

6 Assessment issues in CLIL

The theme of assessment is a difficult and sometimes contentious area amongst CLIL teachers. In some respects it lies at the heart of the question of how to define the level of content–language integration, because, ultimately, no matter what is taught and how it is taught, the mode of assessment determines how the learners perceive the teacher's intention and, of course, also shapes performance data. In this chapter, we are dealing with classroom assessment as opposed to programme evaluation (which is addressed in Chapter 7). Programme evaluation involves looking at a complete CLIL course or an aspect of it and making a judgment regarding its effectiveness, for example through collection of data on learners' performance or attitudes. The distinction between assessment and evaluation is important, as each serves a different purpose. However, there is a potential overlap which is relevant to the question of whether we are assessing content, language or both. Programme evaluation might centre on learners' language attainment (many research reports do so) and this might be an appropriate place and method to carry out discrete language assessment as well.

Assessment processes can be broadly divided into summative and formative and this division forms a major distinction. Summative assessment makes a judgment on the capability of the learner at that point in time and, apart from offering that judgment back to the learner, it often leads to some form of information-giving to another party, for example the school management or the learner's parents. It is therefore associated with testing in a more formal setting or an end-of-unit, 'final' result, even if this is not obtained through an examination. Across the world there are many variations on final course and module testing processes, with a whole range of criteria in use for both content and language outcomes. CLIL units will need to mirror such systems in order to retain credibility as mainstream educational programmes. This point will be addressed again later in the chapter.

Formative assessment is more complex, as its intention is to be directly diagnostic with a view to immediately impacting on the learner's next steps. It is also formative for the teacher, because it can alter planning and practice mid-unit (or even mid-lesson) and not just after all the work is complete, as a summative test might do. Formative assessment was advocated first by Scriven (1967) and Bloom (1968). In common with these writers, Ames and Ames (1984) suggested moving away from a norm-referencing approach; they advocated a task-mastery approach using a learner's performance to structure goals for future improvement on an individual, rather than competitive, basis. This focus then began to develop in both research and practice. It included writers on motivation – for example

Dweck (1986), who argued that summative assessment demotivated learners – and assessment researchers, such as Sadler (1989), who argued for learners to be given authentic evaluative experience, so that they could identify work of high quality and evaluate their own progress towards it. Cohen (1994) brought a language-learning perspective to the issue by recommending formative activity alongside classroom tasks, so that the teacher could better understand students' skills and competences.

Clarke (2001) likens summative assessment to the simple measurement of a plant, and formative assessment to the feeding process which leads to growth. In the UK, as a result of research into assessment (including an important study by Black and Wiliam, 1998), the term 'Assessment for Learning' (AfL) (in Scotland, 'Assessment is for Learning') was coined to describe processes thought to be desirable across the curriculum. In 2002, the Assessment Reform Group in England produced a document of ten principles for AfL, which makes clear that both teachers and learners will benefit from the processes described and that formative assessment should be central to classroom practice. Some of the key features in this document are:

- the sharing of learning intentions (meaning that teachers tell students at the beginning of the lessons what they will learn)
- the use of success criteria (meaning that students will be told what the task will involve and what the outcome will contain)
- the involvement of learners in self- and peer-assessment
- the importance of feedback, which should be sensitive to learners' self-esteem and which should thereby positively impact on motivation.

Zangl, also advocating a formative approach, includes in her article three major conclusions about language assessment (but in a way that could be applied equally well to content assessment). She states that teachers should try to:

- assess the learner's proficiency within a multi-component framework, comprising not only domain-/structure-specific items, but also the use of language within the social context of the classroom;
- capture both the learner's individual profile and the performance level of the class as a whole; and
- trace the learner *along* his or her developmental path where time and experience act as constructive factors.

(Zangl, 2000: 257)

This chapter will focus on such formative assessment approaches, as it seems to us there is a strong case for formative assessment to be used on a regular basis and **summative assessment to be used systematically but rarely**. The strength of formative assessment processes, according to the researchers discussed above, is that they enhance learning to an extent where they actively support better summative outcomes. The pressure on CLIL courses to match first-language test results is immense and it is through this regular occurrence of focused classroom practice that CLIL teachers and learners can work towards

achieving such parity. We will next consider what the specific assessment issues are for a CLIL programme, and then explore how we might address them. We will use examples from practice of different modes of assessment and rationalize them in terms of the broader aims of CLIL as demonstrated by the theorization and Tool Kit offered in Chapters 3 and 4. Finally, we will summarize by giving some exemplars of good CLIL practice in assessment which reflect the principles of this chapter.

6.1 What are the main issues for assessment in CLIL?

Assessment is often a major area of teacher uncertainty in CLIL contexts and, as with other issues relating to CLIL, must be considered with the CLIL practitioners' specific situation in mind. One group of teachers and trainers in Catalonia met in 2007 to collect together and try to address the major questions regarding CLIL assessment. From amongst these teachers one group suggested the following:

What do we assess - CONTENT or LANGUAGE?

In what language do we assess?

Can students answer in Catalan?

What tools can we use for assessment?

How can we assess previous knowledge and/or progression?

How can I deal with learning difficulties?

Provided we assess in English, how can we minimize the effect of the language in the content assessment?

How can we evaluate the skills/processes? Example: planning and investigation / designing a work of art / reaching conclusions.

How can/should we assess group work?

