approach that has evolved in Europe from the new needs for multilingualism set out by the Council of Europe. The term CLIL refers to situations where

subjects, or parts of subjects, are taught through a foreign language with dual-focused aims, namely the learning of content, and the simultaneous learning of a foreign language (Marsh 1994). The approach has been rapidly introduced into mainstream education throughout Europe, yet many issues related to the CLIL teacher remain unaddressed. In particular, the question of CLIL teachers' foreign language proficiency is a little researched area, despite the fact that it is widely considered an essential feature of the success of CLIL: 'One crucial aspect of CLIL should also be spelled out: how good should CLIL teachers' proficiency in the language of instruction be and how could that level be reliably checked?' (Takala 2002:40).

Education authorities throughout Europe have different standards for CLIL teacher foreign language proficiency: the Dutch education authorities recommend at least a B2 level of the CEFR (Common European Framework of Reference), in Hungary a B2–C1 level is required, whilst in Finland the Ministry for Education proposes a C2 level of proficiency, which is also the obligatory level in Poland (Eurydice European Unit 2006:43). Other scholars argue that native speaker skills are a necessary pre-requisite (Smith 2005), while the opinion of one of the leading advocates of the CLIL approach is that 'Teachers do not need to have native or near-native competence in the target language for all forms of delivery, although naturally they need a high level of fluency' (Marsh 2002:11).

Nevertheless, the call for qualified CLIL provision is increasing. In France, additional certification of competence has been required for teachers of a non-language subject teaching in a foreign language since 2003. Germany, too, has introduced additional teacher qualifications for bilingual teaching in some states (Eurydice European Unit 2006:43–44).

As CLIL moves increasingly into mainstream education in Italy, the need for specialised pre-service training and qualification of CLIL teachers is becoming more evident. A project is being developed at the University of Venice to design a performance test to certify the L2 proficiency of CLIL teachers in Italy. The pilot test will be focused on the science classroom, the most common subject taught within the CLIL approach; for the purposes of the initial study, the foreign language used will be English, although aims are to extend the test to other languages to meet the multilingual needs of Italian CLIL.

This paper will examine central issues to be addressed in the development of a language performance test: the definition of the construct to be measured and the specification of the tasks to be administered. An initial analysis of the target language use through a small case study is reported on, and some suggestions for further research are proposed.

Performance language testing

Performance language testing generally tends to follow two main schools of thought. The first is largely a construct-based approach in which performance

is the means by which a language sample is elicited so as to allow evaluation of second language proficiency: McNamara's 'weak' sense of the term (McNamara 1996:43). Test tasks may resemble or simulate real-world tasks, but the real focus of the test is the underlying knowledge and ability that is revealed in the performance, the 'vehicle of assessment' (Messick 1994:14). The meaning of the construct of this kind of performance test 'is tied to the range of tasks and situations it generalises and transfers to' (Messick 1994:15), and provides the criteria used in evaluating task performance.

The construct is generally based on an explicit theory of language and language use, such as the models of communicative language ability developed by Bachman (1990), Bachman and Palmer (1996), and Canale and Swain (1980). Background and topic knowledge, too, are often included in the construct of performance tests for specific purposes, in which test content and test methods stem from an analysis of a specific use situation or context, 'capitalising' on special purpose abilities on the grounds that 'context-based tests may provide more useful information than general-purpose tests when the goal is to make situation-specific judgments about subjects' communicative language ability' (Douglas 1997:18).

The second theory is the task-based approach to performance testing – McNamara's 'strong' sense of the term (McNamara 1996:43) – in which the fulfilment of the test-task is the 'target of assessment', and the second language is the 'medium' of the performance (Messick 1994:14). The test tasks simulate or replicate real-world tasks and the criteria used for evaluation of task fulfilment are based on real-world criteria. In its most pragmatic form, this approach may make no recourse to theoretical models of language use in the definition of the test construct, relying instead on a close analysis of the target language use: 'Task-based assessment does not simply utilize the real-world task as a means for eliciting particular components of the language system which are then measured or evaluated; on the contrary, the construct of interest in task-based assessment is performance on the task itself' (Brown et al 2002, cited in Bachman 2002:455). To what extent a test of this kind can actually provide a basis for interpretations beyond the task or task context, the question of justifying inferences from test performance (McNamara 1996:17), is precisely one of the challenges test designers working in this approach must address.

