



February 2005

DGIV/EDU/LANG (2005) 1

Seminar to calibrate examples of spoken performances
in line with the scales of the
Common European Framework of Reference for Languages

CIEP, Sèvres, 2 - 4 December 2004

REPORT

By

Brian NORTH (Eurocentres)
Sylvie LEPAGE (CIEP)

Language Policy Division
Strasbourg

The seminar was organised by the Centre International d'Etudes Pédagogiques (CIEP) and Eurocentres as part of a project to provide samples of oral performance for French illustrating the levels of the Common European Framework of Reference (CEFR) (Council of Europe 2001). The aim of the seminar was to calibrate performances that could be presented and documented on a DVD illustrating performances at levels A1 to C2 of the CEFR. The programme is given as Appendix 1.

Context

Examination institutes, language schools and university departments concerned with the teaching and testing of the French language are currently in the process of relating their curricula to the CEFR. A problem that arises in this regard is the question of assuring a consistent interpretation of the levels in different contexts.

In July 2002 the Finnish authorities organised a seminar for the Council of Europe (DG IV/EDU/LANG - 2002) to discuss this issue, as a result of which an authoring group was established to produce a preliminary version of a Manual to help institutions to relate their examinations to the CEFR (DG IV/EDU/LANG – 2003/5). That Manual, published for piloting in September 2003 envisages the process of linking an exam to the CEFR in three stages:

- Specification: define the coverage of the examination in terms of the CEFR;
- Standardisation: ensure a consistent interpretation of the CEFR levels related to the interpretation elsewhere, exploiting illustrative samples of performances already calibrated to the CEFR in this process;
- Empirical Validation: check that the results given by the examination relate to the levels of the CEFR in the manner foreseen.

The primary aim of the French DVD project was to calibrate the oral performances that could be used on a standardisation DVD to facilitate a consistent interpretation of the CEFR levels for French. The project was an initiative that arose from the meeting in Strasbourg in April 2004 concerning piloting and case studies in relation to the draft Manual. At this meeting Sylvie Lepage presented the CIEP case study with regard to DELF DALF. Two videos illustrating the levels of the CEFR for English had also been distributed before the meeting. These consisted of samples from the CEFR Swiss project put together by Brian North and Gareth Hughes, plus a cassette illustrating performances at Cambridge ESOL examinations at the different CEFR levels. The Eurocentres Foundation and the CIEP decided at the meeting to produce the DVD in order to provide illustrative samples for the second official language of the Council of Europe: French. A further meeting of a group of experts concerned with the provision of illustrative samples held in Strasbourg in October 2004 was exploited as an opportunity to discuss in depth both the organisation of the programme of the planned seminar and the analysis of the resultant data.

Aims of the Seminar

The notion of a standardisation video exemplifying performances at different levels on a scale of language proficiency is in fact an innovation both in France and in the world of French as a foreign language. As a result there is an interest in such a tool not only from examination boards but also from language schools (e.g. Eurocentres), accreditation associations for language schools (e.g. EAQUALS) and from both teachers and teacher trainers who work with the CEFR and with the European Language Portfolio (ELP).

Because of this wide interest in such a DVD, and because this was the first international benchmarking seminar in the Council of Europe's projects related to the CEFR and the Manual, the seminar had significant process aims in addition to the main objective of producing the actual product – the DVD of calibrated samples. The full aims could be summarised as follows:

Process aims

- Establish a consensus in the interpretation of the CEFR levels in relation to learner performances in French as a foreign language (Français Langue Étrangère = FLE in France and abroad).
- Give participants practical experience of such a seminar so that they would be better able to organise one themselves. This applied both to those institutes involved in the world of FLE in France and also to those institutes invited who planned to produce DVDs illustrating CEFR levels during 2005 (Goethe-Institut for German; Instituto Cervantes for Spanish; University of Perugia for Italian; University of Lisbon for Portuguese).
- Pilot a methodology for running such a seminar, giving practical input for a Guide to organising such an event.

Product aims

- Calibrate examples of spoken performance for young adults from a wide range of nationalities.
- Identify those examples most suitable for inclusion on a DVD illustrating CEFR levels for French.
- Document the selected examples based on discussion at the seminar, the criteria of CEFR Table 3 (Oral Assessment Grid) and statistical characteristics.

