

A QUALITY MATRIX FOR CEFR USE: Examples of promising practices

1. OVERVIEW

Project leader(s) contact: Tim Goodier
Brian North

Country: UK

Institution: Eurocentres Foundation

Type of context: Programme/school level; Classroom level

Educational sector: Adult

Main focus of your project: Assessment by the classroom teacher

SUMMARY

Name: Classroom assessment in small groups

Abstract: Technique to assess CEFR level of learners working on communicative tasks in small groups, using tasks that generate different types of discourse and defined criteria for quality of performance, related to CEFR descriptors.

Stage: Evaluation

Theme: Teaching; Assessment

CEFR aspects targeted: Levels; tasks; assessment with defined criteria

Main features of this example:

- tasks cross-referenced to can-do descriptors (CEFR Chapter 4)
- learners regrouped to create information/opinion gap
- usually three distinct discourse phases (collaboration, exchange, discussion)
- CEFR-based criteria grid: Range, Accuracy, Delivery, Interaction, Overall = RADIO

Quality principles particularly demonstrated: Validity, Coherence



2. PROJECT DESCRIPTION

Background:

This approach was developed in the mid-1980s, and implemented fully in Eurocentres for English from the early 1990s, with attempts to spread to other languages (most successful with German). It was part of a project to develop and implement a system of levels, assessment and certification around a scale of language proficiency (a precursor of the CEFR proficiency scale).

The model for the types of tasks was developed in a Birmingham MA project (North 1986), inspired by John Sinclair's model of classroom discourse, soon after Jane Willis developed her initial task-based model in such a project (Willis 1983).

The first version of the criteria grid was developed in 1989, inspired the CEFR Table 3 criteria grid "Qualitative aspects of spoken language" developed in 1997, and has been revised several times since in order to take account of experience, performance in video recordings and wording of the CEFR descriptors.

The procedure is supplemented by tests of knowledge of the language system drawn from an item bank based on language specifications for each level. The level certificated is a mixture of the teacher assessment and result from the test bank.

Stated aims:

To develop a pedagogically positive form of assessment of level by the class teacher, which would encourage the use of communicative tasks related to previous language work, be administratively feasible and provide reliable results.

Steps/stages:

1. *Defining a model for tasks*

It is important that tasks used for oral assessment generate spoken production (= sustained monologue) and spoken interaction (between peers). Interviews, especially interviews conducted by teachers, almost always reproduce the IRF (Initiation, Response, Feedback) discourse model of the traditional classroom, thus tending to deny the candidate the discourse initiative and a platform to show what he/she can really do. A second problem is the tasks used in pair work and group work often either (a) simulate functional situations, often in roleplays that follow a predictable pattern, or (b) involve little more than communicative drills, giving reinforcement and fluency practice in certain language, rather than authentic discourse. Finally, tasks often generate one single type of discourse which, however long, remains unrepresentative.

In the Eurocentres model, firstly the tasks have a structure that ensures that everyone speaks in groups of 3–5 students, and secondly, the tasks generate three phases with different types of discourse. There are several templates for tasks, but the most popular are "two-tier group work", in which groups are remixed for the second phase, so that each member of the new group comes from a different first group. The second popular type is, "press conference," in which groups will be planning a product or concept of some kind, which they then, in the second phase, present to the class. The three phases are as follows:

(i) *Collaborative phase* in which the group work out what to do (CEFR scales: *Goal-oriented Co-operation*). In this phase, learners plan the task and work on understanding and/or creating text. They produce short, slow turns with high use of cognitive and communication strategies (CEFR scales: *Facilitating collaboration in groups; Collaborating to construct meaning*).

(ii) *Exchange phase:*

In "two tier group work" the groups are then remixed so that each student has unique information and takes their turn to tell the others what their first group suggested or decided (CEFR scales: *Information Exchange*, possibly *Sustained Monologue, giving information; Sustained Monologue, putting a case*, possibly *Processing text*). In "press conference," groups take turns to go to the front and make a presentation (CEFR scale: *Addressing audiences, Sustained Monologue, giving information, and possibly Sustained Monologue, putting a case*). In this exchange phase students produce (in turn) far longer, semi-prepared, coherent turns reporting on and discussing a prepared topic with high fluency.

(iii) *Discussion phase:*

This then inevitably leads into a discussion phase as the group (“two-tier group work”) or the class (“press conference”) compares the proposals of each of the first groups (*Informal Discussion*). At lower levels (A1-B1) students tend to just enjoy having a discussion, with language and strategies similar to the collaboration phase. There is, again, a slower and more conscious use of complex language in spontaneous and usually shorter turns. At higher levels, however, discussion often develops into a debate, including long turns (*Sustained Monologue, putting a case*) at a fast tempo, which sometimes become dominated by certain individuals.

2. *Selecting/designing tasks*

The idea is that the tasks should bring together “can-do” descriptors that are objectives for the current module (week). When the technique was developed, these kind of communicative tasks were not common, but now they are found in many course books, as well as classic supplementary resources. In addition about 60 such tasks were custom written.