Bachman (2002) takes up this challenge and, whilst fully aware of the limitations of generalisability and extrapolation offered by atheoretical task-based testing, proposes that test design take into consideration both construct definition and task specification, attempting to define task characteristics as closely as possible to the facets of the assessment in question on the basis of both the analysis of the target language use domain and either an existing framework or a framework developed *ad hoc* for the test. Bachman also refers to the construct definition of the specific areas of language abilities

to be assessed, suggesting that the construct may include several separate components or involve a global definition. The construct can be rooted in a theoretical model, or based on a course syllabus, or defined from a needs analysis of the target language use domain; it may attempt to measure all or parts of any of these aspects of the construct. 'Planned integration of both tasks and constructs in the way they are designed, developed and used' may provide test developers with 'the full range of validation arguments that can be developed in support of a given inference or use' (Bachman 2002:471).

Defining an appropriate *a priori* construct for a performance language test for Italian CLIL teachers, and considering the test tasks to be designed, will therefore require a careful analysis of the target language use domain. The next section of this paper will examine the target language use in the Italian CLIL science classroom, drawing both on the literature and on a small case study recently carried out in the Italian context.

Background to the study

CLIL methodology

The CLIL approach is rooted in a dual focus on language and content, a vehicular use of the foreign language. The approach draws heavily on strategies taken from models of content-based teaching (Brinton, Snow and Wesche 1989). The 'sheltered approach' to subject matter teaching used in content-based instruction involves a wide range of scaffolding strategies to communicate meaningful input in the content area, as well as adapting the language of texts or tasks and using methods such as visuals, graphic organisers, or co-operative work to make instruction more accessible to students with low levels of L2 proficiency.

Another resource that is drawn on in CLIL methodology is task-based teaching. In this method, teachers 'interactionally support task performance in such a way as to trigger processes such as the negotiation of meaning and content, the comprehension of rich input, the production of output and focus on form, which are believed to be central to (second) language learning' (Van Avermaet, Colpin, Van Gorp, Bogaert and Van den Branden 2006:175). In task-based learning classrooms, the teacher tends to ignore language errors and focus more on the real aim of the task. In this way, the teacher 'puts the initiative for solving comprehension problems, running the conversation and initiating the topic into the hands of the learner' (Van Avermaet et al 2006:175).

CLIL teachers therefore have to possess a level of L2 competence that will enable them to implement CLIL methodology. But the approach calls for further considerations. CLIL teachers are also required to devise and modify materials and tasks that will permit the learners' access to the content subject in the foreign language. Moreover, CLIL teachers and learners are

all simulating an L1 classroom situation in a language that is foreign to both groups and, although a strong L2-only policy is generally advocated, the question must be raised as to what role the teacher's and learners' own L1 will play in the CLIL classroom and what use CLIL teachers can make of it to enhance communication. In addition to the other language skills mentioned above, CLIL teachers need sufficient specialist language knowledge, of both genre and lexis, to teach the content subject in the foreign language.

CLIL in the Italian context

In 1999, education reform in Italy paved the way for a more widespread introduction of the CLIL approach in mainstream education. State schools were given greater autonomy to introduce and develop different forms of teaching that more closely met the needs of their students. Amongst these was the possibility to teach content subjects in a foreign language. Another innovation was the introduction of more flexible forms of teaching, in particular the concept of teaching modules, which may be of variable length, from a few hours to several months, and may have a cross-curricular nature (see Coonan 2002:43–44). The focus on flexible language instruction was further reinforced by *Progetto Lingue 2000*, a project of the Italian Ministry for Education to improve the quality of foreign language teaching in the state school system (MPI 2000). CLIL is currently delivered in over 100 pilot projects throughout the country, generally in a modular format. Although English is the most popular foreign language, all the Italian projects have a strong multilingual policy, and CLIL in French, German and Spanish is strongly encouraged (Eurydice European Unit 2006:34).

Since the early 1990s, Italian education authorities have organised projects for CLIL teacher development. In the Veneto region, for example, the University of Venice has run training courses in CLIL methodology for in-service teachers in collaboration with regional education authorities since 2002, and is working to introduce the training of pre-service teachers (Coonan 2004a). In addition, pan-European CLIL projects, under the Socrates scheme, have funded teacher mobility programmes for language and subject teachers alike, to improve their language skills or to follow CLIL teacher training courses abroad (Coonan 2002:107–108).

There are some content teachers who teach CLIL on their own; however, in Italy CLIL is mainly provided through a teaching team of subject and foreign language teachers. In the Italian CLIL classroom, the teaching partnership seems to be characterised by features of the complementary/supportive teaching team, defined by Maroney (1995) as one in which 'one teacher is responsible for teaching the content to the students, while the other teacher takes charge of providing follow-up activities on related topics or on study skills'. In some cases, the collaboration takes place before the lesson and the

content teacher manages the lesson on his/her own. More commonly, in addition to shared preparation, both teachers are always present in the classroom at the same time (see the examples reported in Coonan 2004b).