Essentially, in an international benchmarking seminar of this type there are two different levels of aims that are to some extent incompatible:

1. Calibrate and document the performances

In a language testing context, such a task is normally carried out by a relatively small group of expert raters trained and experienced in the test context concerned. Essentially the examination institute gives the experts the authority to dictate the way the criteria are to be interpreted and the experts justify their decision by documenting the relationship of the performances to the criteria and by demonstrating the degree of consistency (intra-rater reliability) and agreement (inter-rater reliability). In the case of a seminar aiming to calibrate to a *common* (not institute-specific) framework, it was not so obvious whether calibrations based on the opinions of a closed group of experts would be representative of different perspectives in the wider world of FLE.

2. Establish a consensus in interpreting the CEFR levels in the pedagogic culture(s) concerned

The alternative is to assemble a wider group of experts and establish a consensus. The establishment of such a consensus cannot be separated from the process of training. A larger number of raters from different backgrounds will need training in order to follow the same defined procedure and to make judgments in relation to the same CEFR criterion descriptors. After training, “consensus” may be interpreted to mean:

- a. an averaging of differing opinions from individual judgments;

- b. convergence to agreement through discussion.

Ideally both approaches should produce the same result.

In the inclusive tradition of the Council of Europe's modern languages projects, the decision was taken to invite a wide range of participants representing different perspectives and seek consensus. It was decided to collect data *both* on individual judgments representing such different perspectives *and* on the consensus formed after discussion.

Participants

Efforts were therefore made to ensure attendance by a large group representative of different perspectives on the interpretation of CEFR levels for French. The 38 participants listed in Appendix 2 represented four different groups:

- experts from the French examinations boards for the French language: CIEP and Alliance Française (10 persons);
- teachers from French language schools in French-speaking countries (10 persons);
- experts in the French language from other European educational systems (11 persons);
- experts in the CEFR levels who were not experts in the French language (7 persons).

The decision to invite such a large group representing radically different perspectives was taken for several reasons. Because of the fact that little standardisation discussions had taken place in relation to French as a foreign language it was felt politically important to include as many people from the world of French as a foreign language as possible. Secondly, the examination institutes and French language schools in France were simultaneously starting to standardise on the CEFR, and it was felt better to seek convergence rather than to impose the view of one institution on others. Thirdly, since use of the CEFR was advanced in other countries in Europe, it was felt wise to include approximately 50% of participants from contexts abroad.

In addition to conventional analysis, more sophisticated statistical analyses in which the four groups are separately identified were undertaken. This would show whether there are significantly different interpretations between the different groups of raters represented. The organisers are grateful to Cambridge ESOL and ALTE (Association of Language Testers in Europe), and in particular to Neil Jones, for the support offered in carrying out these analyses.

Performances

The video recordings were made between May and September 2004 at Eurocentres Paris, the CIEP studio at Sèvres, the Centre International d'Études de Langue at Brest and the Collège International de Cannes. Other recordings were also made in DELF DALF test centres at Prague and Madrid, but these had to be abandoned because of technical problems with sound and vision quality. The recordings shown at the seminar had performances from young adults from 15 countries: Belgium, Brazil, China, Columbia, Germany, Great Britain, Italy, Mexico, Peru, Serbia-Montenegro, Sweden, Switzerland, the Ukraine, the United Arab Emirates and the United States of America.

Each learner had signed a form authorising the transmission of their video image on such a standardisation DVD and in the internet. The learners filmed were selected in a systematic

fashion and their language level in relation to the CEFR was documented with teachers' evaluations, with questionnaires and with test results – especially for the new CEFR-based DELF DALF.

The filmed performances were then viewed at a workshop on 14th September by an expert group consisting of Béatrice Dupoux (CIEP), Gareth Hughes (Migros Club Schools – and member of the Portfolio Validation Committee), Sylvie Lepage (CIEP), Marie-Claude Moyer (Eurocentres France) and Brian North (Eurocentres – and coordinator of the CEFR Manual authoring group). At this workshop the performances were rated onto the CEFR levels and discussed before a final selection of performances for the seminar was made. The performances are listed in Appendix 3 from level C2 to A1.

Tasks

The tasks filmed were a further development of those developed for standardisation videos in the Swiss research project that had produced the CEFR levels and descriptors. These tasks had been used for the English video circulated in April 2004 and were recommended in the Council of Europe's "Brief for Recording."

Each recording shows two learners, with no native-speaker examiner/interlocutor. There are three phases:

- a production phase by the first learner with a sustained monologue, which may be followed by questions from the other learner;
- a similar production phase by the other learner;
- a phase of spontaneous interaction between the two learners.

This same test format was used at all levels. The production phase is "semi-prepared" in that the learners can choose a topic and have about 10 minutes in which to reflect on what they want to say. The interaction phase is completely spontaneous, elicited by cards (or strips of paper) containing discussion questions. As with the production phase there is an element of learner choice in the topics for the interaction phase in that the two learners could discard topics that did not interest them.