These are the key questions asked by the majority of CLIL teachers when they meet to discuss practice. The starting point usually centres on three basic issues: *Do we assess content, or language, or both? Which is more important? How do we do this?* We can divide this set of questions into a series of more generic questions which probe the needs and demands of a specific CLIL context. For example:

- What do we mean by assessment in CLIL?
- Do we assess language or content first?
- Do we sometimes assess one and not the other? If so, which and when (and, more crucially, why and how)?
- What about cognition and culture?
- Who assesses?

- When do we assess?
- How do we assess?
- What is the role of standard examination systems?
- Is there a role for the Common European Framework (2001)?

The next section of this chapter begins to confront these questions by looking at whether to assess language or content, followed by the issues involved in assessing each of these.

Language or content?

This central dilemma was summarized early in an article by Short (1993), in which she explored alternatives to standard testing in CLIL. Short also raised the two essential questions which lie behind teacher uncertainty about assessment, both the *what* question and especially the *how* question:

The many varieties of alternative assessment include performance-based tests, portfolios, journals, projects, and observation checklists. Although these measures allow better demonstration of student knowledge, they can nonetheless confound teachers of language minority students. Complications arise first because teachers must determine whether the language or the content is being assessed in these alternative measures. Then teachers must distinguish between the language and content knowledge of the students and decide if one is interfering with the demonstration of the other.

(Short, 1993: 633)

Here we see that the two questions are linked: firstly – as mentioned as a key question in the previous section – should we assess *language* or *content*? Secondly, what methods can we use which will give us *reliable* assessment information – that is, will one element (content or language) impede the other?

The *how* is the bigger question and will rightly occupy a larger proportion of the chapter, but we will address the first immediately. CLIL units will all contain clear objectives, possibly fashioned around the 4Cs. Even if a different approach is taken by the CLIL planners, they will still at some point have had to construct statements regarding the content (concepts, knowledge and possibly skills) which is to be covered by the unit and one or more statements regarding language. The language objectives may relate simply to communicating the content effectively, or they may include notions (such as specialist vocabulary from the unit) or functions (such as the ability to discuss effectively) or even be form-focused (for example, concerning effective use of the past tense). The teacher designing the unit will know what she or he wishes to teach and what the overall purpose of the CLIL module is. Therefore, the answer to the ‘language or content’ question is determined by the relative priority within those objectives. It is important to have a clear head about that priority; we have taken a position in this book that the content should always be the dominant element in terms of objectives, even though we intend that language will be learned securely alongside the content’s concepts and skills. With this perspective in mind,

we will turn now to the second question, assuming that it is content first and foremost that is being assessed. However, as we discuss later on in the chapter, many of the principles involved in assessing content can also be applied to the assessment of language, so even practitioners with different priorities should find the information useful.

Assessing content

Assessing content is potentially very challenging. Genesee and Upshur are clear:

Generally speaking, the same content objectives should be used to assess the achievement of second language and native speakers alike – lower standards of achievement should not be established for second language speakers.

(Genesee and Upshur, 1996: 47)

However, this is not necessarily easy to achieve – content may be understood by a learner, but she or he may not be able to express it sufficiently clearly if the language forms needed are not known, or if anxiety prevents it. Pinker summarizes:

Any particular thought in our head embraces a vast amount of information. But when it comes to communicating a thought to someone else, attention spans are short and mouths are slow.

(Pinker, 1994: 81)

A practical example would be if a learner were offered two parallel tables of statistics about two different countries being compared in a geography module. Inside the student's head, comparisons would be made instantly and a concept formed relating to this comparison/contrast. The essential knowledge intended to be gained would be gained. The learner's language competence would next determine whether this understanding could be communicated back to the teacher. If the student 'failed' to communicate understanding during the assessment process, then the teacher would not be sure whether this was due to limited language competence, or whether the student had really not understood.

We also need to define which aspect of the content we are assessing. We could be interested in any of the following:

- factual recall (detail)
- general understanding (major points)
- ability to manipulate the content, using higher-level thinking skills such as interpretation, analysis, synthesis or application. This will also reflect objectives regarding cognition (refer to Chapter 4 for some concrete examples), which are best assessed through content assessment, as without it they become simply abstract skills
- ability to research more independently and extend the topic knowledge beyond what has been presented by the teacher.

How should we assess?

While assessing simple detail may be uncomplicated, the other aspects in the list above are more complex for both teachers and learners. For this reason, when designing the

means of assessment, teachers should choose – whether assessing learners individually or in pairs/groups – the most direct method which uses the least language. Examples of this are that the learners should complete grids, draw diagrams or pictures, decide if bulleted statements are true or false, correct facts which are wrong, make simple presentations linked to visuals or answer content-based questions with a simple yes/no response. This point will be developed later in the chapter.

But the *how* of assessment also raises other issues. With the current strong focus on concrete objectives and purposeful learning activities which involve students in thinking and problem solving – sometimes in pairs or groups – come regular assessment opportunities, as long as alternative formats are accepted. Creating a specific ‘test’ may not be necessary if the activities themselves deserve monitoring and can provide concrete evidence of learning. So an ongoing approach to assessment in each lesson can become the norm, as Short’s (1993) article suggests. As well as observing learners at work on the tasks set by the teacher, if a three-part lesson structure is implemented (with a starter, main activities and a plenary) then the whole-class plenary can double as an opportunity to both monitor understanding and to re-teach the material for those who need to hear it (content or language) again. If the notion of assessment is truly formative, then the teacher wants to monitor the understanding at all the different levels – not to make a judgment on individuals, but to inform her or his own actions and future planning. So it is not a matter of ‘catching people out’, but of repairing misconceptions and filling gaps.

