

CEFR Teacher Assessment

A guide for Eaquals Members on implementing CEFR-referenced teacher assessment

Brian North April 2009



Contents

Introduction	Page 2
Standardisation:	
• Training in performance expected at the CEFR levels	Page 3
Teacher Assessment:	
• Continuous Assessment? Guiding teacher impression	Page 5
• Assessment Tasks: Evaluating speaking and writing	Page 9
Moderation:	
• Techniques to limit teacher subjectivity	Page 18
Case Studies:	
• Eurocentres	Page 20



Introduction

This guide supports teacher assessment of CEFR level in a CEFR-related curriculum. For developing such a curriculum, readers are referred to *Eaquals Self-Help Guide on CEFR Curriculum Development*, and the set of Curriculum Case Studies associated with it.

Even when teachers are operating a CEFR Curriculum and have a common understanding of the content of the different levels, it is necessary to carry out standardisation training to show them what type of performance is typical at different levels.

Even after such training, moderation techniques are necessary to adjust for excessive severity/lenience by any individual teacher so to guarantee comparability. Such techniques may include the use of progress test results to confirm teacher impressions about the position of individuals in the group, but when good assessment checklists and grids based on CEFR descriptors are used by well-trained teachers, then the use of formal tests is not essential. Developing progress tests is the subject of a guide parallel to this paper.

This guide does not seek to cover all aspects of teacher assessment. CEFR Chapter 9 discusses assessment from many different perspectives, many of whom have little to do with stating what level a learner is. The most important aspects of teacher assessment in fact concern issues like the evaluation of successful learning at the end of a lesson, informal quizzes to reinforce memory and to motivate revision, and the use of assessment to raise learners' language awareness and accuracy. These aspects of assessment are taken into consideration in an *Eaquals* inspection, but they are not directly related to the certification of a particular CEFR level and are therefore not discussed in this guide.

The guide is therefore organised in four sections:

- Standardisation: a short review with reference to *Eaquals* standardisation packs and other materials and reference points available
- Continuous Assessment: judgement by the teacher on the basis of course work, guided by descriptors
- Assessment Tasks: suitable tasks and assessment grids
- Moderation: techniques to limit the subjectivity of judgements

Tips

There are no golden rules but the following axioms may be useful:

- Start from your curriculum.
- Keep it simple.
- Insist on the CEFR descriptors and DVDs as the sole point of reference in disagreements
- Require teachers to assess a specific performance to confirm their general impression
- Ensure everybody understands and follows simple, defined procedures
- Implement systematic quality control as part of academic management



Standardisation: Expected performance at the CEFR levels

Standardisation training with videos is the most effective way to achieve a common interpretation of the CEF levels. In introducing standardisation, it is important to place the training exercise in context. The CEFR has not emerged from nowhere and the levels are already operationalised in examinations that the teachers will be familiar with; they are not starting from scratch. The next page gives a chart showing the relationship of ALTE examinations offered for the major European languages.

Nevertheless, standardisation training is necessary for three reasons:

- a) People often think they "know" the CEFR levels without having looked at either the CEFR descriptors that define the levels or samples that illustrate them.
- b) People often make judgements based on criteria they are unaware of. In particular, some teachers have an exaggerated idea of the level of accuracy (or fluency) it is reasonable to expect. Using shared, defined criteria is a way of addressing this problem.
- c) People can interpret the written word (the descriptors) in different ways; some people are just stricter than others. Teachers do not realise this; they naturally think that they and their colleagues share the same interpretation.

Discussing concrete examples of performances in relation to common criteria, supported by detailed documentation that explains why a performance is one particular level, is a very effective and enjoyable way of counter-acting these problems.

The Standardisation pack is available for French and German as well as English. It is vital to download the documentation explaining the levels of the illustrative samples. It is logical to do standardisation training with videos before using scripts because everyone can watch and then discuss the same video performance. DVDs of illustrative samples for standardisation training can be obtained either through the Equals Office or direct from the Council of Europe. The DVD for German "Mündlich" can only be purchased from language bookshops.

Standardisation training may need to be carried out in a series of workshops. Normally the following 3 steps would be followed:

- a) Illustration with documented illustrative samples (DVD, scripts): "This is level B1, because" Here one should hand out documentation on the samples concerned as part of this process.
- b) Small group discussion of other documented illustrative samples: "What level do you think this is?" Here the group receive the documentation after they have made their decision - and compare.
- c) Individual rating of documented samples, followed by small group, followed by plenary. In the plenary at the end, the workshop animateur could read out part of the comments in the documentation.

One might transfer the experience gained in order to benchmark performance samples from the school concerned, thus creating local calibrated CEFR samples for future reference.



ALTE Examinations and the CEFR

	A1	A2	B1	B2	C1	C2
Deutsch	Start Deutsch 1	Start Deutsch 2	Zertifikat Deutsch (ZD)	B2	C1	Zentrale Oberstufenprüfung (ZOP) Kleines Deutsches Sprachdiplom
English		Key English Test (KET)	Preliminary English Test (PET)	First Certificate in English (FCE)	Certificate in Advanced English (CAE)	Certificate of Proficiency in English (CPE)
Español		-	Diploma de Español (Nivel Inicial)	Diploma de Español (Nivel Intermedio)	-	Diploma de Español (Nivel Superior)
Français	Diplôme d'Études en Langue Française DELF A1	Diplôme d'Études en Langue Française DELF A2	Diplôme d'Études en Langue Française DELF B1	Diplôme d'Études en Langue Française DELF B2	Diplôme Approfondi de Langue Française DALF C1	Diplôme Approfondi de Langue Française DALF C2
Italiano		Certificato di Conoscenza della Lingua Italiana, Livello 1 (CELI 1)	Certificato di Conoscenza della Lingua Italiana, Livello 2 (CELI 2)	Certificato di Conoscenza della Lingua Italiana, Livello 3 (CELI 3)	Certificato di Conoscenza della Lingua Italiana, Livello 4 (CELI 4)	Certificato di Conoscenza della Lingua Italiana, Livello 5 (CELI 5)
Português		Certificado Inicial de Português Língua Estrangeira (CIPEL)	Diploma Elementar de Português Língua Estrangeira (DEPLE)	Diploma Intermediário de Português Língua Estrangeira (DIPEL)	Diploma Avançado de Português Língua Estrangeira (DAPLE)	Diploma Universitário de Português Língua Estrangeira (DUPEL)



Continuous Assessment: Guided teacher impression of curriculum achievement

There are two different ways of thinking about teacher assessment in relation to the CEFR.