Recordings with a total duration of up to 12 minutes were shown in their entirety. Longer recordings (a couple lasted 20 minutes) were shortened to extracts of 3-4 minutes for each of the three phases.

Rating Instruments and Procedures

The main rating instrument was the criteria grid presented as CEFR Table 3 (Appendix 4), which defines for each of the 6 CEFR levels the requirements in terms of Range (Étendue), Accuracy (Correction), Fluency (Aisance), Interaction (Interaction) and Coherence (Cohérence). This grid had been distributed to participants before the seminar with the request to study it.

Participants were also provided with a supplementary grid giving descriptors for the "plus" levels in the middle of the CEFR scale A2+, B1+ and B2+ (Appendix 5). The addition of these 3 "plus" levels provided 9 levels. These 9 levels reflect the linear scale produced in the Swiss research project. In the CEFR, descriptors on the scales that are at these "plus" levels are presented above a horizontal line, with the descriptors for the criterion level underneath this horizontal line.

Before giving the final, Global judgement, it was stressed several times that participants should consult other relevant CEFR scales with which they had also been provided (Appendix 6). These consisted of a selection of scales for Spoken Interaction and for Spoken Production from Chapter 4 as well as the scale for Phonological Control from Chapter 5, since this qualitative category was not included in CEFR Table 3.

Rating was first done on paper using Form B2 (analytic rating form) from the Manual given as Appendix 7. Votes were recorded electronically with CEFR levels corresponding to buttons on the keypad as shown in Appendix 8.

The same basic pattern of rating was followed once the initial familiarisation and training phases later described had been completed.

1. Watch the sequence (Production: Learner A; Production: Learner B; Interaction between Learners A & B).
2. Consult criteria grids and record on Form B2 the ratings for both learners for the 5 categories: Range (Étendue), Accuracy (Correction), Fluency (Aisance), Interaction (Interaction) and Coherence (Cohérence).
3. Consult other scales provided, reflect and rate the Global Level for both learners on Form B2.
4. Electronic Voting: individual votes: Range (Étendue), Accuracy (Correction), Fluency (Aisance), Interaction (Interaction) and Coherence (Cohérence), Global (Note Globale)
5. View Histogram of the individual global judgements.
6. Group discussion in groups of 6-7 mixed by background and with a CIEP person as chair/rapporteur. Normally 10 minutes discussion.
7. Reports from Group leaders.
8. Plenary discussion.
9. Electronic Vote: Global (Note Globale).
10. View Histogram of global judgements after discussion.

The rationale was to:

- a) capture the full data of participants' individual judgements before any conferring;
- b) attempt to establish consensus through (i) seeing one's own anonymous vote in the context of the total pattern of votes, and (ii) discussion.

As mentioned before, the desired result for sequences that would be included on the DVD was that the product of the objective analysis of the individual data (a) would coincide with the consensus arrived at in the discussion (b).

Familiarisation Exercises

At the beginning of the seminar, however, before moving on to rate performances, participants completed two familiarisation exercises along the lines suggested in the Manual, modified by suggestions made at the October seminar in Strasbourg, modified again to meet the requirements of the time available.

The first exercise involved identifying the CEFR level of descriptors from the two CEFR scales “Overall Spoken Interaction” and “Overall Spoken Production” when these were presented in a mixed list in random order. A version of the worksheet with the answers added is given as Appendix 9. The follow up was an explanation of why each descriptor was the level at which it was assigned, by drawing attention to the salient features characteristic of the level concerned. This exercise was effective and drew attention to the fact that some participants were misleading themselves by taking one word out of context (e.g. “can interact.....” in a descriptor for A1, despite the fact that this descriptor is qualified with provisos and conditions).

The second familiarisation exercise involved the grid to be used as rating criteria in the seminar: CEFR Table 3. Participants were asked to work with their neighbour in pairs. The exercise was an ambitious one in which participants were given a blank grid and 24 pieces of paper to place in the correct cells, with the warning that 6 of the 30 cells were missing. Some participants completed the task astonishingly quickly and appeared to thoroughly enjoy it. Many, however, had difficulty seeing which descriptors were defining which category. This was particularly the case with Range (Etendue) and Accuracy (Correction). This may have been at least partly caused by the fact that, judging by later discussion in the seminar, it appears to be customary in French as a foreign language to interpret the former as “Vocabulary” – rather than as the range of language resources available - structures, turns of phrase, words, knowledge of collocation and colligation etc.