The case study

The principal objective of the case study was to pilot a classroom observation checklist developed for the purpose of identifying the language features involved in CLIL teaching performance as a tool for defining a framework for the construct of the test. A second goal of the study was to begin to examine what minimum level of language proficiency a content teacher working in a team-taught CLIL context might need to have.

The school chosen for the case study was an Istituto Tecnico Statale, a technical secondary school that trains students for employment in the sectors of trade, tourism and surveying. It was chosen as the context for the case study because English-language CLIL in the science classroom has been implemented here for several years, generally in the first two years of secondary school with students aged 14–16.

The class observed was made up of 20 students aged 15: four boys and 16 girls. The students had already received science instruction in CLIL the previous year with the same teachers and were therefore familiar with the procedures and classroom rules regarding the use of English, as well as with pair and group work activities.

Methodology

A qualitative approach was adopted in the study, incorporating methods of data collection to build up as rich a picture as possible of the CLIL learning and teaching environment. Semi-structured interviews were carried out separately with content and EFL teachers both to obtain background information on the classes to be observed and to put together a portfolio of the subject teacher's English competences.

Four CLIL lessons were observed and audio recordings made, which were then transcribed and coded for a close study of the CLIL classroom discourse and, in particular, the CLIL teacher's use of language. As the observations of the CLIL classrooms were exploratory and diagnostic, an observation checklist was chosen as a useful tool as a framework for the observation. Often used to provide a sampling frame to classroom observation (see, for example, Montgomery 2002) this instrument has also been used for both *a priori* and *a posteriori* analysis of output in speaking test tasks (O'Sullivan, Weir and Saville 2002).

Two checklists devised for classroom observation of non-native English speaking teachers were examined in the preliminary stages of the development

of a similar tool for Italian CLIL classrooms. De Graaff, Koopman, Anikina and Westhoff (2007) report on the development of an observation tool based on principles from second language pedagogy. The checklist covers several aspects of CLIL methodology: focus on form, focus on meaning and different kinds of scaffolding. It does not, however, look at the fields of general language proficiency, subject-specific language or classroom management, as the specific aim of De Graaff et al's study is to detect effective CLIL pedagogy.

Closer to the aims of the Italian testing project, albeit in a non-European context, Elder (1993) illustrates an observation schedule developed 'to assess the English language proficiency of non-native speaker graduates training as secondary mathematics and science teachers' in Australian schools (Elder 1993:235). The schedule contains features of both language and language-related behaviour based on the literature of classroom communication, considered crucial for effective teacher performance and revised to include only those features which were found 'to discriminate among non-native speaker teachers' (Elder 1993:237). The schedule was produced for use by teachers of mathematics and science, so was formulated to be meaningful to non-language experts and designed to be used during a 15-minute observation of teacher performance.

As Elder's 1993 schedule contained some of the main categories of language features considered relevant to the Italian CLIL context, it was decided to use this schedule as a starting point for the CLIL classroom observations. A group of Italian experts in CLIL methodology, teacher trainers, CLIL teacher trainers, and trainee teacher supervisors was asked to indicate what aspects of the original schedule they considered to be important features of the Italian CLIL classroom. Their evaluations were then incorporated into a revised version of the schedule that attempted to focus more precisely on the foreign language needs of the Italian CLIL teacher. An additional section was added to the schedule, which took into consideration code-switching, intended here as any kind of alternation between L1 and L2, not specifically switching, borrowing or mixing. Although L2-only interaction is encouraged, the effective use of L1 is an important feature in CLIL classroom discourse (see, for example, the studies by Butzkamm 1998 and Nikula 2005 for further research on code-switching practices in CLIL classrooms). The two descriptors added concerned the teacher's effective use of L1/L2 code-switching and the teacher's encouragement of effective code-switching by the students. The CLIL observation schedule used in this pilot study can be found in the Appendix.

In addition to teacher interviews and classroom observation, various documents used in the CLIL module were examined: handouts prepared by the content teacher and used by the EFL teacher to prepare students for the CLIL module, and the tasks set for students during the module. The

end-of-module test was also looked at. The test combined multiple-choice and true/false items with open-ended questions and was marked by both science and EFL teachers, with separate grades given for content and language.