- Assessment on the basis of the teacher's impression of the learners achievement based upon their contribution in class, homework, test scores etc. This is dealt with in this section.
- Assessment of speaking and writing by the teacher in relation to performance on a specific assessment task. This is dealt with in the next section.

The two approaches can of course be combined and supported by the kind of moderation techniques to ensure consistency.

As regards the teacher impression judgements themselves, there are basically two approaches:

- With descriptor scales for each skill (one descriptor per level) which each cover all levels; the teacher makes a judgement as to the learners level on each skill
- With a checklist for the level concerned, which probably has sub-sections for each skill (several descriptors for each skill), and possibly for language points covered, but which is confined to the relevant curriculum level.

1. Scales

This approach is very simple. In effect teachers are judging at the end of a course the learner's achievement of the "global objectives" for the level the class is aiming at. Towards the end of a B2 class, most students will probably be getting "B2" in most skills, but some will have B2+ / B1+. Some might even have a full level above in a certain skill (e.g. C1 in reading) or a full level below in a certain skill (e.g. B1 in writing). Teachers hand in their grades on a "grade sheet" and the Academic Manager compares the grades given to different classes to check that they make sense.

Advantage: Simple, quick and non-bureaucratic.

Disadvantage: Very unreliable unless supported by assessment tasks and/or an anchor test.

2 Checklists

Checklists are used for genuine continuous assessment: checking off what the learner can do at certain stages of the course.

The main points to make about checklists are that:

- (a) the students have to be with the teachers for a long time for the teacher to have the necessary detailed knowledge of the students;
- (b) they can include language points as well as communicative "Can Dos;"
- (c) they should not be long. Experience at Bell Krakow suggests that teachers are overwhelmed when asked to assess 15-20 students on circa 35 descriptors. Completing a single checklist in relation to the content covered by the class is one thing; completing a checklist showing individual



achievement for each member of the class is quite another. This suggests the checklist should perhaps be restricted to "Can-Do" descriptors for real world communicative tasks.

Advantage: Direct link to the syllabus; encourages students to understand that they must accumulate competences; can therefore be used for self-assessment.

Disadvantage: Teachers may lack the detailed knowledge of the performance of individual students on individual descriptors.

Examples (Level A2): IH La Spezia; IS-Aix en Provence; Bell Krakow. The IH La Spezia example has 27 items including language points and is used mainly by teachers. The Is-Aix and Bell Krakow ones have 12 items, and are both also used for self-assessment as well as teacher assessment. Generally speaking, when the learner "can do" 70-80% of the descriptors they are considered to have achieved the level.

Teacher Report Form: IH La Spezia: Level A2

	Giovani Bianchi	Luisa Verdi
I can talk about a friend or family member	1 2 3 4 5	1 2 3 4 5
I can describe a picture or photo	1 2 3 4 5	1 2 3 4 5
I can describe a picture or photo	1 2 3 4 5	1 2 3 4 5
I can use other words to explain a word I don't know	1 2 3 4 5	1 2 3 4 5
I can write an introductory email about myself	1 2 3 4 5	1 2 3 4 5
I can talk about a past holiday	1 2 3 4 5	1 2 3 4 5
I can talk about the music I like	1 2 3 4 5	1 2 3 4 5
I can read and understand a simple story	1 2 3 4 5	1 2 3 4 5
I can deal with basic hotel situations and problems	1 2 3 4 5	1 2 3 4 5
I can talk about future arrangements and intentions	1 2 3 4 5	1 2 3 4 5
I can make predictions about the future	1 2 3 4 5	1 2 3 4 5
I can make promises, offers and decisions	1 2 3 4 5	1 2 3 4 5
I can deal with problems in a restaurant	1 2 3 4 5	1 2 3 4 5
I can write an informal letter	1 2 3 4 5	1 2 3 4 5
I can talk about past experiences I have had	1 2 3 4 5	1 2 3 4 5
Present Simple	1 2 3 4 5	1 2 3 4 5
Present Continuous	1 2 3 4 5	1 2 3 4 5
Defining relative clauses	1 2 3 4 5	1 2 3 4 5
Past Simple	1 2 3 4 5	1 2 3 4 5
Past Continuous	1 2 3 4 5	1 2 3 4 5
Subject/ Object questions	1 2 3 4 5	1 2 3 4 5
So, but , because, although	1 2 3 4 5	1 2 3 4 5
Present Continuous for future arrangements	1 2 3 4 5	1 2 3 4 5
Will and won't	1 2 3 4 5	1 2 3 4 5
Present Perfect	1 2 3 4 5	1 2 3 4 5
SimpleComparative adjectives	1 2 3 4 5	1 2 3 4 5
Superlative adjectives	1 2 3 4 5	1 2 3 4 5

Is-Aix en Provence: Auto-evaluation Niveau A2

OBJECTIFS

AUTO- EVALUATION

1) Comprendre / Ecouter:

Je peux:

- Comprendre des exposés spécialisés, complexes.
- Comprendre tous les locuteurs natifs sans d'efforts.(accents, terminologies, régionalismes)

Oui	Un peu	Non

2) Comprendre / Lire :

Je peux :

- Comprendre tous les types de textes longs, abstraits, complexes ou riches en expression familières
- Distinguer le style et le sens explicite ou implicite.

Oui	Un peu	Non

3) Parler / Communiquer :

Je peux :

- Communiquer par un discours élaboré, limpide et fluide.
- Communiquer de façon logique et structurée en m'adaptant aux auditeurs.
- Gérer tout type de questionnement.