The first evaluation sequence was also a form of familiarisation task. Participants were asked to watch a first short sequence showing two learners considered to be B1, consult the criteria and rate just the overall, global level of each learner. This task aimed to get participants accustomed to the rating instruments and to the process of first recording judgments on paper and then voting electronically, but it also served to show the degree of agreement in the group.

Training in Procedures and Criteria

There then followed a phase of training in which the production phases of three sequences were rated separately by each of three criteria.

Two such sequences (Margarida and Mariana; Valérie and Sophie) were rated on the Thursday. The third sequence (Deborah and Iryna) was evaluated on the morning of the main rating day (Friday) after the explications on the use of the criteria.

The same procedure was followed for each of the three sequences:

- First the production phase of Learner A was viewed and rated individually with RANGE (with no discussion), then it was repeated and rated for ACCURACY (again no discussion), then it was repeated a third time and rated for FLUENCY – without discussion.
- Exactly the same was then applied to Learner B
- Finally, the Interaction phase – showing both learners – was played once, and participants were asked to finalise their decisions on the three criteria used and to make a global judgement on the overall level.

Discussion in groups of 5-6 followed for approximately 15 minutes, followed by a second viewing of the Interaction phase, and a final judgement (the vote after discussion).

Participants found discussing learner proficiency and justifying their opinions in relation to specific criteria during the seminar a very positive experience, as shown in their feedback comments in Appendix 11. Several also commented that they much appreciated the explanations of the Table 3 criteria and other CEFR descriptors during the seminar. However, the process of focusing on detailed formulations for one aspect of performance rather than starting with the question: “What level is this person in the general scheme of things?” did appear to cause some participants difficulty. This point is returned to at the end of the report.

Analysis Method

The statistical analysis includes the provision of conventional descriptive summaries of ratings (mean, modal). However the main analysis employed a multi-faceted Rasch model scaling analysis, using the program FACETS (Linacre 1989). The illustrative video for English showing Swiss adult learners contained examples calibrated with FACETS in the original CEFR research project (100 raters: North 2000) and confirmed at a final conference of the Swiss project (circa 25 raters again with a FACETS analysis).

Although not exactly user-friendly, FACETS has the following advantages:

- **Rater Consistency**

Raters who are inconsistent in their judgements (sometimes lenient, sometimes strict) can be identified and removed from the analysis.

In initial analysis runs it appeared that 3 or 4 raters in Sèvres were close to or outside the conventionally acceptable criteria for rater consistency. FACETS, like all Rasch programs, produces “fit statistics” that quantify the extent to which something or someone “fits” the expectations and assumptions of the statistical model. A rater who “misfits” is being inconsistent: rating good people low, and weak people high – or in some other way changing their standard for judgments. The quality and accuracy of the measurement is improved if data from such raters is removed from the analysis. The discovery that 5-10% of raters show inconsistency is not unusual. There can be many reasons. Sometimes the “misfit” is concentrated in just one or two judgements, and just that unreliable data can be removed.

- **Rater Severity**

Differing severity (leniency/strictness) of the raters can be taken into account in calculating the “ability estimate” for the learners. Extreme cases can be eliminated if desired, though FACETS actually compensates for severity, thus objectifying the judgements. Provided lenient or severe raters are being consistent, therefore, they can be safely retained in the analysis.

In the Sèvres data, once the raters had tuned in, the range of severity they showed was not remarkable. There were some differences between individuals, but the program FACETS is designed precisely to take this into account in arriving at an objective result.

- **Difficult Performances**

Learners who prove difficult to assess because of some extraneous factor can also be identified. Lack of “fit” with certain learners can be due to poor audio quality, lack of

familiarity with learners speaking the mother tongue concerned, bad pronunciation, or some other factor in the performance confusing to raters.

Among the samples shown Luis (a Peruvian paired with Aleksandar) misfitted seriously, reflecting people's extreme difficulty concentrating on what was an idiosyncratic, nervy performance with poor elucidation. The only other performances with any noticeable level of misfit was in the sequence with Ambrogio and Silvia. This interaction was somewhat more intimate than the others and slight misfit on both learners may have been caused by fast diction from Ambrogio and some embarrassment from Silvia.

- **Identifying Significantly Differing Interpretation Among Groups**

Since the 4 different groups to which participants belong are defined by context and were identified as a "facet" in the analysis, it is possible to measure whether any group:

- tends to rate the performance of any learner significantly differently to the other groups, or
- as a group is significantly more lenient or more strict than the other groups.

In the event there were no significant differences in these two respects in the interpretations of the four different groups of experts. Indeed the only noticeable distinction was that the group "*French language experts in European projects*" tended to be more severe on the criterion *Correction* (accuracy) and more lenient on *Interaction*, whereas the group "*French language schools in France*" rated more evenly across all the criteria.