As a result, teachers can select such tasks that relate to ‘can-do’ descriptors from a resource finder (done in Excel; printed on a poster and stuck on the wall in the staff room). Nowadays, however, with authentic material so accessible with interactive white boards (IWBs), many teachers invent their own tasks.

The best tasks are those in which each group has to actually produce a final product or concept – otherwise the discussion phase can become a bit aimless.

3. *Assessment procedure*

The three distinct phases are used for three stages of assessment: initial impression, detailed analysis with criteria, considered judgement, as summarised in the table below:

	Assessment Procedure	Instructions for a standardisation training activity
1. Collaborative Phase Group works out what to do (short, slow turns with high use of communication strategies);	Impression: Write down the overall impression of the global level of each candidate that you have after about 5 minutes.	While viewing, after 4-5 minutes, write a single level – your overall, initial impression – in the space at the top of the rating form.
2. Exchange Phase in which each student has a chance to take the floor (long, coherent turns which are semi-prepared)	Analysis: Consciously read the descriptors for that level across the assessment grid. If you confirm that the candidate does meet the criterion description for a category at that level, look at the level above in that same category to see if they are even better than that. Write a result for each assessment category (Range, Accuracy, Fluency, Interaction, Coherence if using CEFR Table 3; Range Accuracy, Delivery, Interaction if using Eurocentres’ grid).	While viewing, after marking that initial judgement, consciously read the descriptors for that level across the assessment grid, for the level above and the level below. After viewing, read the criteria closely and mark your decision for each category on the form in the space provided
3. Discussion Phase in which some members of the group take things further (spontaneous, short turns).	Judgement: Compare your analysis result to your original impression and make a considered judgement.	Consult the CEFR scales for “Overall Spoken Interaction” and “Overall Spoken Production” for a final check. Write your final decision at the bottom of the form in the space provided.

North 2014: Table 5.3 Eurocentres oral assessment procedure

4. Developing an assessment grid

There are lots of ways to develop an assessment grid. There is a fundamental difference whether

- One single grid is used for all levels for all tasks
- One single grid is used for all tasks at one level
- Every task at every level has its own assessment grid (of tasks of specific types share the same grid)

The first option is the case here. This makes things simple in a programme in which there will be learners at different levels and teachers teach into different levels (because they work in teams). Another advantage is that it means that the grid used for standardisation training (which covers all levels) is the same as that used operationally.

The Eurocentres RADIO grid was developed with the following steps:

- Take ideas and formulations from respectable sources to produce a first draft. Nowadays, the existence of the CEFR and defined criteria for exams makes this process much easier; but at that time it was more difficult.
- Consult colleagues and ask all involved to try using the grid to assess real performances. Refine the grid based on feedback from this process.
- Consciously compare the wording of the grid to key characteristics of performances that can be viewed together (videoed or live).
- Review the grid every few years. This was done circa every 5 years in Eurocentres.

	Range	Accuracy	Delivery	Interaction
B1+	Able to use a range of simple language flexibly, and explain a point with reasonable precision, but can't always say what they would like to. Familiarity with main tenses, modals and major sentence patterns.	Reasonable accuracy with basic tenses etc. in everyday contexts. Frequent errors and inappropriate expressions occur, partly due to mother tongue influence, but it is clear what he/she is trying to express.	Gives extended descriptions, able to keep going effectively without help, despite some problems with formulation resulting in pauses and "cul-de-sacs". Stress and intonation may be very foreign, but can generally be followed okay.	Handles structured discussion on familiar topics easily, inviting others in, commenting on views, comparing and contrasting alternatives. Participation more restricted in freer or unfamiliar contexts.
B1	Relatively wide repertoire of simple language, for familiar subjects, but limited alternatives. Normally requires simplification of intended message. Good level of familiarity with basic tenses and sentence patterns.	Reasonable accuracy with a repertoire of frequently used "routines" and patterns. Tendency otherwise to mix up tenses and pick the wrong word or expression - may be conscious of this and try to self-correct.	Can keep going clearly and comprehensibly, though perhaps slowly, pausing especially in longer stretches. Frequent reformulations and hesitations and/or heavy interference from L1 may make comprehension difficult.	Initiates, maintains and closes simple interaction with some cooperation from the interlocutor. Can exchange information and repeat back part of what someone has said to confirm mutual understanding.

Extract from Eurocentres spoken assessment grid: RADIO (Note: red = exact CEFR content)

The cells for Range and Accuracy have comments about the use of particular language, which would need to be developed separately for different languages and contexts.