Oui	Un peu	Non

4) Parler / Exprimer :

Je peux :

- Faire des descriptions claires et détaillées sur des sujets complexes.
- Reformuler, restructurer, revenir sur une difficulté sans interrompre la communication

Oui	Un peu	Non

5) Ecrire :

Je peux :

- Produire des textes élaborés, fluides, et claires.
- Utiliser des structures logiques qui aide le destinataire à remarquer les points importants.
- Adapter mon style au destinataire.

Oui	Un peu	Non

CEFR LEVEL: A2

BELL

**KRAKOW LEVEL: 3
COMMUNICATIVE OBJECTIVES
2008/2009**

STUDENT:.....

GROUP:.....

Indicate the student's performance regarding the following partial competences by ticking the appropriate column: poor (P), good (G), excellent (Ex)	P	G	Ex
Listening and Speaking*			
understand what is said clearly, slowly and directly in simple everyday conversations			
participate in short social exchanges, introduce him/herself, get information about travel, order something to drink or eat, make and respond to invitations, etc.			
Reading*			
understand simple written messages from friends or colleagues			
understand simple user's instructions for equipment			
understand a simple personal letter in which the writer tells or asks about aspects of everyday life			
Writing*			
write short, simple notes and messages			
fill in a questionnaire about his/her educational background, job, interests and specific skills			
briefly introduce him/herself in a letter with simple phrases and sentences (family, job, hobbies)			
write simple sentences, connecting them with words such as "and", "but", "because"			
Grammar and Vocabulary*			
produce simple grammatical structures that have been learnt and practised in class			
describe past activities and personal experiences (e.g. the last weekend, his/her last holiday)			



refer to future plans			
-----------------------	--	--	--

**The European Language Portfolio, CODN Warszawa: accredited model No.6.2000, <http://culture2.coe.int/portfoio>*

Assessment Tasks

Teacher assessment can be made more accurate if judgements are made in relation to performance in a specific assessment task. One written homework – or a task set in class – is consciously assessed; one particular integrated skills or spoken communicative activity is used to observe and assess students' performances. Introducing such a set task requires the teacher to actually observe the work of the individual, rather than just assuming – on the basis of their impression of the student's contribution in class – that they know enough about the student to make a judgement.

It can happen that a particular student then has an "off day" doing the assessment task, performing surprisingly badly as a result. There are any number of reasons why this might happen. Unlike in an examination situation, a class teacher is in a position to spot this as an atypical performance and to decide to record the results as they occurred, but to make allowances when actually giving the student their grades. This is not so different from the procedure that most examination bodies have for spotting when an examination group is disadvantaged for some reason (Cambridge screen data for this). But even the cleverest examination bodies can only spot it with a group of scores; a trained teacher can spot it with an individual.

An alternative to the above is to weight results from continuous assessment 50:50 with results from the assessment task(s).

Evaluating performance in an assessment task involves two issues:

- eliciting a sample
- judging how good it is.

1. Eliciting an adequate spoken or written sample

The performance the student gives a limited sample. One wishes to make valid generalisations about the learner's overall competence on the basis of that sample. No amount of sensitive and well-trained judgement can give an accurate result if the sample is unrepresentative. In an examination situation, when nothing is known about candidates in advance, great care must therefore be taken with the design of tasks to ensure that they elicit different kinds of discourse that are representative of the learners full repertoire of skills. It is true that in teacher assessment this is less crucial – because if the assessment goes wrong one can compensate with the detailed knowledge of the class as discussed above. However, it is still very important.

In practice it is not always feasible to set two different pieces of written work. That is all the more reason to ensure that the task set is an appropriate one, related to one or more CEFR descriptors for the level, and challenging enough in relation to the level.

With spoken tasks it is more necessary to consider seriously the structure of the task in order to get a sufficient spoken sample from every student. It is no good just having a "class discussion;" some people will not speak. Any assessment of spoken proficiency requires a task that has phases which elicit different types of discourse; there is not so much point in eliciting 90 minutes of exactly the same thing.

Some key variables to take into consideration are:

- production or interaction;



- prepared talk or spontaneous talk;
- colleague as interlocutor or teacher as interlocutor.
- given themes or chosen themes;
- descriptive (reporting) or evaluative (stating opinion and arguing);

a) One simple approach is just to give learners a platform to talk, guided by prompt cards. This method was used successfully in the production of the calibrated samples produced for the Council of Europe. Learners are put in pairs. A prepared production phase is given by each candidate on a chosen theme and the other candidate(s) ask questions at the end. After the last speaker's production phase, the pair draw cards with discussion topics. They are allowed to discard themes they don't like and to go on to a new theme when they have nothing more to say, so again have choice of theme. The interaction phase is 100% spontaneous. The entire activity takes 12-15 minutes.

This style of elicitation leaves the learners a great deal of autonomy to show their best. It is very simple to set up in classroom contexts, it avoids the complications of the teacher/examiner being a part of the performance they are evaluating, and it combined the philosophies of "bias for best" and learner autonomy inherent in the Portfolio.

b) Eurocentres use a range of types of small group activities that conform to a loose template¹. What the tasks have in common is that they provide three phases, each generating different kinds of discourse: (a) a Collaborative Phase working out what to do (short, slow turns with high use of communication strategies); (b) an Exchange Phase in which each student has a chance to take the floor (long, coherent turns which are semi-prepared), and (c) a Discussion Phase (spontaneous, short turns). Such activity phases cover the CEF spoken interaction categories "Goal-oriented Co-operation", "Information Exchange" and "Informal Discussion," with the potential for the spoken production categories "Sustained Monologue: Describing experience" and "Sustained Monologue: putting a case," depending on the level of the learners and the precise nature of the activity.

c) Cambridge ESOL focus on defining tasks for phases that elicit different types of discourse that can be explained easily to candidates, that can be mass produced and carried out in a standardised fashion. The result is slightly reminiscent of a TV quiz show: very clear rules, an authenticity created by the very acceptance of the unnaturalness of the situation and a lot of talk from the "moderator."