- **Mathematical scaling to CEF Levels**

The cut-off points on the mathematical scale dividing the CEF levels (CEF Manual Table 6.2 given as Appendix 10) can be used to "anchor" the analysis to the (logit) scale created in the original CEF research project so that a mathematical "ability estimate" is made for each learner. This can be useful when discussing whether a learner is, for example, a strong B1, a typical B1, a borderline B1, etc.

There are issues with regard to the best interpretation of the results from a FACETS analysis of rating scale data and those who are interested are referred to the report available in English by Neil Jones, the ALTE data analyst. The FACETS scaling of the independent judgements before discussion produced an illuminating result that is interesting to compare with the result based purely on aggregating raters' opinions.

In making a final decision for each sample, the FACETS result from individual judgements was compared to the consensus opinion established after discussion.

Analysis Result

Inter-rater reliability: A first point to be made is that the ratings show an impressive level of inter-rater reliability. The mean inter-rater correlation for the independent judgements before discussion was 0.886, and that for the more consensual view after discussion was 0.967. In interpreting these high coefficients, it must be borne in mind that the full range of language proficiency was being assessed, from below Level A1 to Level C2. It is easier to achieve such high correlations in these circumstances. Nevertheless, such inter-rater reliability coefficients

compare well with those reported in the literature for ratings by expert, trained judges across the range of proficiency.

Halo Effect: A second point is that there was an extremely high inter-correlation of 0.992 to 0.998 between the five criteria: Range (Étendue), Accuracy (Correction), Fluency (Aisance), Interaction (Interaction) and Coherence (Cohérence). The criteria were clearly not being used in an independent fashion. To quote the analysis report: “On this evidence either the rating criteria are not distinct, or the raters were not able to identify that they were distinct (=halo effect) or the subjects do not display distinct profiles of skill.” This result is not unusual: such criteria give raters a focus to help them to arrive at a considered judgement and may not actually represent truly independent variables.

Training Effect: The analysis suggests that during the course of the seminar the participants as individuals became more consistent in their interpretation of the criteria (CEF descriptors). This can be seen in the way that the “fit” statistics for the facet “Day” improved over the three days as the participants became more familiar with the criteria and the procedures. Such an effect is to be expected in a seminar which had the aim of developing a consensus view.

CEF Levels of Samples: The placement of the samples at CEF levels is best considered by comparing the FACETS analysis of the independent judgements before discussion with the consensus reached after discussion. The latter cannot properly be subjected to a FACETS analysis because the judgements in the data set are not independent and because such an analysis cannot handle 100% scores - as happened for Josue and Rachel, for example, who were unanimously considered to be C2.

The following table compares the conclusions reached from the two sets of votes, and makes a final recommendation in the right hand column.

		Independent Judgements (FACETS)	Judgements after Discussion	Definitive CEF Level
33	Josue	C2	C2	C2
32	Rachel	C2	C2	C2
31	Aleksandar DALF	C1	C1	C1
15	Ambriogio	C1	C1	C1
25	Aleksandar	C1	C1	C1
8	Xi	B2+	B2+	B2+
26	Luis	B2+	B2+	(B2+)
16	Silvia	B2+	B2+	B2+
7	Nataliya	B2	B2	B2
37	Gu Jung	B1+	B1+	B1+
4	Sophie	B1+	B1+	B1+
3	Valérie	B1+	B1	B1
13	Evelyne	B1	B1	B1
1	Margarida	B1	B1	B1
2	Mariana	B1	B1	B1
14	Andrea	B1	A2+	(A2+) / B1

5	Deborah	B1	A2+	A2+
61	Katell	A2+	A2+	A2+
38	Aamer	A2+	A2+	A2+
6	Iryna	A2+	A2+	A2+
62	Sun Ying	A2+	A2	A2
50	Viggo	A1 (high)	A1 (high)	A1
42	Suzanne	0	A1	A1
41	Sally	0	A1	A1
49	Jessica	0	0	0

Luis is shown in italics because his performance showed a high level of misfit denoting inconsistent rating. His nervousness and a strong Hispanic accent to which some of the participants were not accustomed made him very difficult to focus on. The real-life Luis was clearly C1 and passed the DALF.

In the majority of cases the results from the two different approaches are identical. As can be seen from the table below, there are differences with six performances. These are for the following samples, for which a final decision has to be made:

- Valérie (B1+ or B1)
- Andrea (B1 or A2+)
- Deborah (B1 or A2+)
- Sun Jing (A2+ or A2)
- Sally (A1 or Tourist)
- Suzanne (A1 or Tourist)

Sally and Suzanne: The type of proficiency found immediately below A1 is referred to in the CEF as “Tourist.” Sally and Suzanne, however, should be regarded as A1. The fact they come out lower in the FACETS analysis is probably caused by a known weakness in the Rasch model employed by the program.