The **plenary** is the section of a lesson where the teacher and learners together summarize the learning up to that point in order to move on. This is often towards the end of the lesson.

Alongside this, and in accordance with the principles of AfL or its equivalent, is an understanding that assessment should not always be of individuals, but will sometimes be of groups of learners. Although it may be difficult to decide who has contributed what and who knows what, this is seen as less important, given that there are other gains to be made through collaborative work. The final output may be more than the sum of all the parts with more sophisticated use of language after group negotiation and editing. Research, divided between members of a group and then shared, can also contribute to this refinement. In addition, such tasks potentially raise different areas for assessment, such as teamwork, project management and capacity for self-assessment.

Who should assess?

The possibility of expanding assessment beyond the teacher looking solely at individual learners links partially to the question of *who* assesses. Clearly, teachers wish to retain the major role in this, but we can consider the following factors in establishing the possible range of teacher, self- and peer-assessment methods available:

- Clear success criteria enable learners to peer-assess or self-assess in certain kinds of tasks.

- Assessment can be collaborative within the whole-class setting if the teacher shows anonymous extracts from work and invites constructive amendments.
- Presentations can be assessed for a range of factors; for example, the communication of certain items of content, use of media, use of effects to scaffold understanding and contribution of members of a group.
- Self- and peer-assessment can be used as a platform to elicit comments about the learning process by asking why the judgments are as they are. This, when well established, can lead to insights into cognition, which is the most difficult *C* to assess.
- Cultural content can be something which learners feel adds interest and which can be peer-assessed through a more subjective system such as, for younger learners *three stars and a wish*, or an equivalent age-appropriate mechanism (this involves the assessor finding three aspects to praise and one to suggest for development).
- Peer-assessment can lead to better self-assessment. If a learner has formulated ideas about a piece of work sufficiently well to communicate and justify those judgments to another learner, she or he will be more able to look at her or his own work in the same objective manner.

The points above all demonstrate that relying on teacher assessment alone could impoverish a CLIL classroom. We will state again that a teacher will still be the main assessor, but there are numerous possibilities to vary this in appropriate circumstances. In considering how and where to add this variety, it is also necessary to weigh up how well learners can assess from a linguistic perspective: is their language capability sufficient to make valid judgments? Will a teacher need to re-assess everything? Collaborative assessment in a whole-class context managed by the teacher will always give an indication as to student capacity for the process.

Assessing content in the first language

We have so far avoided the notion of content assessment carried out in learners' (or the school's) first language. Some CLIL courses have built in the practice of addressing the second-language 'language barrier' issue by monitoring comprehension through a test given in the first language. We should note immediately that this becomes difficult or even impossible in classrooms with a wide variety of first languages and may actually disadvantage some learners if the majority language is assumed to be every learner's first language. But even in classrooms where all students share a first language, it can be problematic for both practical and pedagogical reasons. It can fail on a practical level when the specialist vocabulary needed for the content area is simply not known in the first language, because the topic has been taught through the CLIL language. This is yet more pronounced if the full subject is CLIL-taught for a year or more, as the first-language specialist terminology will be less related to current topics. On a philosophical and pedagogical level it can fail, because the intention of the CLIL programme is to build capacity to cope fully in an additional language, which includes finding strategies to communicate and developing

thinking as far as possible in that language. The proponents of this system will argue, of course, that the use of the first language still allows a deeper understanding to be communicated and that the practical problems can be overcome. The issue needs careful thought by those developing the programmes.

This issue does, however, become very difficult when we bring any nationally set testing into play. One of the pioneering schools which developed CLIL approaches in the UK stopped their programme a year away from national examinations, because the vehicular language was not accepted for testing. As the students were less confident with the subject matter in English, they had to carry out a revision programme in English in the lead up to the assessments so as to be able to reach their potential grades. A detailed report by Serra (2007) addresses many of the above issues, focusing carefully on what she calls 'language alternation' (also called 'translanguaging', mentioned in Chapter 2), specifically because of the need to manage first- and second-language capability in the content area (in this case, mathematics).

Assessing language

We have already mentioned the need for CLIL courses to seek parity with first-language programmes by using recognized local testing frameworks. There is clearly a case in language assessment for summative attainment at the end of courses to be stated in terms of levels in an internationally recognized system such as the Common European Framework of Reference for Languages (2001). The self-assessment level descriptors from B1 upwards (ibid.: 26–7) refer to elements of content which could encompass CLIL material. But in common with the rest of this chapter we want to look more at the earlier stages before programme assessment and to answer the question: *How do we assess language on an everyday basis?* To begin with, just as with content, we need to be sure which aspect of language competence we are assessing. It could be the ability to:

- recall subject-specific vocabulary
- operate functionally, using appropriate language structures and forms to discuss and disagree, ask effective questions, report in appropriate language structures, and so on
- listen or read for meaning
- present or discuss effectively
- demonstrate thinking/reasoning in the CLIL language
- show awareness of grammatical features of the language.