Tasks from the Cambridge orals lend themselves well to adaptation for classroom group work similar to the Eurocentres approach described above. Eurocentres London Central has considerable experience of this.

d) The focus in the CIEP's DALF examination for French is on the use of one or more authentic texts as a springboard for discussion. This is a good approach for intermediate and advanced learners. The students read the text(s). Then they (a) summarise the main points made, glossing that report with their opinion at higher levels, (b) answer follow up questions from an interlocutor and then (c) engage in a discussion with the examiner on the subject.

¹ North, B. (1991): Standardisation of continuous assessment grades. In Alderson, J. C. and North, B. (eds.): Language Testing in the 1990s: Modern English Publications/British Council, London, Macmillan: 167-177.

North, B. 1993: L'évaluation collective dans les Eurocentres. In Evaluations et Certifications en Langue Etrangère, numéro spécial, *Le Français dans le Monde - Recherches et Applications*, août-septembre 1993: 69-81.



Again such an approach can be easily adapted to group work in the classroom. Each group receives a different text – on the same subject. They read it and discuss it (collaboration phase); then the groups are reformed and each person in turn summarises their text, taking follow up questions (exchange phase). Finally there is a discussion prompted by one or more questions, as in “The Story of a Marriage” (discussion phase).

Group work or interviews?

There are arguments for and against the use of an examiner/interlocutor in interviews.

For interviews:

- Unless the classroom task is well selected, the resulting discourse may be too “chatty” to show what the learner is capable of.
- In the classroom task, lots of people talking at once. In structured activities, the teacher can focus on students doing their “party piece,” but it is nonetheless difficult. It has nevertheless worked well in Eurocentres for 20 years.
- Learners will take an interview more seriously

Against interviews:

- Most teachers talk far too much when giving interviews - often up to 80% of the joint talk with “normal” students; even trained examiners tend to talk 50% of the time.
- Teachers find it very difficult to avoid just asking questions. This prevents students having any initiative, and it is often self-perpetuating as the learner starts to give only short answers.
- It is very difficult to consult criteria whilst conducting an interview. Interviewers therefore tend to form their judgement during the interview based on their internalised memory of what the criteria are, or even stop using the criteria altogether.
- The rest of the class are not “being taught” while an interview is going on; this means either that they are taught separately, which is expensive, or that the class is occupied with something to keep them quiet (e.g. the written test). In the latter case, however, some students may complain about interviewing being done in class time.

Tips for interviewing:

- The interviews can be carried out by the teacher outside the classroom door while the other students are completing a written test or writing task.
- In interviews it is crucial to say “Tell me about X” rather than asking questions. The DALF approach with the text as springboard has the advantage that the learner automatically has the initiative and it is noticeable that DALF examiners speak far less than in a conventional (e.g. Cambridge) interview. Goethe Institute examinations combine both approaches with a production phase (more like DALF) and an interaction phase (more like Cambridge).
- One way to structure an opportunity to look consciously at the criteria during the interview is to ensure that the interview has distinct phases. One stops the interaction at the end of each phase by giving the student(s) something to read or to look at in preparation for the next phase. In the breathing space provided, the teacher can consult the criteria, make notes and assign grades.

2. Awarding a level or mark

There are essentially two ways of judging the performance:

- awarding one level (e.g. B1 + - or even B1 ++) rather than another (e.g. B2, B2+)
- giving a mark (e.g. 17/20) in relation to what was expected at the level concerned.

The two approaches can also be combined into a third, hybrid approach.



Awarding levels/marks without specific reference to defined criteria is less reliable. Teachers need to be constantly reminded what the criteria for their decisions are – and the best way to ensure this is through regular use of the criteria in the assessment procedure.

The CEFR itself is a source of descriptors that can be adopted or adapted as criteria. As a short cut, the Equals Standardisation pack provides CEFR Table 3: Qualitative aspects of performance. The descriptors for aspects of quality in the November 2008 Equals descriptor bank (given at the end of this section) may also be useful.

2.1. Directly awarding a level

This is “holistic assessment:” the performance is compared to a grid of descriptors – for speaking, perhaps 3 categories (e.g. Range, Accuracy, Spoken Fluency) defined at 4 levels (e.g. A1 + A2 A2+ B1 B1 +) as below.

	RANGE & PRECISION	ACCURACY	FLUENCY
B1+	Can describe unusual situations and to express thoughts on abstract or cultural topics (such as music, films). Can explain the main points relating to an idea, problem, or argument with reasonable precision.	Can communicate with reasonable accuracy in familiar contexts, though with noticeable mother tongue influences.	Can express self relatively easily when talking freely and keep the conversation going effectively without help, despite occasional pauses to plan and correct.
B1	Can talk about family, hobbies and interests, work, travel, news and current events. Can make the other person understand the most important points.	Can express self reasonably accurately in familiar, predictable situations.	Can keep a conversation going, but sometimes has to pause to plan and correct.
A2+	Can talk about familiar everyday situations and topics, with searching for the words; sometimes has to simplify.	Can use some simple structures correctly in common everyday situations.	Can participate in a longer conversation about familiar topics, but often needs to stop and think or start again in a different way
A2	Can communicate in a simple and direct exchange of limited information in everyday situations; otherwise has to compromise the message.	Can use correctly simple phrases learnt for specific situations, but often makes basic mistakes – for example mixing up tenses and forgetting to use the right endings.	Can make self understood with short, simple phrases, but often need to stop, try with different words – or repeat more clearly what was said.
A1+	Can talk about self, family and job in a simple and direct exchange in common everyday situations.	Can use correctly some simple memorized structures.	Can speak slowly in a series of very short phrases, stopping and starting as he/she tries to say different words.

Essentially teachers assess their class exactly the same as they assessed samples in CEFR standardisation training – using a criteria grid. It may be exactly the same grid covering all levels (as in Eurocentres); it may cover only a range of levels as in the example above.

This criteria grid would be used with a separate marking grid like the example below. The teacher enters a level for each student. Here, it is important to standardise what notation is acceptable (e.g. B1 ++ is okay, but B2- is not).