Being rated on a scale of 0-10, Sally and Suzanne each scored 0.7 or 7% (missing “93” marks). Josue scored 89% (missing 11 “marks”). Calibrations with the Rasch model are known to be suspect at “extreme scores;” the problem is to define where this “extreme score” distortion begins. In terms of the logit scale underlying the CEF levels Josue comes out massively above the C2 cut-off point with a mathematical value that is clearly no longer on a linear scale: the calibrated value is distorted. This doesn’t matter: he is still C2 because there is nothing on the scale above C2. Sally’s and Suzanne’s ratings of 7% are more extreme than Josue’s (at 7/93 rather than 11/89) – and the logit values calculated for them are therefore also suspect. Yet they only just miss the A1 cut-off score. Unfortunately this has a different consequence because, unlike the case at the top of the scale, beyond A1 going down the scale is no longer A1.

Sun Ying: Sun Ying comes out more or less exactly on the cut-off between A2 and A2+ (more properly described as A2.1 and A2.2). In other words, if one ignores the “plus level” distinction, she is *exactly* in the middle of the Level A2. For the final decision, she has therefore been described as A2, because she has not demonstrated that she is A2.2.

Deborah: Deborah is a Brazilian law student who was in one of the first sequences to be used, whilst participants were still getting used to applying the criteria. She was clearly better than Iryna, with whom she was paired. She spoke very confidently and probably benefited from a nascent tendency to rate the two candidates in relation to each other, rather than solely in relation to the criteria. After discussion focusing more closely on what language she was using, 26 of 37 raters agreed she was A2+.

Andrea: Andrea is a problematic case. Some participants thought she was reading in the Production Phase; certainly she had copious notes. Matters were not helped by the fact that the pair (Evelyn and Andrea) then misinterpreted a key word in the prompt card for the Interaction Phase that was shown, went off on completely the wrong track and were stopped by the teacher (who was off camera). Many participants referred to this lexical misunderstanding in their arguments during the group discussions as another reason for marking Andrea down, though others were convinced she was meeting the B1 criteria. After the discussion, 19 participants voted A2+ and 19 voted for B1 (including 4 for B1+), yet in the analysis she came out just under the B1 cut-off at a very high A2+. She is best regarded as a low B1.

Valérie: In the independent votes, Valérie received an almost perfectly balanced number of votes at B1 and at B1+. After discussion, 21 participants voted for B1, 9 for B1+ and 5 for B2. As with Sun Ying in relation to Level A2, if one ignores the “plus level” distinction between B1 and B1+ (more properly called B1.1 and B1.2), Valérie is a good example of a learner right in the middle of the full B1 band. Again, for the final decision she has been put at B1 as she had not clearly demonstrated that she was B1.2.

In relation to the other, more straightforward samples one or two interesting points can be made in relation to the position of some of the candidates in the CEF proficiency band concerned, on the basis of the FACETS logit values from the independent judgements before discussion:

Aleksandar comes out on the assessment task used for all the performances at the bottom of the C1 band just above the cut-off point. On the DALF task, on the other hand, he comes out at the top of the C1 band with a far stronger, more focused performance.

Ambrogio and Silvia are each only just above the cut-off points on the bands C1 and B2+ respectively. Yet judgements after discussion were almost unanimous that they are indeed those two levels.

Katell and Aamer both come at the top of the A2+band: they are extremely strong A2 performers.

Viggo is at the top of the A1 band: a very strong A1 performance.

Feedback from Participants

Participants clearly appeared to have profited from the seminar. The time planning had been over ambitious, the need for discussion of and training in interpreting the levels themselves before getting into different detailed aspects of individual proficiency (Range, Accuracy, Fluency etc.) had been rather underestimated, but nevertheless the seminar had been a rewarding experience that produced a concrete result.

The extended group discussions that followed the individual votes were the defining feature of the seminar. Many participants commented how much they enjoyed them. There was a strong feeling voiced during the seminar that the quality of the discussion and consideration of criteria was vital and that a genuine consensus should be reached through argument, rather than decisions being made primarily through arithmetical averaging in the data analysis. Therefore the discussions took considerably longer than planned and the evaluation of a 10-15 minute sequence took approximately one hour, but participants definitely wanted to invest this amount of time.