Teachers need to be clear both *why* they are assessing language as opposed to content and *how* they wish to do this. If we speak firstly about formative assessment of language, then we could mean ongoing correction in the classroom as well as assessment of written language in workbooks, or of the oral language of presentations after they have been completed. It could be argued that such language correction and assessment should be used specifically to improve the communication of content. If a student is told, as part of 'live correction', that changing the language in a certain way would make the content clearer,

then there is also a clear motive for that language assessment. If it is simply made as a correction of a detail of language accuracy, then it will inevitably halt the flow of content communication and could frustrate learners. **It is important to be clear that this does not mean we should ignore all errors and never assess language, but we can create specific opportunities to do this rather than offer continual corrective feedback which undermines content confidence.** The 'language clinic' is a potentially useful version of this practice: **from time to time, the teacher gathers language errors which need to be addressed as a class and holds a language clinic in a lesson, explaining to learners that this is a necessary step to support better communication of content.**

When looking at how to assess language, we should note that – as with content – language can be assessed through a variety of approaches. Brown and Hudson present the following as types of assessment:

... (a) selected-response (including true-false, matching, and multiple-choice assessments); (b) constructed-response (including fill-in, short-answer, and performance assessments); and (c) personal-response (including at least conference, portfolio, and self- and peer assessments).

(Brown and Hudson, 1998: 658)

This links back to the Short article in which she also lists assessment instruments which offer a better range of opportunities for CLIL students to demonstrate understanding:

... skill checklists and reading/writing inventories, anecdotal records and teacher observations, student self-evaluations, portfolios, performance-based tasks, essay writing, oral reports, and interviews.

(Short, 1993: 629)

In this article, Short was setting out a new view of assessment for bilingual teaching in America which did not relate to the existing English as a Second Language schemes. The emphasis on classroom processes which lies behind many of these methods is still not completely accepted across the world, but, as we have maintained so far, such methods are vital tools for teachers to gain a full understanding of student progress. In terms of continuous language assessment, the European Language Portfolio scheme (<http://www.coe.int/t/dg4/portfolio/>) offers a range of material developed in different countries which teachers may find useful, but at present this is not directly inclusive of a CLIL approach. We will next explore some assessment contexts in order to exemplify some of these tools.

6.2 Assessment in action: Examples of practice

In this section, the intention is to develop the threads opened up in the chapter so far, exploring rationale and methods of assessment, and to select assessment types which exemplify certain issues. This cannot be a full guide to CLIL assessment, as both the scope of different methods and the many different levels on which CLIL courses operate would make

that impossible. The points made here, however, should be transferable to related types of assessment and to levels of work and ages other than those directly referred to.

Sharing objectives and success criteria

Sharing the objectives and offering success criteria are important first steps towards effective assessment, as learners begin to find out in this way not just what they are likely to be learning, but also how their work will be assessed, both as they work and when they have completed it. It is important to use concrete statements in framing these intentions, not just because of the potential linguistic constraints contained in a CLIL context, but because this is good assessment practice. The older and more advanced learners are, the more complex this stage can be made, so that it remains cognitively appropriate. For example, the objectives / learning outcomes and the success criteria can be referenced more fully to previous knowledge if the linguistic knowledge can accommodate this. The primary-age structures of **WALT** (*we are learning to*) and **WILF** (*what I'm looking for*), comprising criteria outlining what the finished work will contain, sometimes personified into two cartoon characters, provide direction for making the statements concrete. These basic concepts can be adopted in a less 'primary' form for use with older learners. We may be addressing something as simple as: 'Today we are learning to see the differences between the landscapes of La Réunion and the Isle of Skye, so we can decide which pictures show which place'. Or we may be handling more advanced concepts such as: 'Building on last week's work on zonal soils and how Northwest Europe and a tropical environment such as La Réunion show differences, we are looking more closely at intrazonal soils and a feature called podsol in the tropical region. By the end of this week's work you will have a clear view of the soil characteristics of that area and why they might differ from local soils'. In both cases, students start the lesson knowing what they are going to learn, and in both cases the CLIL teacher will need to use some visual support to ensure that all learners follow the content of those learning intentions. Whether it is pictures of two environments, maps of locations, key vocabulary or diagrams, those statements are better supported by these visual elements than if they were just spoken. Success criteria can also be given for a piece of homework, such as the production of a presentation. The example given on page 122 ('Preparing a presentation') acts on several levels, clarifying content (as in the third bullet point), the presentation conventions and the quality expectations. The subject of this task was *Aspects of the weather systems in the Pacific Ocean*, so the checklist of points included explanation of the thermocline and the features of El Niño / La Niña. The set of bullet points here acts as an overall checklist for students when they have completed the task, making the assessment process more overt:

Example: Preparing a presentation*About the PowerPoint presentation:*

- There should be a title summarizing what you are explaining.
- There should be the names of the authors.
- There should be all the points of the outline I have given you.
- The explanations should be concise and clear.
- The drawings and/or diagrams should clarify the explanation.
- The presentation of the PowerPoint should be attractive and well organized.

Source: Roser Nebot (2008)

Link to worksheet [Accessed 27 April 09]: http://www.xtec.cat/cirel/pla_le/nottingham/roser_nebot/students.pdf

The grid in the example 'Drawing and painting a landscape' acts as a checklist for a final task, consolidating a unit. This is a good example of where success criteria refer to a non-linguistic outcome, but contain within them a reference to much of the key vocabulary of the unit, so checking comprehension and even language – if the piece of work matches all criteria, the teacher can be sure that the content and the language of the unit have been established. The language does not need to be produced for this process and therefore a discussion with the student about the finished painting would reveal her or his capacity to use the language effectively, but teachers can decide to what degree receptive and productive competence are desirable or required.