Candidate	RANGE	ACCURACY	FLUENCY	OVERALL
Marisa	B1	A2	A2+	A2+
Angeles	B1	B1+	A2+	B1
Cristina	A2	A2+	A2+	A2+
Yolanda	A2	A1+	A2	A2
Albert	A2	A2	A2+	A2

2.2. Hybrid approaches

An alternative approach would focus more closely on the range of level in the class, awarding marks based on



fulfilling descriptors. The grid now shown below is a variant on the previous "holistic" one. A2+ is now set as the "standard;" what is expected from the class. It is therefore given "3" marks per category. The maximum score is 15 (5 x 3). The "pass mark" is 8 (A2+ for two categories; A2 for one). Strong students will get a "4" in some categories, exceptional students, who the teacher thinks are achieving "B1" in this respect, a "5".

	RANGE & PRECISION	ACCURACY	FLUENCY
5 (B1)	Can talk about family, hobbies and interests, work, travel, news and current events. Can make the other person understand the most important points.	Can express self reasonably accurately in familiar, predictable situations.	Can keep a conversation going, but sometimes has to pause to plan and correct.
4			
3 (A2+)	Can talk about familiar everyday situations and topics, with searching for the words; sometimes has to simplify.	Can use some simple structures correctly in common everyday situations.	Can participate in a longer conversation about familiar topics, but often needs to stop and think or start again in a different way
2			
1 (A2)	Can communicate in a simple and direct exchange of limited information in everyday situations; otherwise has to compromise the message.	Can use correctly simple phrases learnt for specific situations, but often makes basic mistakes - for example mixing up tenses and forgetting to use the right endings.	Can make self understood with short, simple phrases, but often need to stop, try with different words - or repeat more clearly what was said.

The same content could also be presented with the categories on the left and the grades across the top, as follows

	1 (A2)	2	3 (A2+)	4	5 (B1)
RANGE & PRECISION	Can communicate in a simple and direct exchange of limited information in everyday situations; otherwise has to compromise the message.		Can talk about familiar everyday situations and topics, with searching for the words; sometimes has to simplify.		Can talk about family, hobbies and interests, work, travel, news and current events. Can make the other person understand the most important points.
ACCURACY	Can use correctly simple phrases learnt for specific situations, but often makes basic mistakes - for example mixing up tenses and forgetting to use the right endings.		Can use some simple structures correctly in common everyday situations.		Can express self reasonably accurately in familiar, predictable situations.
FLUENCY	Can make self understood with short, simple phrases, but often need to stop, try with different words - or repeat more clearly what was said.		Can participate in a longer conversation about familiar topics, but often needs to stop and think or start again in a different way		Can keep a conversation going, but sometimes has to pause to plan and correct.

The marking grid for this hybrid approach would look the same as that for the holistic approach, but with marks as entries rather than levels. The final step will be to convert the total marks to an overall level. The marks for the individual criteria in this worked example might not be exact translations of the entries for levels shown for this worked example in the previous, holistic, grade grid. This is because now, rather than being forced to chose



between levels, the teacher has the possibility of awarding a mark between the defined levels (i.e. 2 or 4).



Candidate	RANGE	ACCURACY	FLUENCY	OVERALL
Marisa	5	2	4	11 = A2+
Angeles	5	5	3	13 = B1-
Cristina	2	3	3	8 = A2+ (low)
Yolanda	2	1	2	5 = A2 (good)
Albert	1	1	3	5 = A2 (good)

2.3. Giving a mark

One may prefer to rate learners' performance in terms of their success at achieving the task targeted at a specific CEFR level. In this approach other levels are irrelevant. The question is the degree to which the performance of the learner reflects the descriptor(s) for the target level. This is reflected in marks for each criterion. The approach is "analytic".

The worked example above might be presented for such a grid as follows:

	Candidate A					Candidate B				
RANGE & PRECISION Can talk about familiar everyday situations and topics, with searching for the words; sometimes has to simplify.	1	2	3	4	5	1	2	3	4	5
ACCURACY Can use some simple structures correctly in common everyday situations.	1	2	3	4	5	1	2	3	4	5
FLUENCY Can participate in a longer conversation about familiar topics, but often needs to stop and think or start again in a different way	1	2	3	4	5	1	2	3	4	5

- Key:**
- 5 = Is above the level described; is the next level
 - 4 = Yes, even unprepared, in difficult circumstances
 - 3 = Yes, under normal circumstances
 - 2 = Sometimes, with help and encouragement
 - 1 = Is below the level described; is still the previous level

It is important to note that with this approach, the criteria do not necessarily need to be the same for each level. One might take a different approach at the A levels, at the B levels and at the C levels. One might decide that at A1, grammatical accuracy is pretty irrelevant- but that getting student's pronunciation off to a good start is crucial. We might decide that global task completion is more important than range, ending up with:

- Task completion (i.e. Can they do it)
- Fluency
- Comprehensibility (i.e. - is their pronunciation clear?)

One might decide that Fluency is a given at C2, but that both Thematic Development and Discourse Coherence are really important – giving four rather than three categories:

- Range and Flexibility (i.e. do they express themselves in varied and appropriate ways)
- Accuracy and precision (i.e. do they use language correctly to pinpoint their exact meaning)
- Thematic development (getting ideas across; the equivalent of "task completion" at A1/A2)
- Discourse coherence (is the language well-structured with good linking; does it relate well to previous contributions, weave in different perspectives).