A group interview with six teachers implementing the CLIL approach in the school in question was organised to discuss the specific questions of CLIL teachers' language needs and levels, drawing on the experience of both foreign language and content subject teachers. Coonan (2007) has conducted considerable research on the 'insider' view of the CLIL classroom, working with subject and language teacher teams implementing the approach in Italian classrooms, to record their perceptions of the CLIL classroom. Her results show that, due to a high degree of teacher awareness, useful information can be gleaned from CLIL teachers' experience in the classroom. In this study, the group was made up of three content teachers (two science teachers and one history teacher), two Italian EFL teachers and one native speaker 'conversation' teacher. All six had at least two years' experience of the CLIL approach and all had completed, or were currently following, CLIL methodology training at the University of Venice.

The CLIL science teacher

A semi-structured interview was carried out with the science teacher before the observations with the purpose of gathering data on his CLIL experience and to put together a portfolio of his English language background. Male, an Italian national and native speaker, PP has a degree in chemistry and has been teaching science at secondary school for over 26 years. After studying English at school for eight years, he then conducted most of his university studies using English language textbooks. Since then, his use of English has largely been limited to personal use (mainly television and film), consultation of online scientific journals, and attendance at European conferences. He has, however, been involved in Comenius exchange programmes, working with visiting teachers from schools in Wales and Lithuania, even though he has not spent time in either country.

PP is a strong advocate of the CLIL approach and has been instrumental in introducing it into the school. He completed a CLIL training course at the University of Venice and was involved in a research project involving CLIL teachers in Italy. He uses internet resources to provide material for his CLIL module, including Massachusetts Institute of Technology videos of science lessons.

In an attempt to establish PP's level of English proficiency two different tools were used. Firstly, he was asked to provide an evaluation of his level of language competence using the Common European Framework of Reference self-assessment grids (see CEFR, Tables 2 and 3, Council of Europe 2001:26–29). He placed himself within the B1 level for all skills, with the exception

of reading comprehension, which he evaluated as B2. Secondly, PP assessed his English proficiency using the DIALANG diagnostic language tests. In reading, grammar and vocabulary, his results were at the C1 level, whereas his listening comprehension score was slightly lower at the B2 level.

When asked about his perception of his strengths and weaknesses, PP spoke of the amount of care with which he prepares his teacher-fronted laboratory lessons, with which he feels confident as, familiar with his subject, he can concentrate on his use of English. His greatest difficulties arise in unplanned interaction, the unpredictable lexis that he might require during the lesson to respond to student requests for information.

PP also outlined the structure of the CLIL module planned. It was to last 18–20 hours and would be delivered in the last five weeks of the school year. The students would first be made familiar with some of the vocabulary to be used by the EFL teacher in the English lessons. Then a series of four lessons would be held in the physics laboratory. In these teacher-fronted lessons, PP would carry out demonstrations and experiments related to the theme of the module. The next four lessons would be group work held in the multimedia laboratory. The students would work in pairs on a task that involved retrieving information from the internet. The students would then prepare a PowerPoint presentation of the completed task.

PP also provided insight into the role of the two teachers in the CLIL teaching team. He explained that he chose the materials to be used in class, mainly from the internet for its greater flexibility, adapting them slightly, mainly by reducing the length. He then passed the material on to the EFL teacher who devised exercises to be used in her EFL lessons. He stressed that the role of the EFL teacher in the CLIL classroom is that of providing language support, intervening when she sees students in difficulty, or when a lexical problem occurs.

The EFL teacher

The formal interview with the EFL teacher took place after the observations, although several informal conversations had taken place before and during the observations. She had team-taught CLIL with PP and another science teacher at the school for two years. The teacher confirmed that during the CLIL module, all her EFL lessons were given over to preparation of the CLIL science lessons. Her assessment of the class's English language skills was that they had an overall good level of comprehension with varying levels of written and oral production.

With regard to PP's language needs, the EFL teacher saw the shift from his working within his subject (what she called ESP) to other registers, such as class management, as being his greatest difficulty, as he lacked the 'lexical richness and flexibility' to answer student questions with ease. She also

mentioned the issues of intonation and pronunciation as being areas of difficulty for this particular teacher.

Observing the CLIL classroom

The first two lessons observed were held in the physics laboratory and involved PP explaining a process and illustrating it through a series of practical demonstrations. The topic of the module was electrostatics. These lessons were science teacher-fronted activities, while the EFL teacher stood at the whiteboard providing written support (for example, irregular verbs, specialist lexis) and occasionally intervening orally.

A further two lessons observed were held in the multimedia laboratory. In these lessons, the students worked in pairs retrieving information from the internet to respond to a series of questions they had been assigned while the teachers monitored and assisted them.

Classroom management

A first observation was that in both lesson types (teacher-fronted and group work), most of the classroom management, such as introducing the lesson, setting up activities, and disciplining the students, was carried out by the EFL teacher. The EFL teacher opened lessons with revision of content material dealt with in the previous lesson; she also closed lessons with instructions for the next meeting. Many of the descriptors in the CLIL classroom observation schedule could therefore not be related very closely to the CLIL content teacher's performance.