Timeline:

- 1986 Analysis of activities
- 1987 Development and piloting of techniques for informal observation and assessment of level
- 1989 1st Criteria grid RADIO developed
- 1990 1st Standardisation video produced
- 1992 Introduced in Eurocentres for English
- 1994-8 Attempts to transfer to French, German, Italian, Spanish (mixed success)
- 2002 Criteria grid revised to incorporate CEFR wording
- 2004 2nd Standardisation video produced, rated to CEFR as well as Eurocentres
- 2005 Procedures formalised and standardised; resource finder produced

People/roles:

Development (1986-1990)

- Analysis of activities as part of a Birmingham MA project (Brian North, 1986)
- Initial piloting at Eurocentres Bournemouth involved the Director of Studies (Brian North) and teams of teachers. The criteria (range, accuracy, delivery, interaction, overall) were decided at this point – but not yet defined with descriptors
- The Head of the Eurocentres then research and development department (John Arnold) then suggested defining descriptors for each criterion for each level.
- Descriptors were developed with the team of teachers piloting the approach
- A small team put together an initial set of suitable tasks (some written, some from books like *Discussions that work* (Penny Ur)
- An expert team was put together to produce a set of standardisation videos by rating videos extracts from the Birmingham MA project onto the scale of levels with the criteria. The team consisted of:
 - o the Eurocentres Bournemouth teachers piloting the approach in class
 - o other Eurocentres teachers who were Cambridge examiners
 - o the then Chief Examiner for Cambridge First Certificate (Heather Daldrey, Eurocentres London)

Dissemination:

- Brian North (by then a member of the Eurocentres research department in Zurich) was tasked with implementing the approach across all Eurocentres as one measure related to the introduction of a new leaving certificate in 1992.
- Representatives from all Eurocentres schools met for a seminar in Amboise in 1991, at which tasks were also written for different languages
- A teacher in London, in Washington, Paris and Cologne were each tasked with preparing a suitable set of tasks.

Further Development:

- Brian North revised the initial assessment grid in 2002 and in 2007 in relation to (a) the CEFR and (b) feedback received.
- A teacher was commissioned to produce/collect more tasks several times.
- A teacher in London (Nareene Kaloyan) produced a chart cross-referencing available tasks to the CEFR-based 'can do' objectives for each level.

In practice:

- The academic manager of a school organises a standardisation training session at the start of each summer season.
- An experienced assessor (usually a senior teacher) acts as second assessor with less experienced teachers, with grades negotiated between the two assessors after the lesson.

Other resources needed: None

Quality Assurance procedures employed:

Standardisation training is carried out before every high season (summer).

A senior teacher acts as second assessor with less experienced teachers, with grades negotiated between the assessors after the lesson.

Publications that have been used or produced related to this example:

North, B. (2014): *The CEFR in Practice*, Cambridge: Cambridge University Press, (Sections.4.3.2 & 5.1.4).

North, B. (1993): "L'Evaluation 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.

North, B. (1991): "Standardisation of Continuous Assessment Grades" in *Language Testing in the 1990s*, Alderson, J.C. and North B 1991, London, Macmillan/British Council: 167–177.

3. RESULTS

What was achieved:

A popular, standardised approach to classroom assessment introduced across schools in UK, USA, Canada, Australia.

Impact:

Eurocentres experience over 20 years suggests that classroom assessment with small group tasks as described above can:

- be very motivating;
- incorporate authentic materials and issues that learners have views on;
- offer a natural monologue that is embedded in interaction as is the case in real life, since all learners have a right to a long turn, encouraging coherent articulation of thoughts, even at A2;
- generate spontaneous discussion in which learners develop interaction strategies (CEFR scales: *Turntaking, Cooperating, Asking for clarification*).

The assessment procedure (separating initial, holistic impression from later considered judgement with defined criteria) also impacted on the procedure for Cambridge orals, initially for the CAE (C1) and later for the other levels of the Cambridge main suite of exams.

Resources on this theme:

Not available

4 ADVICE AND LESSONS LEARNT:

Advice on this theme; things to remember:

- It is really important to give very clear instructions to the students, backed up by a written version (on paper or board)
- It can be difficult with classes over 20 – but not impossible; it has been used with classes of 40!! With large classes, the teacher must choose which groups to focus on in detail each time.
- The regrouping is essential to ensure that (a) everyone gets their turn to produce a longer turn of coherent discourse.
- The teacher can use that longer turn to home in on particular students.
- Often, the teacher does not have a fully completed grid for all students on every occasion. In practice this turns out not to be such a frequent problem with classes of 12-15 (teachers quickly get used to the technique). A few blanks, or “hedging” between, for example B1+/B2 on certain criteria (quite common) does not in the end matter, unless feedback is given on individual criteria. It is the final, considered judgement at the end that counts.
- It is important to have occasional (e.g. yearly) standardisation sessions in which teachers watch videos and rate the learners with the criteria. In Eurocentres this happens in early summer every year.

Advice on this theme; pitfalls to avoid:

- Avoid very lengthy instructions for teachers: they don't read them.
- Avoid putting people that dislike each other into groups together.
- Don't intervene, unless (a) a group gets lost or (b) one student is clearly dominating a group: otherwise stick to a role as observer/assessor.
- Ensure teachers have had standardisation training with available CEFR videos showing levels (see CEFR or Equals websites).

Lessons learnt/Issues to watch out for:

- Teachers may have a tendency to revert to simple 'discussion' activities, if not steeped in the rationale behind the three-phase design.
- Avoid overambitious tasks for which there is not sufficient time!!
- There may be a danger of not immediately noticing that a group has misunderstood the task.