The following two examples are real grids for Speaking and for Writing at A2 in an extensive teaching context, from Avo-Bell, Sophia:



Speaking	A2					Candidate A					Candidate B				
Discourse management Can make her/himself understood in very short utterances, even though pauses, false starts and reformulation are very evident. Can link groups of words with simple connectors like "and", "but" and "because"	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Pronunciation Can produce words / phrases in a recognizable manner and make her/himself understood in simple everyday situations but there may still be noticeable interference from mother tongue in terms of individual sounds, word and sentence stress	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Interaction Can ask and answer questions and respond to simple statements. Can indicate when s/he is following but is rarely able to understand enough to keep conversation going of her/his own accord.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Grammar and vocabulary Uses basic sentence patterns with memorized phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations. Uses some simple structures correctly, but still systematically makes basic mistakes.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Global achievement Overall impression mark	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

Writing	A2					Candidate A					Candidate B				
Text Management Can make her/himself understood in short sentences. Can produce a short but logically connected text which is relevant to the task. Can link groups of words and sentences with simple connectors like "and", "but", "when" and "because"	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Communication strategies / Effect on the target reader Can convey more complex meaning using strategies like: reporting events in chronological order; describing aspects of everyday life; filling in questionnaires	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Layout and organisation Can use more confidently opening / closing expressions in a limited number of written tasks, e.g. simple letters, postcards, descriptions Can link ideas in clear paragraphs	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Grammar and vocabulary (accuracy and appropriacy) Uses basic sentence patterns with memorized phrases, groups of a few words and formulae in order to communicate limited information in simple texts on everyday topics. Uses some simple structures correctly, but still systematically makes basic mistakes.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5



Global achievement														
Overall impression mark / Task achievement (all points covered)	1	2	3	4	5	1	2	3	4	5				



2.4. Which approach

The three approaches discussed above have different appeals. One cannot say that one approach is better than the other. It is partly a question of style and tradition and partly a question of context. The first (holistic) focuses on proficiency as a continuum, which for convenience is chopped into levels. The last (giving marks) focuses on the level in question as an educational objective. These represent the two classic perspectives on criterion-referenced assessment. The hybrid approach is a compromise between the two, with at least the potential to combine the advantages of both.

Holistic: This approach tends to mean using the same criteria for all levels, though as CEFR Table 3 shows, entries under any heading (e.g. "Range") can have a different emphasis at different levels.

Advantages: Simplicity and coherence: teachers – and potentially learners - learn one set of criteria; the whole school judges on the same basis, comparability between levels is easy, profiling uneven competence across the categories is easy.

Disadvantage: Criteria not focused on the special "salient" features of the level(s) concerned. As described in CEFR 3.6, simplified in Equals CEFR Reference Sheet 2, there are significant changes in the nature of the game (e.g. between B1 and B2).

Verdict: Particularly suited to intensive schools that may:

- be large, with considerable staff turnover and a need for simple, coherent systems
- be small, with a certain range of level in each class
- be offering short summer courses for learners to activate their competence at a particular CEFR level rather than progressing through syllabus levels
- be profiling aspects of student's competence for diagnostic reasons linked to guided self study (= balance your profile to fully activate your competence)
- have students progressing up syllabus levels in a relatively short period, compared to extensive schools

Analytic: Assigning marks to each of a set of criteria; the criteria may be the same at each level or different at successive levels or groups of levels.

Advantage: Criteria can be focused on the salient issues at the level concerned

Disadvantage: Some danger of dislocation between different departments of the school; difficulty of crediting an exceptional student with achievement of a level above the aim of the course.

Verdict: Very popular in French and German pedagogic culture. Also particularly suited to extensive contexts in which:

- the same learners may climb up a school's syllabus over a number of semesters and years
- there tend to be a larger number of sub-levels to show progress
- there may be a need for formality (e.g. reports for parents) and hence evaluation of attendance, effort, homework; these could also be rated 5-1 with definitions and added to the grade/report form

	RANGE	PRECISION	LINKING TEXT AND	ACCURACY	FLUENCY	INTERACTION
C2	I can reformulate ideas in differing ways to give emphasis, to differentiate and to eliminate ambiguity	I can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of expressions to qualify statements and pinpoint the extent to which something is the case.	I can make full and appropriate use of a variety of organisational patterns and a wide range of connectors in order to organize what I say and write.	I can consistently maintain grammatical control of complex language even when my attention is otherwise engaged.	I can express myself naturally and effortlessly; I only need to pause occasionally in order to select precisely the right words.	I can interact naturally, picking up and using non-verbal and intonational cues without effort, and interweaving my contribution into the joint discourse with fully natural turn taking, referencing, etc.
C1+	I have a good command of a broad vocabulary, including collocations and idiomatic expressions; I can overcome gaps in my vocabulary with alternatives.	I can reformulate ideas differently in order to ensure that people understand exactly what I mean, eliminating the possibility of misunderstandings.	I can produce clear, well-structured speech and writing, showing control over ways of developing what I want to say in order to link my ideas into smoothly-flowing, coherent text.	I can consistently maintain a high degree of grammatical accuracy; errors are rare and difficult to spot.	I can express myself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.	I can use fluently a variety of appropriate expressions to introduce my remarks in order to get the floor, or to gain time and keep the floor while thinking. I can relate my contribution skilfully to those of other speakers.
C1	I have a good command of a broad vocabulary. I do sometimes have to search for expressions but can then find alternatives to express what I want to say.	I can formulate opinions and statements in a very precise manner in order to indicate my degree of agreement, certainty, concern, satisfaction, etc.	I can produce clear, well-structured speech and writing, linking my ideas into coherent text.	I maintain a high degree of grammatical control in speech and writing.	I can express myself fluently and spontaneously, except occasionally, when speaking about a conceptually difficult subject.	I can select from a readily available range of expressions to preface my remarks appropriately and to follow up what other people say. (New)
B2+	I can express myself clearly and without much sign of having to restrict what I want to say. I can reformulate ideas in different ways to ensure people understand exactly what I mean.	I can focus my attention effectively on how I formulate things, in addition to expressing the message.	I can use a variety of linking words efficiently to mark clearly the relationships between ideas.	I can maintain good grammatical control. I may sometimes make mistakes but I can correct them afterwards.	I can communicate fluently and spontaneously, even when talking at length about complex subjects.	I can intervene appropriately in discussion, using a variety of expressions to do so. I can help the development of a discussion by giving feedback, follow up what people say and relating my contribution to theirs
B2	I have a sufficient range of vocabulary to vary formulation and avoid repetition when expressing myself on matters connected to my field and on most general topics.	I can explain the details of an event, idea or problem reliably.	I can link what I say or write into clear, well-organised text, though I may not always do this smoothly so there may be some jumps. (New)	I can communicate with reasonable accuracy and can correct mistakes if they have led to misunderstandings.	I can produce stretches of language with a fairly even tempo; although I can be hesitant as I search for expressions, there are few noticeably long pauses.	I can use standard phrases like 'That's a difficult question to answer' to gain time and keep the turn while formulating what to say. I can help a discussion along on familiar ground confirming comprehension and inviting others in.
B1 +	I have a sufficient range of language to describe unusual and predictable situations and to express my thoughts on abstract or cultural as well as everyday topics (such as music, films).	I can explain the main points relating to an idea, problem, or argument with reasonable precision.	I can use connecting words to link sentences into a coherent sequence, though there may be some jumps ⁵ .	I can communicate with reasonable accuracy in familiar contexts, though with noticeable influences from my mother tongue.	I can express myself relatively easily when talking freely and keep the conversation going effectively without help, despite occasional pauses to plan and correct what I am saying.	I can join in a discussion on a familiar topic, using a suitable phrase to do so. I can sum up what has been said in order to help the discussion to move forward.
B1	I know enough vocabulary to talk about my family, hobbies and interests, work, travel, news and current events.	When I explain something, I can make the other person understand the points that are most important to me.	I can link a series of short phrases into a connected, sequence of points.	I can express myself reasonably accurately in familiar, predictable situations.	I can keep a conversation going, but sometimes have to pause to plan and correct what I am saying.	I can ask someone to clarify or elaborate what they have just said. I can repeat back part of what someone has said to confirm that we understand each other.
A2+	I know enough vocabulary for familiar everyday situations and topics, but I need to search for the words and sometimes must simplify what I say.	I can generally communicate the main points of what I want to say, though I sometimes have to simplify it.	I can use the most important connecting words to tell a story (for example, first, then, after, later).	I can use some simple structures correctly in common everyday situations.	I can participate in a longer conversation about familiar topics, but I often need to stop and think or start again in a different way	I can start, maintain, or end a short conversation in a simple way. I can ask somebody to repeat what they said in a simpler way.
A2	I have enough vocabulary to communicate in simple everyday situations.	I can communicate what I want to say in a simple and direct exchange of limited information; in other situations I generally have to compromise the message.	I can link ideas with simple connectors. For example: 'and', 'but' and 'because' ⁵ .	I can use correctly simple phrases I have learnt for specific situations, but I often make basic mistakes - for example mixing up tenses and forgetting to use the right endings.	I can make myself understood with short, simple phrases, but I often need to stop, try with different words - or repeat more clearly what I said.	I can start a conversation. I can say what exactly I don't understand and ask simply for clarification.
A1 +	I have a basic repertoire of phrases to talk about myself and communicate in common everyday situations.	I can communicate limited information about myself, my family and my job in a simple and direct exchange.	I can join phrases with words like 'and', 'but', or 'because' 'then'.	I can use correctly some simple structures that I have memorized.	I can speak in slowly a series of very short phrases, stopping and starting as I try and say different words.	I can simply ask somebody to speak more slowly.