The fact that much of the classroom management was carried out by the EFL teacher meant that it was hard to evaluate the validity of the CLIL schedule on the basis of these observations. Further observations will be needed to verify whether this division of tasks by the EFL and content teachers is common to many teaching teams or whether it was specific to this particular pair.

Using subject-specific language

During the experiments in the teacher-fronted lessons, a useful sample of subject-specific language was recorded. Regarding the knowledge of subject-specific terms, PP appeared to have good control of the specialist lexis of electrostatics. Moreover, not only did he pronounce specialist terms clearly and correctly, he also corrected the EFL teacher's mispronunciation of specialist terms. In addition, he also helped the EFL teacher in the correct spelling of subject-specific words.

PP linked ideas using very simple connectors, such as, for example, *if* and *so*. The students appeared to have little difficulty in understanding the simple demonstrations of the principles of electrostatics. In addition PP's

experiments were scaffolded by the handouts distributed to students and to which the EFL teacher invites students to refer during the lesson.

The section of the observation schedule regarding the use of subject-specific language seems to correspond to the content teacher performance observed.

Using L1 and L2

In their respective interviews, both content and EFL teacher claimed that they attempted to maintain a strong L2 policy in the CLIL classroom, using Italian only as a last resort after several attempts to reformulate have been made. Both teachers repeatedly issue instructions to students to use English.

In the classroom interaction recorded, Italian was used in four different ways:

1. by students in response to an explicit request for the Italian translation by the content teacher:

   Content Teacher (CT): This is the box full of Styrofoam chips. OK? It what is Styrofoam in Italian?

   Student (S): *Polistirolo.*

2. in 'private' interaction between the content and EFL teachers:

   CT: This is a bakelite rod. OK. [*sotto voce* to EFL Teacher who is writing the term on the board] *bakelite con kappa*. [bakelite written with a k]

3. by the content teacher in response to a student question in Italian:

   S: And what mean 'drive' in this case?

   CT: Uh you'll see you'll see.

   S: *Qual'è il significato di 'drive'?* [What does 'drive' mean?]

   CT: *Immergere.*

4. the content teacher provides translation of previous statement in English.

   CT: Bakelite rod. Bakelite is *bachelite* in Italian.

The descriptors seem to capture the use of L1/L2 code-switching in the lessons observed. If further observations confirm that the different ways in which Italian is used by the content teacher are systematic to CLIL science classrooms, the descriptors might be articulated to take this into consideration. In the case of this study, the interactions recorded do not seem to correspond to the explicit policy advocated by the teachers in their interviews.

General language proficiency

I have left to last the most complex section of the CLIL teacher's performance, that of general language proficiency. This is the section in which issues

such as intelligibility, fluency and accuracy come into play, requiring clear definitions if qualitative judgments are to be made.

Native-speaker language proficiency does not seem to be the model in the CLIL approach. Many countries have indicated minimum levels of teacher foreign language proficiency using the Common European Framework of Reference, and educationalists promoting the CLIL approach have explicitly advocated the non-native speaker model (see Graddol 2005 and Marsh 2002).

Performance on language tests is typically judged with reference to a native speaker ideal. Some scholars have, however, challenged the concept that the native speaker is an appropriate model of English for language testing, and teaching, outside Kachru's (1990) 'Inner Circle' (see work by Elder and Davies 2006, House 2002, Jenkins 2006, Seidlhofer, Breiteneder and Pitzl 2006, Taylor 2006). CLIL would seem to be a clear example of English used as a lingua franca in the classroom, albeit between non-native speakers sharing the same first language. A discussion of the issues involved in terms of what model is to be used in the evaluation CLIL teacher performance will be necessary.

Intelligibility of expression

The concept of intelligibility is a complex one involving both speaker and hearer. The hearer's understanding may depend on whether the speaker's accent is familiar, on the hearer's inferencing skills and on knowledge of the topic. The speaker's production may depend on pronunciation (stress, rhythm, intonation, voice quality and sounds), delivery (hesitations, uncertainty, volume), grammar, sound symbol relationships.

Although a discussion of how the processes by which understanding is achieved in ELF interaction may be qualitatively different from native speaker-based interaction will be necessary (see Brown and Lumley 1998, Elder and Davies 2006, Han and Singh 2007, Pickering 2006), for the purposes of this initial study, intelligibility of the content teacher in the CLIL classroom in this study is considered in terms of the researcher's perception of the Italian student listeners' understanding based on classroom observations. In other words, whether the teacher was intelligible to students with limited English proficiency working with a teacher whose accent, pronunciation and delivery are familiar, as is the topic of the lesson.