Example: Drawing and painting a landscape*During and after your work, check the following points:*

Draw the horizon line and add the vanishing point.	
Set the background and the foreground.	
Objects appear smaller as they get further away and with less detail.	
Overlapping tells us which object is in front, closer.	
Objects get higher on the foreground and closer to the horizon line.	
Warm colours advance and cool colours recede.	
Objects in the distance appear pale.	
Do not forget the way light and shadow create forms with colour and shading techniques.	

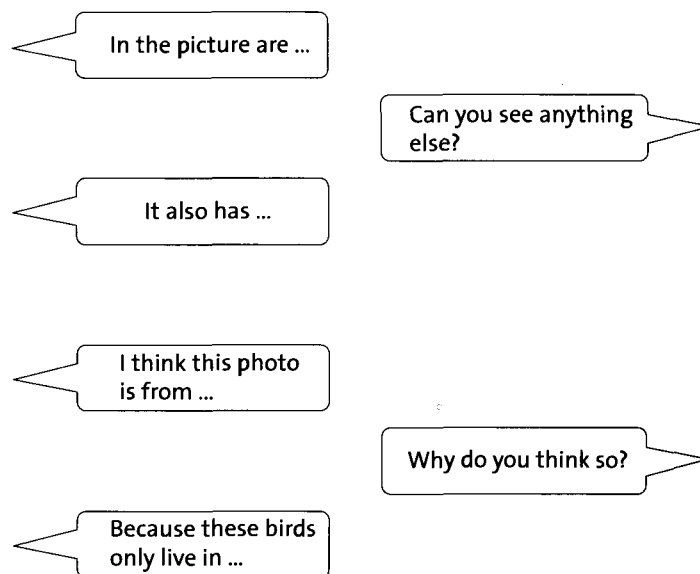
Source: Isabel Palomares Cots (2008)

Link to worksheet [Accessed 27 April 09]: http://www.xtec.cat/cirel/pla_le/nottingham/isabel_palomares/student.pdf

Alternative assessment formats

It is important to allow learners to express their responses to tasks in the most direct way possible so that language is not a barrier to demonstrating understanding of content. Simple assessment formats such as recording to a grid have several advantages. The format itself requires little language knowledge to stimulate content recall; it activates and organizes thinking to support maximum demonstration of knowledge, thus forming part of the process of working within a student's 'zone of proximal development' (Vygotsky, 1978), which for any individual learner will also involve dialogic interaction with the teacher and/or more able peers. It is therefore part of the formative structure.

In our first geographical example on page 121 (comparing the islands of La Réunion and Skye), learners might have a grid system with individual columns for each of ten photographs and rows, labelled simply with items which might be visible in the photographs, such as *a volcano, a sparrowhawk, a whiteye, the Cuillin Ridge*. Learners tick any items from the list that they see in each photo in turn. This establishes some specialist vocabulary knowledge demanded by the topic, and is at a basic level of comprehension. Once complete, the grid can be used for a further task involving pair work, in which learners produce a short, oral description of a photograph and then come to a decision about where it has been taken. At the simplest level, this may be between two locations, but a comparison of three environments (perhaps the two islands and the school locality for the younger children) makes it a more complex and more cognitively challenging task. The teacher can eavesdrop during this stage of the work to listen for correct location decisions and to evaluate language use beyond the single-word structure which might result from learners' referring to the grid. The language *for* learning (see Chapters 3 and 4) demands the fuller sentence structure which accompanies a description:



A grid checklist for more advanced work still performs the same function. It allows information, perhaps on a more complex level, to be assembled easily, with thought (rather than language) highlighted at that stage, and then for that assembled information to be used to stimulate language production once the concepts are securely in place.

In the assessment of content comprehension through *receptive* tasks, a major learning tool, and so also a major assessment tool, is reading. Naturally, at primary level, this has to be restricted and carefully planned, and may involve listening rather than reading, especially with the younger ages. But from late primary onwards, visual texts of all types (see Chapter 5) are an integral part of CLIL classrooms. Many task types involve simply reading – for example, matching pictures to vocabulary, ‘heads and tails’ sentence halves, true/false decision tasks, gap-fill where the missing items are given in a box, decision tasks where two versions are given and the correct one has to be chosen and, in more practical subjects, following instructions to create an outcome. Most writing tasks also begin with reading, as we will see later in this section.

Matching information

The assessment instrument which involves matching information, for example by ‘heads and tails’ (joining two halves of several definitions or sentences), also serves more than one purpose simultaneously. In this type of learning/assessment task, demonstrating comprehension should always involve real decisions based on concept understanding and not on other elements, such as linguistic forms. In the example ‘Identifying coordinates’ – a simple task at CLIL beginner level – the 11 target sentences often have the same sentence structure. This means that, when pairing the sentence halves, learners are faced with between two and six possible tail matches for each head, each of which would produce a structurally sound sentence. Only the simplest pair of sentences is open to a straightforward 50/50 choice (they being the first and the fifth sentences). Learners must therefore focus on meaning in order to match the correct tail to each head. The assessment is designed to be carried out in pairs, so offering the teacher another opportunity to listen to dialogue and assess to what extent learners’ understanding seems to be based on concept knowledge, as well as whether the learners have internalized the language needed to explain that understanding. Additionally, other elements can be evaluated, such as the pronunciation of key vocabulary. The intention signalled in the task rubric is for the pair work to be followed by a plenary, during which the rationalization of choices can be tested in open class discussion. For those who were less sure either of their choices or of the reason for their choices, this will offer another chance to consolidate learning.