A1	I have a very basic repertoire of words and phrases about family and personal details, plus simple everyday situations.	I can communicate very basic information about myself and my family in a simple way.	I can join simple phrases with words like and or then.	I can use memorised, short phrases for specific purposes with reasonable accuracy.	I can speak in very short phrases and isolated words.	I can say when I do not understand. I can very simply ask somebody to repeat what they said.
-----------	---	--	--	--	---	---

Moderation: Techniques to limit teacher subjectivity

The most obvious way to support teacher judgements is by backing them up with assessment tasks for speaking and writing as discussed in the last section.

The following techniques are also recommended:

- Progress tests
- Second assessors
- "Tuning in" before assessment by individual assessors
- Scores from an "anchor test" calibrated to the CEFR
- Quality control by the academic manager

1. Progress tests

Progress tests can be used to support the judgements the teacher makes. Provided that the tests reflect the content of the syllabus for the CEFR level, they give teachers a good idea of each learner's mastery of the relevant content and skills, even though the tests scores have not been calibrated to CEFR levels. Progress tests can be produced primarily with materials from course book and examination past papers, as described in the parallel guide on CEFR Progress Tests.

It is important that:

- Simple "teacher tests" of grammar covered are not used exclusively; if progress tests are used as a source of information for certification (rather than just to reinforce learning), it is important to test skills as well as knowledge
- Teachers do not create "ceilings" through unrealistic expectations of learner accuracy. Research shows that accuracy typically declines at B1 as learners struggle to deal with more complex range of situations and tasks and the wider range of language to deal with them. Accuracy typically picks up again at B2, B2+.

2. Second assessor

When teachers teach into each others classes, which is common in intensive schools, the second teacher can be involved in a negotiation of grades for the students in the class.

Other possibilities are:

- The class teacher rank orders pieces of written work from the class and gives the top, bottom and middle script to another teacher to double mark independently.
- Alternatively, each 5th or 10th script is assessed by a second assessor
- With classroom assessment, a second assessor (e.g. the Academic Manager) comes into the class for that lesson, assesses in addition to the class teacher and then later negotiates final grades with them.
- With interviews, each 5th or 10th student is assessed by a second assessor.

3. "Tuning in"

This approach avoids double-marking by substituting a live standardisation session at the start of the assessment.



- With interviews, the teachers start by interviewing one or two students as a panel of 2-3 teachers in order to standardise amongst themselves before assessing the rest of their class. It is important that explicit reference is made to the criteria grid in the discussion, referring back to the illustrative samples (DVD) where necessary. It may be a good idea to have the documentation to the DVD samples for the level(s) concerned handy.
- In marking written scripts, the same approach can be taken - necessitating the photocopying of 2-3 scripts for each person beforehand. This approach is easier when assessing writing because the illustrative samples themselves can be available for reference and comparison. The classic approach would be to look at the illustrative samples before starting and to select one or two as "benchmarks" particularly relevant, particularly relevant to the scripts that are going to be marked from the point of view of task and level.

4. Anchor tests

Tests calibrated to the CEFR can be used as reference points. There is no reason to expect a one-to-one relationship between the teacher assessment results of any individual learner and their result on an anchor test, but for a school of 150, a very definite pattern should emerge.