In the observations recorded, the students seem to have little difficulty in understanding the content teacher working in English. Their problems appear to be in understanding the subject content, or specialist lexis, rather than the science teacher's delivery.

PP stresses important words and makes easy transition from one experiment to the next. His use of non-verbal strategies would seem to be appropriate to the science classroom situation.

Fluency and flexibility of expression

In the literature, fluency as a descriptor for oral language performance has been defined in various ways. Crystal, for example, defines fluency as 'smooth, rapid, effortless use of language' (Crystal 1987:421); Brumfit talks of 'natural language use whether or not it results in native speaker-like language comprehension or production' (Brumfit 1984:56). Chambers (1997) looks at recent research into definitions of fluency, including speech rate and pauses, and compares the difference in fluency in native and non-native speakers.

In the presentation phase of the lessons, PP maintains a fluent description of the experiments and procedures he is carrying out. Less fluency is noted in the teacher–student interactions during the monitoring of students in group work activities.

Turning to the descriptor regarding flexibility of expression, PP appears to have little lexical flexibility and variety, and often tends simply to repeat his previous statement with little variation on the original, as in this example:

CT: Like charges repel each other. For example positive and positive they are like charges or negative and negative. Like charges are of the same sign. OK? Both positive or both negative.

Accuracy of expression

In the testing of oral performance, accuracy is generally perceived as being based on grammatical correctness, and is often contrasted with fluency. Within the CLIL approach, however, as in other forms of content-based instruction in which there is a dual focus on language and content, fluency is favoured over accuracy and native-speaker competence is not aimed for (see, for example Marsh 2002:36). Student errors are generally only corrected when communication breakdown or misunderstanding occurs. This explicit policy regarding accuracy will therefore also affect the model the CLIL content teacher is expected to provide, especially in the co-taught classroom.

Assuming, for the purposes of this initial study, grammatical correctness as a definition of accuracy, PP makes very few errors in the teacher-fronted lessons. Some examples of grammatical and syntactical inaccuracy are:

CT: I'm going to do some experiments and then I'm going to commenting them with you.

CT: The electroscope told that the Cavendish hemisphere is charged.

CT: When I shake fast the rod . . .

CT: What kind of forces are they between the paper and the white board?

None of these inaccuracies caused any misunderstanding with the students.

Planning, monitoring and repair

PP has evidently planned his teacher-fronted lessons with great care. In the interview, he claims each lesson takes 3 to 4 hours' preparation. When he

does make mistakes or slips, he is very often able to correct himself, whether they are minor grammatical slips or errors of pronunciation.

Conclusions and further research

This initial analysis of the observation schedule seems to indicate that it is a useful tool in the observation of the CLIL science teacher's performance. Most of the descriptors seem relevant and capture salient moments of the classroom interaction.

In future research, validation of the checklist will continue with repeated observations of science classrooms and through focus group discussions with other teachers using the CLIL approach in Italy aimed at further clarifying and refining the checklist. This framework drawn from the target language use domain will form the basis for the construct underlying the test, and guide the construct-based scoring criteria used for performance evaluation.

The case study also provided insight into the tasks that might be designed for the performance test, operationalising the construct. The teacher needs to be able to prepare and deliver teacher-focused presentation of subject-specific material, with the aid of practical demonstrations, board work, and written handouts; to set up and monitor pair and group work task-based activities, interacting with the students on issues regarding both content and language; to evaluate student performance, both oral presentation of group work tasks and written test production. Establishing the nature of the test tasks and defining the task characteristics will require careful consideration of what degree of authenticity and interactivity is desired and can be achieved in a performance test simulating a classroom situation.

As regards the issue of starting to establish what minimum level of language skills a teacher implementing CLIL in Italy needs to possess, from this brief case study it would appear that in the science classroom, a content teacher with a language profile that ranges between a B2–C1 level of general English competence seems to be able to function in the presentation phase of the CLIL lesson. His delivery of prepared material shows considerable accuracy – grammatical, syntactical, lexical and of pronunciation – and he is able to monitor and correct both his own errors, and those of students. When dealing with subject-specific terms, he is also able to monitor the students' and the EFL teacher's speech. Fluency and pronunciation also appear to be appropriate for the level of the students being taught.

However, it would also seem that this level of general English competence is insufficient for some aspects of the teaching performance. The content teacher depends heavily on the presence of the EFL teacher for the phases of the lesson that require more flexible and interactive language use, such as opening and closing the lesson, and setting up activities. Yet his familiarity

with his subject and teaching experience provide him with tools that assist his teaching performance in English.