Many participants gave personal thanks and suggestions towards the end of the seminar, and during or after the closing session 35 of the 36 participants other than the two organisers completed the questionnaire for which the results are collated in Appendix 11.

The feedback from the questionnaires can be briefly summarised as follows:

Most participants were best acquainted with the A and B levels, especially level B1. Only nine of them said they knew all levels equally. The guests were meant to be familiar with the CEFR but several of them said they gained a better appreciation of the levels especially the plus levels during the seminar.

What participants found most useful were the group discussions, which enabled them to share their opinions with professionals from different backgrounds, and the requirement to justify their opinions in those discussions in relation to transparent criteria.

The clarifications of some descriptors and the explanations on the differences between the levels and the plus levels were also welcome. Some participants reported finding the plus levels difficult to use. Others appreciated the opportunity thus offered to take note of points from the following level that was in the process of acquisition.

Most people would have liked to have more time either for familiarisation (more time spent on defining range, accuracy and fluency), more time to agree on the meaning of some descriptors and more time to discuss the performances. A few people said they would have liked to see the extracts twice (only a few extracts were replayed). Five people said they also wished there had been more training at the beginning of the seminar.

It would have helped to make clear from the start whether we were assessing if the student had acquired some of the skills described at the level or whether he had attained the level.

Two participants made the point in detail that it was difficult to commence by assessing in detail with 5 criteria - and taking into account plus levels. One participant suggested first viewing sequences and putting them into rank order of proficiency, and then on a second day re-viewing the sequences with the criteria in order to decide where to put the boundaries between levels. The second participant suggested that the rating procedure for such seminars should involve:

- initial impression of level: A B or C
- decision on main level, with criteria (Table 3): A1, A2, B1, B2, C1 or C2
- refinement by consideration of the relevance of plus levels (A2+, B1+, B2+).

There was no agreement on what was most difficult: including pronunciation into the assessment, understanding the difference between the levels and the plus levels, assessing the first sequences, having the same interpretation of the descriptors, assessing two students at the same time...

Although the goals of the seminar were explained in the letter of invitation, some people appreciated the fact they were stated again clearly at the beginning of the second day.

Most of the participants think they will use the experience from this seminar to train teachers and examiners once the DVD is published. In addition to the Goethe and Cervantes Institutes, the Centre de Langues du Luxembourg and the Collège International de Cannes are also thinking of organising a similar seminar but on a smaller scale for the latter institutions.

Finally, it is perhaps worth quoting the conclusions of Neil Jones, the ALTE data analyst, at the end of his report:

“The Sèvres seminar provides an excellent model for how similar events should be run in future. Important features included:

1. The participation of a large number of well-qualified language professionals, including those with authority to interpret the video extracts in relation to the CEFR,
2. A technical infrastructure (electronic voting) which enabled the error-free capture of a large amount of data;
3. A planned sequence of stages, beginning with familiarisation with the rating criteria.

Inevitably perhaps the approach was revised during the event. My interpretation of this process is that:

1. the early ratings were felt to be excessively individualistic, failing to create the feeling of a group endeavour to create a shared understanding of the levels;
2. the remedy for this was to focus more on the text of the CEFR descriptors, that is, to present a more authoritative view of how particular extracts should be interpreted in relation to them;
3. in consequence the standardisation aspect, focussing on the interpretation of scale descriptors, took prominence over a more exploratory, construct-focussed approach based on the direct comparison of performances.

It may be that it is difficult or impossible to pursue both goals within the context of a single event. In this case the emphasis placed on standardisation was certainly the correct one, given the requirement to produce a set of accurately-rated exemplars for the DVD planned as support material for the CEFR Manual.

None the less, the direct comparison of performances remains an important goal, because equating levels across languages can probably only be done satisfactorily by suitably qualified judges making such direct comparisons. Comparisons mediated by the text of the CEFR are indirect and subject to a problem evident in the seminar: that the wording of the illustrative scales is to an extent relative rather than absolute.

It was very important that two sets of ratings were captured, both before and after discussion. The discussion in small groups was particularly useful because this seemed to favour useful

reflection on features of the performance and on why one was inclined to a particular rating. The plenary discussion that followed the reports from groups and preceded the final vote tended to have a more prescriptive tone. The difference is evident in the data: the data which are amenable to analysis using FACETS are the independent ratings, because they contain enough variability to enable scale construction. The post-plenary data provide the best view of the final consensus as to the level of each extract, but are quite unsatisfactory from the point of view of Rasch analysis, because of the evident lack of local independence in the ratings – they strongly violate the assumptions of the Rasch model. It is crucial that if part of the purpose of a rating exercise is to construct a credible measurement scale then there must be an opportunity for rating both before and after plenary discussion.”