The use of an anchor test has two positive influences:

- it reminds the teacher of what level grade most of the students should be receiving, and hence over time it dampens down the teacher's own tendency to strictness or lenience
- given a record sheet showing both the anchor test results and the teacher assessments, the Academic Manager can easily "eyeball" the columns and spot whether any particular teacher is awarding a set of results that are systematically above those from the test (= too lenient) or systematically below those from the test (= too strict).

The anchor test needs to have validity as regards the content of the level, and will work better if it is a test reporting a "global" level. Cloze or modified cloze tests or other integrated tasks from past papers for Cambridge or other examinations could also play this role.

5. Quality Control by the Academic Manager

The school has accumulated experience of a CEFR curriculum. As a result, the Academic Manager should be able to predict the *expected* results at the certification point in question and use this as a point of reference.

The class is one of a series of classes, some above the class concerned and some below it. Clearly there will be an overlap between the CEFR results of adjacent classes, but one would not normally expect reversals, except with a teacher known to be exceptional.

The learner is a student in a class at a certain CEFR level. The teacher knows who the stronger and weaker students are. The question for the teacher is: Is this student so strong that they have already achieved the next level? At A1/A2 it is difficult for this to happen as the learner is being introduced to language for the first time. From B1 upwards, however, what is happening is a broadening and deepening of the learners competence, which may be accelerated by autonomous practice (e.g. a reading habit) or external events (e.g. an exchange trip to Canada). Such students are exceptions that the teacher can and should recognise. A "grade sheet" is produced for each class and checked carefully by the Academic Manager or

Centre Manager. Exceptions happen, but the onus should be on the teacher to argue the case convincingly for the exception, providing evidence in the process. The teacher can be required to show the Academic Manager work samples from students who show exceptional progress and to whom the teacher wishes to give a higher CEFR level.



Case Study: Eurocentres

Standardisation:

Eurocentres is an exception because there is a direct relationship between the levels used in Eurocentres since 1993 for curriculum organisation, assessment and certification and the CEFR, since Eurocentres produced the CEFR descriptor scales. One third of the CEFR descriptors related to the active skills come exclusively from Eurocentres and two-thirds partly from Eurocentres.

Documented videos have been available for staff training since 1991, recently replaced by the CEFR illustrative samples. A set of written scripts calibrated in 1989 is also available for reference.

The levels are explained to teachers as part of the induction process and occasional staff seminars are used for standardisation training with videos.

Teacher Assessment:

Written homework is graded at certification points with a 10-level assessment grid definition Language, Organisation and Communication (task fulfilment) at each level.

Spoken language is assessed in group activities in the classroom. An example of a very simple, prototypical activity of the Eurocentres type is given below. It was written by Ricardo Piancastelli of Eurocentres Florence in 1991, originally for Italian. Ideally one selects/invents an activity that brings together the communicative objectives of the current learning module. Communicative speaking activities of this type are a very positive learning experience for students. They react positively to the teacher using them for assessment, and to an external assessor sometimes also being present. The three phases of the activity are used for three phases of the assessment process (collaboration phase: impression judgement; exchange phase: analysis with the criteria; discussion phase: considered judgement). The criteria defined on a 10-level assessment grid are Range, Accuracy, Delivery (=Fluency & Pronunciation) and Interaction.

Curriculum checklists (Can dos & language points) have been discussed with students at assessment points since 2007, but in practice this process has been found to produce too much information to be acted upon; the procedure will be reviewed for 2009.

Teachers provide grades for 6 aspects: Listening, Reading, Spoken Interaction (from the classroom assessment task), Spoken Production (from the exchange phase of the classroom assessment task), Writing (from graded homework) and Language Resources (from the anchor test). Each aspect is defined on a 10-level CEFR-related scale. Assessment of Listening and Reading is thus purely by teacher judgement, supported by performance in class, informed by the profile the student already has for the other aspects.



STORY OF A MARRIAGE

By Riccardo Piancastelli, Eurocentres Florence
(with notes by teachers at Eurocentres Cambridge).

This is an information-gap activity. For the first phase, each of four groups of students is given a photo of both a man and a woman (the "couple who get married").

Collaborative Phase:

Each group has to construct the love story which led up to Janet and John's marriage. They should then decide how things changed after marriage, describing the character defects and bad habits of each partner which would eventually lead to their divorce. Each group should provide as much detail as possible about their story.

Exchange Phase:

Re-group the students so that there is one representative of each of original groups in each new group. The representatives of each of the first groups take it in turns to tell their version of the drama of Janet and John to the rest of the group.

Discussion Phase:

Divide the class into two groups to discuss their "ideal family". Suggested prompts:

- what age should men/women get married?
 - how many children would you like to have?
 - nuclear vs. extended family?
 - conventional families vs. communal living?
 - is marriage an out-of-date idea?
- etc.

Moderation

Moderation is through an anchor test, occasional second assessors and quality control by the Academic Manager.

1. Progress Tests: Formal progress tests are not used for certification purposes. Classes have quizzes at the end of each week; course book materials are exploited to reinforce learning; French schools also set "tests de verification" at intervals to check that grammatical themes have been mastered.
2. Second assessors: With inexperienced teachers, the Academic Manager or an experienced teacher acts as second assessor for the classroom assessment. Final grades are then negotiated between the two teachers after the activity.
3. Tuning-in: This technique is only used in placement test interviews.
4. Anchor test: Eurocentres uses 36-40 item tests of knowledge of the language system ("recognition", "one word" and "complete the sentence" items of various kinds) produced from an item bank called "Itembanker" which was developed between 1991 and 1998 for English, German, French and Spanish. The system translates scores from all possible tests into levels and plus levels on the Eurocentres scale and the CEFR. For Italian, tests in the same style are done on line, but the items are not formally calibrated to the CEFR. The results from the classroom assessment and the anchor test are averaged to give the learner's global level. This is the first step, before determining the full 6-aspect profile.
5. Quality Control: Student entry level and progress profiles are recorded on Student Record Forms that are then used to produce certificates. Academic Managers consciously check grades; teachers are required to provide evidence in the form of homework to support the award of exceptional grades.