Example: Identifying coordinates

Join the following heads with the correct tails (working in pairs, and later in a plenary):

The horizontal axis is called positive x and positive y coordinates.
The point $(-2, -3)$ is 2 units to the left, 3 units up.
The first quadrant contains all the points with the x-axis.
The fourth quadrant contains all the points with 2 units to the right, and 3 units up.
The vertical axis is called the y-axis.
The point $(2, 3)$ is negative x and positive y coordinates.
The point $(2, -3)$ is 2 units to the left, 3 units down.
The point $(-2, 3)$ is on the x-axis.
The second quadrant contains all the points with negative x and negative y coordinates.
The point $(2, 0)$ is 2 units to the left, 3 units down.
The point $(0, 2)$ is on the y-axis.

Source: M. Luz Esteve (2007)

Link to worksheet [Accessed 27 April 09]: http://www.xtec.cat/cirel/pla_le/nottingham/mluz_esteve/worksheet1.pdf

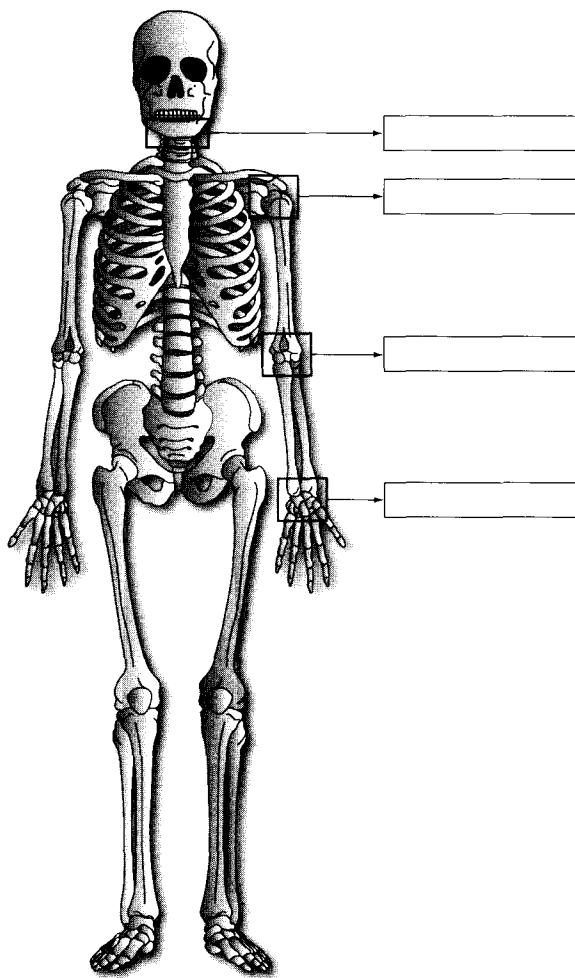
Productive assessment tasks

Those productive tasks which elicit content from students either orally or in written format are clearly the more difficult assessment instruments to structure, because they require not just recognition of key language, but also accurate memory for it. Students need not only to understand the topic, but to be able to use language in a way which communicates that understanding, and this will rarely be in single-word form (except in a simple labelling task). Often the material for labelling is given either in an accompanying text or in a box (and so it is another example of the reading-based tasks described in the previous section), but there will be times when the teacher will wish to establish whether the class has properly internalized the key vocabulary and the associated concepts. In this case, the teacher will use an open labelling task for this purpose, such as that shown in the example 'Labelling a diagram' on page 126.

Beyond such simple labelling, students' use of speaking/writing to express understanding needs to be scaffolded. With primary-age children, learners early in a secondary-level CLIL unit, or CLIL beginners in secondary education, this scaffolding is best achieved

Example: Labelling a diagram

Label the diagram with the joints:



Link to similar worksheets [Accessed 27 April 09]:

http://www.xtec.cat/cirel/pla_le/nottingham/francesc_niella/worksheets.pdf

by using a modelling approach. Diagrammatic structures are still the most useful ways of starting a writing process, as they require key vocabulary and an understanding of processes, but do not necessarily need connected text. A branched or statement key which uses yes/no questions to lead the reader to the correct definition of, for example, an animal, is an example of a real-purpose comprehension task which can also be used as a

model for the construction of a different key. A similar way of eliciting key vocabulary is to use a Venn diagram for classification with visuals as a source. By locating the items into separate or joint sections of the Venn diagram (which could consist of between two and five circles with a range of overlap possibilities), learners are demonstrating a conceptual understanding, but without the more complex language which a branched key requires. In this way, the essential descriptive or definitive terms can be tested along with the understanding of how they link and differ, without the need for other language which might divert attention. This works especially well as a group task, because it involves an initial brainstorming of relevant ideas, which should inevitably produce a more comprehensive outcome if shared by a number of students. This will then lead to a group reasoning process in order for decisions to be made about the placement of the assembled ideas onto the diagram. The need to state the reasons for the decisions 'out loud' supports the deeper concept comprehension of individuals and of the group collectively. However, for the production of longer, connected texts, a simple task brief which begins with instructions to *describe*, *explain* or – at a higher level – *justify* is rarely sufficient to elicit a response which will truly represent as full an understanding as learners may actually possess (except with more linguistically advanced students). Using a heard text as the model (such as a short clip of a documentary) is a more demanding bridging task which will ultimately allow learners to produce a fuller, richer text. This is because the task requires the information to be captured as it is spoken and in context rather than through multiple readings carried out at the student's own speed. 'Watching a documentary' gives an example of a heard-text bridging task.