Some of the questions that future research will therefore have to explore are what factors contribute to task fulfilment: language proficiency or teaching strategies/classroom competence, and how teaching experience can be separated from language performance in the test situation in order to avoid construct-irrelevant variance, a major threat to construct validity in performance tests.

Acknowledgements

The author would like to thank the CLIL teachers at Istituto Tecnico Statale per il Commercio, il Turismo e per Geometri Girardi, Cittadella, for their kind collaboration.

References

Bachman, L F (1990) *Fundamental considerations in language testing*, Oxford: Oxford University Press.

Bachman, L F (2002) Some reflections on task-based language performance assessment, *Language Testing* 19 (4), 453–476.

Bachman, L F and Palmer, A S (1996) *Language testing in practice*, Oxford: Oxford University Press.

Brinton, D M, Snow, M A and Wesche, M B (1989) *Content-based second language instruction*, Mass: Heinle & Heinle.

Brown, A and Lumley, T (1998) Linguistic and cultural norms in language testing: a case study, *Melbourne Papers in Language Testing* 7 (1), 80–96.

Brumfit, C J (1984) *Communicative Methodology in Language Teaching: The Roles of Fluency and Accuracy,* Cambridge: Cambridge University Press.

Butzkamm, W (1998) Code-switching in a bilingual history lesson: The mother tongue as a conversational lubricant, *International Journal of Bilingual Education and Bilingualism* 1 (2), 81–99.

Canale, M and Swain, M (1980) Theoretical bases of communicative approaches to second language teaching and testing, *Applied Linguistics*, 1 (1), 1–47.

Chambers, F (1997) What do we mean by fluency?, *System* 25, 535–544.

Coonan, C M (2002) *La lingua straniera veicolare*, Torino, UTET.

Coonan, C M (2004a) La formazione avanzata a livello universitario dei docenti CLIL in Italia. Retrieved 17.5.2006. <gold.indire.it/nazionale/content/index.php?action=read_cnt&id_cnt=5967>.

Coonan, C M (2004b) *CLIL: un nuovo ambiente di apprendimento*, Venice: Libreria Editrice Cafoscarina.

Coonan, C M (2007) Insider Views of the CLIL Class Through Teacher Self-Observation and Introspection, *The International Journal of Bilingual Education and Bilingualism*, 10 (5), 625–646.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.

Crystal, D (1987) *The Cambridge Encyclopaedia of Language*, Cambridge: Cambridge University Press.

De Graaff, R, Koopman, G J, Anikina, Y and Westhoff, G (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.
Douglas, D (1997) *Testing Speaking Ability in Academic Contexts – Theoretical Considerations*, TOEFL Monograph Series 9, Princeton, NJ: ETS.
Elder, C (1993) How do subject specialists construe classroom language proficiency?, *Language Testing* 10 (3), 235–253.
Elder, C (1994) Performance testing as a benchmark for LOTE teacher education, *Melbourne Papers in Language Testing* 3 (1), May 1994, 1–25.
Elder, C (2001) Assessing the language proficiency of teachers: are there any border controls?, *Language Testing*, 18 (2), 149–170.
Elder, C and Davies, A (2006) Assessing English as a Lingua Franca, *Annual Review of Applied Linguistics* 26, 282–301.
Eurydice European Unit (2006) *Content and Language Integrated Learning (CLIL) at School in Europe*. Retrieved 28.05.2006. <www.eurydice.org/portal/page/portal/Eurydice/showPresentation?pubid=070EN>
Graddol, D (2005) CLIL debate questions and answers, *Guardian Weekly*, Wednesday April 20 2005. Retrieved 18.08.2007 <www.guardian.co.uk/theguardian/2005/apr/20/guardianweekly.guardianweekly13>
Han, J and Singh, M (2007) World English Speaking (WES) student-teachers' experiences of schools: curriculum issues, transnational mobility and the Bologna Process, *Transnational Curriculum Inquiry*, 4 (1). Retrieved 13.05.2008. <http://nitinat.library.ubc.ca/ojs/index.php/tci>
House, J (2002) Developing pragmatic competence in English as a lingua franca, in Knapp, K and Meierkord, C (Eds), *Lingua franca communication*, Frankfurt a. M. et al.: Lang.
Jenkins, J (2006) The spread of EIL: A testing time for testers, *ELT Journal*, 60 (1), 42–50.
Kachru, B (1990) World Englishes and applied linguistics, *World Englishes* 9 (1), 3–20.
Maroney, S (1995) *Team Teaching*. Retrieved 12.02.2008. <www.wiu.edu/users/mfsam1/TeamTchg.html>
Marsh, D (1994) *Bilingual Education & Content and Language Integrated Learning*, Paris: International Association for Cross-cultural Communication, Language Teaching in the Member States of the European Union (Lingua) University of Sorbonne.
Marsh, D (Ed.) (2002) *CLIL/EMILE. The European Dimension*, Finland: UniCOM Continuing Education Centre, University of Jyväskylä.
McNamara, T (1996) *Measuring Second Language Performance*, Harlow: Longman.
Messick, S (1994) The interplay of evidence and consequences in the validation of performance assessments, *Educational Researcher*, 23 (2), 13–23.
Ministero della Pubblica Istruzione (2000) *Progetto Lingue 2000*, Roma: Gruppo Lingue Straniere, MPI.
Montgomery, D (2002) *Helping Teachers Develop through Classroom Observation*, 2nd ed, London: David Fulton Publishers.
Nikula, T (2005) English as an object and tool of study in classrooms: Interactional effects and pragmatic implication, *Linguistics and Education* 16 (1), 27–58.
O'Sullivan, B, Weir, C J and Saville, N (2002) Using observation checklists to validate speaking-test tasks, *Language Testing* 19 (1), 33–56.