Example: Watching a documentary

Watch the video and list the sources of CO₂ emissions that appear in it.

While listening, read the transcription of the video and complete the gaps.

Energy-dependent appliances are part of our modern way of life. Most of the energy they use comes from burning gas,, which emit carbon dioxide, CO₂, into the atmosphere, the planet's climate . . .

Source: J. Miquel Montesinos (2008)

Link to worksheet [Accessed 27 April 09]: http://www.xtec.cat/cirel/pla_le/nile/miquel_montesinos/students_worksheets.pdf

Another variant on this is a task which requires learners to take notes or fill in a diagram or grid, whilst listening to the teacher give a presentation which consolidates and synthesizes previously learned material from the unit. Shorter writing or speaking tasks are appropriate once the modelling is partially or wholly removed. The example task 'Thinking about a problem' scaffolds the language of conclusion but not the actual mathematical reasoning – this needs to come from the students, either individually or in groups.

Example: Thinking about a problem

It's impossible to fold a piece of paper more than eight times!

Sounds odd, doesn't it? What is the reason for that?

Try it yourself and try to answer. Think about the thickness of the paper, the number of layers and the mathematical rule.

I think that the reason for this is that . . .

I think it is impossible because . . .

This is due to . . .

Source: Imma Romero (2007)

Link to worksheet [Accessed 27 April 09]: http://www.xtec.cat/cirel/pla_le/nottingham/imma_romero/student.pdf

Science investigations offer opportunities for short pieces of writing or speaking from notes. Once the language of report has been established, the scaffolding can be at least partly withdrawn. This is a good example of the integration of teaching, learning and assessment, as there will be a series of stages involved in the whole process.

- 1 First of all, modelling or instruction-giving will set out the objectives and will establish the success criteria (not for the investigation, but for the reporting of it).
- 2 During this process, scaffolding will take place as the teacher circulates and encourages pairs or groups to discuss what they are doing. At this stage, the teacher will sample individuals' and groups' understanding of the concepts behind the investigative work, as well as their ability to see *what* is happening and *why* as the investigation proceeds.
- 3 The teacher will also become aware during the modelling stage of any really specific language needs which might prevent accurate and full reporting of the investigation.
- 4 As an assessment opportunity, the reporting stage will be divided into two sections. Firstly, the pairs/groups will create the report using peer scaffolding. Individuals will write this formally or make notes for an oral report. The teacher will then either see the written reports and assess them or will listen to oral reports and offer feedback. In either case, the assessment will still be formative and so form part of the ongoing teaching and learning process.

6.3 Peer- and self-assessment

We should lastly explore the subject of peer- and self-assessment, which has been alluded to throughout the chapter. It was noted earlier that there needs to be a close link to success criteria for this to be effective, and that quality and accuracy of expression will not

be included in these judgments, except for the most advanced and able learners (although clarity can certainly feature in them). There are numerous reasons for using peer- and self-assessment in the CLIL classroom. From a long-term perspective, we can assert that learners who understand what they are learning, as well as how to demonstrate high-quality understanding, will make greater progress than they might otherwise do if 'kept in the dark'. Black and Wiliam make these two comments:

[S]elf-assessment by pupils, far from being a luxury, is in fact an essential component of formative assessment. When anyone is trying to learn, feedback about the effort has three elements: recognition of the desired goal, evidence about present position, and some understanding of a way to close the gap between the two.

(Black and Wiliam, 1998: 4)

Peer-assessment which refers to specific criteria and is carried out in discussion between two partners in a class is valuable, because it centres on a process where each student puts into words – and therefore also rehearses – their individual understanding of the topic material. Negotiation takes place and a finer understanding of that material by both parties is possible as a result. This can also be modelled before being completely handed over to students. Language or content 'clinics', as suggested earlier, provide an opportunity for whole-class discussion of issues concerning aspects of the CLIL programme, in which the use of success criteria can be properly explained and demonstrated, and also a model for positive and constructive statements can be given. Self-assessment and self-evaluation are both likely to be better informed if they follow peer-assessment, meaning that target-setting will subsequently also be more relevant. Peer-assessment can also be a larger-scale exercise, including the whole class listening to presentations by other groups and 'marking' them all with reference to a set of criteria. We include on page 130 an example of criteria included in a peer-assessment grid used to assess a PowerPoint presentation (Figure 10).

6.4 Summary of assessment principles

This chapter has attempted to provide a discussion of issues in and potential approaches to the difficult question of assessment in CLIL. It cannot of course do justice to the enormous range of possible differences between contexts, but it has taken a philosophical line which we hope is coherent. We conclude with a set of summary principles which we feel have underpinned the discussion throughout, and which, echoing Short's (1993) plea, advocate alternative assessment methods:

- Clear learning objectives are needed before an assessment focus can be chosen. Learning objectives/outcomes should use a format which acknowledges the different areas of learning in the classroom (such as the 4Cs approach) – this will usually include content/skills first, then language in some form. In a CLIL classroom there are likely to be more possible angles of assessment at **any one** point because of the integrative nature of content and language. **Therefore, even** more than in first-language lessons, we cannot always assess **everything**.