Pickering, L (2006) Current research on intelligibility in English as a lingua franca, *Annual Review of Applied Linguistics*, 26, 219–233.
Seidlhofer, B, Breiteneder, A and Pitzl, M-L (2006) English as a Lingua Franca in Europe: Challenges for applied linguistics, *Annual Review of Applied Linguistics* 26, 3–34.
Smith, K (2005) Is this the end of the language class?, *Guardian Weekly*, January 21, 2005. Retrieved 11.11.2006. <education.guardian.co.uk/tefl/teaching/story/0,15085,1394830,00.html>
Takala, S (2002) Positioning CLIL in the Wider Context, in Marsh, D (Ed.) *CLIL/EMILE. The European Dimension*, Finland: UniCOM Continuing Education Centre, University of Jyväskylä, 40–42.
Taylor, L (2006) The changing landscape of English: implications for language assessment, *ELT Journal*, 60 (1), 51–60.
Van Avermaet, P, Colpin, M, Van Gorp, K, Bogaert, N and Van den Branden, K (2006) The role of the teacher in task-based language, in Van den Branden, K (Ed.) *Task-Based Language Education*, Cambridge: Cambridge University Press, 175–196.

APPENDIX

CLIL classroom observation schedule

General language proficiency

1. Intelligibility of expression

1.1 pronounces words/sounds clearly
1.2 utters sentences clearly, with suitable rhythm and intonation
1.3 stresses important words/ideas
1.4 clearly marks transition from one idea/lesson stage to the next, using words such as *so*, *now*, *right*
1.5 uses appropriate facial expressions, gestures, body movement

Fluency and flexibility of expression

1.6 speaks at a speed appropriate to the level of the class
1.7 speaks fluently, without too much uncertainty
1.8 can express ideas in different ways: rephrasing, elaborating, summarizing, exemplifying

Accuracy of expression

1.9 grammar of spoken and written language is generally accurate
1.10 uses correct spelling and punctuation in board work

Planning, monitoring and repair

1.11 plans what is to be said and the means to say it, exploiting any resources available
1.12 uses circumlocution and paraphrase to cover gaps in vocabulary and structure
1.13 backtracks when a difficulty is encountered and reformulates
1.14 corrects own slips and errors if s/he becomes aware of them or if they have led to misunderstandings

2. Using subject-specific language

2.1 demonstrates knowledge of subject-specific terms
2.2 pronounces specialist terms clearly
2.3 uses specialist terms judiciously, writing on board when necessary
2.4 makes clear the connection between ideas, stressing link words *if*, *since*, *in order*
2.5 explains concepts and processes in ways appropriate to the level of the class, using simple language and familiar/concrete examples
2.6 explains diagrams, models, graphs clearly
2.7 links new information to the students' previous knowledge

3. Using the language of classroom interaction

3.1 poses questions to check understanding of previously learned material/ new information
3.2 grades questions appropriately for the level of the class and the learning task: simpler to more complex; closed/open
3.3 responds appropriately to students' questions, requests for assistance
3.4 deals effectively with wrong answers, non-response, using scaffolding techniques such as requests for clarification and recasts
3.5 gives clear instructions for activities
3.6 makes effective use of teaching materials

4. Using L1 and L2

4.1 makes effective use of L1/L2 code-switching, clarifying rules with students
4.2 encourages students' effective use of L1/L2 code-switching