Figure 10: A grid for peer-assessment

PowerPoint	Beginning 1	Developing 2	Accomplished 3	Excellent 4
General aspects of slides	Disorganized and difficult to follow	Organized but difficult to follow	Disorganized but easy to follow	Organized and easy to follow
Pictures and graphics	Small and impossible to understand	Big but difficult to understand	Small but easy to understand	Big and easy to understand
Texts	Small and impossible to understand	Big but difficult to understand	Small but easy to understand	Big and easy to understand
Content	Does not cover all appropriate topics	Covers some of the appropriate topics	Covers most of the appropriate topics	All topics covered. Also interesting facts
Speech	Beginning 1	Developing 2	Accomplished 3	Excellent 4
Matching between speech and images	Speech has nothing to do with slides	Speech is substantially different from slides	Only a few items of the speech are not reflected in the slides	Speech and slides match perfectly
Language	Many pronunciation and grammatical errors	A few errors	Only one or two errors	Pronunciation and grammar are perfect
Communication	The speech is read all the time	The speech is read most of the time	The speech is read sometimes	The speech is not read
Timing between team members	Only one member speaks	One member speaks most of the time	One member speaks more than the other	The two members share speech equally

(Adapted from Alberich, 2007)

- We should use a mixture of formal and informal assessment which is both task-based and assignment-based, and a mix of specific test times and classwork sampling.
- We should familiarize the learners with the assessment measures and success criteria, expressed in a student-friendly format.
- Content knowledge should be assessed using the simplest form of language which is appropriate for that purpose.

Excellent
4
Organized and easy to follow
and easy to understand
and easy to understand
Topics covered. Also interesting
Excellent
4
and match
stly
Association grammar
effect
tech is
id
rs share
equally
rich, 2007)
h
sswork
ess
e

- Language should be assessed for a real purpose in a real context – sometimes this will be for form/accuracy, sometimes for communicative competence and/or fluency.
- If the assessment is orally based, 'wait time' is crucial, as in CLIL contexts we should be asking students to think, and thinking takes time and the expression of that thinking takes longer.
- Scaffolding is not 'cheating' – we need to assess what students can do with support before we assess what they can do without it.
- Students need to be able to take some responsibility for their own assessment, both in terms of self- and peer-assessment. This will enhance their longer-term learning potential.

References

- Alberich, J. (2007) *English through Science* [Online lesson plans and worksheets]. Available at: http://www.xtec.cat/cirel/pla_le/nottingham/joan_alberich/index.htm [Accessed 29 April 09].
- Ames, C. and Ames, R. (1984) 'Systems of student and teacher motivation: Toward a qualitative definition', *Journal of Educational Psychology*, **76**, 4, 535–56.
- Assessment Reform Group (2002) *Assessment for Learning: 10 Principles. Research-based principles to guide classroom practice* [Online]. Available at: http://www.qca.org.uk/libraryAssets/media/4031_afl_principles.pdf [Accessed 27 April 2009].
- Black, P. and Wiliam, D. (1998) 'Inside the black box: Raising standards through classroom assessment', *Phi Delta Kappa*, [Online] **80**, 2. Available at: <http://www.pdkintl.org/kappan/kbla9810.htm> [Accessed 28 April 2009].
- Bloom, B. S. (1968) *Learning for Mastery. The Evaluation Comment*, Los Angeles: University of California.
- Brown, J. D. and Hudson, T. (1998) 'The Alternatives in Language Assessment', *TESOL Quarterly*, **32**, 4, 653–75.
- Clarke, S. (2001) *Unlocking Formative Assessment: Practical Strategies for Enhancing Pupils' Learning in the Primary Classroom*, London: Hodder and Stoughton.
- Cohen, A. (1994) *Assessing Language Ability in the Classroom* (Second edition), Boston: Newbury House/Heinle and Heinle.
- Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (2001) Cambridge: Cambridge University Press.
- Dweck, C. S. (1986) 'Motivational processes affecting learning', *American Psychologist*, **41**, 1040–8.
- Genesee, F. and Upshur, J. A. (1996) *Classroom-Based Evaluation in Second Language Education*, Cambridge: Cambridge University Press.
- Pinker, S. (1994) *The Language Instinct: How the Mind Creates Language*, New York: Harper Collins.
- Sadler, D. R. (1989) 'Formative assessment and the design of instructional systems', *Instructional Science*, **18**, 119–44.
- Scriven, M. (1967) 'The methodology of evaluation', in Stake, R. E. (ed.) (1967) *Curriculum Evaluation* (American Educational Research Association Monograph Series on Evaluation, 1), Chicago: Rand McNally.

- Serra, C. (2007) 'Assessing CLIL at primary school: A longitudinal study', *International Journal of Bilingual Education and Bilingualism*, 10, 5, 582–602.
- Short, D. (1993) 'Assessing integrated language and content instruction', *TESOL Quarterly*, 27, 4, 627–56.
- Vygotsky, L. (1978) *Mind in Society*, Cambridge, MA: Harvard University Press.
- Zangl, R. (2000) 'Monitoring language skills in Austrian primary (elementary) schools: A case study', *Language Testing*, 17, 250–60.