Conclusion

One can state with confidence that all concerned found the seminar a very positive experience and that the aims of the seminar have been fulfilled:

Process aims

- Establish a consensus in the interpretation of the CEFR levels in relation to learner performances in French as a foreign language (Français Langue Étrangère = FLE) in France and abroad. Although this consensus was achieved, it is, however, clear that participants had many different starting points and that a considerable number of workshops, plus exams calibrated to CEFR levels, will be necessary before such a consensus becomes common.
- Give participants practical experience of such a seminar so that they would be better able to organise one themselves. This applied both to those institutes involved in the world of FLE in France and also to those institutes invited who planned to produce DVDs of samples illustrating CEFR levels during 2005 (Goethe-Institut for German; Instituto Cervantes for Spanish University of Perugia for Italian; University of Lisbon for Portuguese). Participants appreciated this practical experience. 75% said they would use the DVD for training teachers, test developers and examiners.
- Pilot a methodology for running such a seminar, giving practical input for a Guide to organising such an event. See reflections below in relation to events designed to produce DVDs of samples illustrating CEFR levels.
- Pilot a methodology for running such a seminar, giving practical input for a Guide to organising such an event. See remarks below.

Product aims

- Calibrate examples of spoken performance for young adults from a wide range of nationalities Samples are calibrated for all CEFR levels showing learners from 14 nationalities.
- Identify those examples most suitable for inclusion on a DVD illustrating CEFR levels for French 23 individual performances covering all levels from below A1 to C2 are included on the DVD.

- Document the selected examples in relation to discussion at the seminar, the criteria of CEFR Table 3 (Oral Assessment Grid) and statistical characteristics. Documentation will be produced on the basis of reports from the group leaders and further reflections in relation to performances and criteria.

Experience with this seminar suggests that organisers of benchmarking and standardisation events should not underestimate the need at the beginning of the seminar for lengthy, detailed training in judging levels and relating the set of criteria to samples. If, in a benchmarking seminar as in this case, the training cannot be done with reference to an authority (published calibrated samples), then samples rated at the beginning of the seminar need to be used as points of reference during discussion over further performances. This may mean replaying such reference samples – which again has implications for time management.

Secondly, several more experienced participants pointed out that going from detailed criteria to a global judgment was the reverse of what is an intuitive order for many people: going from general to particular. Two suggestions by participants to deal with this issue were mentioned above. It is also worth mentioning that the CEFR Manual includes a Global assessment scale (Table 5.4) intended to be used precisely for gaining this first impression of level before turning to the detailed criteria in CEFR Table 3.

Finally, the use of “plus levels” (A2+, B1+ and B2+) had disadvantages as well as advantages. Reference to the criteria for the “plus levels” led to much discussion about levels being attained (“acquis”) as opposed to being in the process of acquisition (= previous plus level). Many participants said they much appreciated the richness of these discussions, and came to see that the “plus levels” were a far more interesting concept than they had realised. However, the plus levels – and the discussions they led to – were also a complication which cost a lot of time.

The feedback from participants on this and other issues, the experience and notes of the group leaders and the reflections of the organisers will be fed into a practical guide for the organisation of future such events.

Brian North, Eurocentres, Zürich
 Sylvie Lepage, CIEP, Sèvres
 17.01.05

Appendices

- 1. Programme**
- 2. List of Participants**
- 3. Performances evaluated**
- 4. Oral Assessment Criteria Grid (CEFR Table 3)**
- 5. Supplementary Grid with criteria for A2+, B1+, B2+**
- 6. Analytic Rating Form: Draft Manual for "Relating language examinations to CEFR" - Form B2**
- 7. Supplementary CEFR Scales**
- 8. Electronic Keypad and CEFR Levels**
- 9. Familiarisation Exercise : identify level of descriptors**
- 10. CEFR Cut-offs on the underlying Statistical Scale**
- 11. Summary of Questionnaire Feedback**

Appendix 1

PROGRAMME DU SEMINAIRE SUR LE CALIBRAGE DE PERFORMANCES ORALES

Sèvres, 2-4 Décembre 2004
Grande Bibliothèque et Salle des conférences du CIEP

Jeudi 2 décembre

14.30-15.00 OUVERTURE (Grande Bibliothèque)
Discours d'accueil

Albert Prévos, directeur du CIEP
Johanna Panthier, administratrice au Conseil de l'Europe
Brian North, Responsable du Développement Pédagogique
de la Fondation Eurocentres

Présentation du programme Sylvie Lepage, Chargée de programme DELF DALF

15.00-16.15 FAMILIARISATION

- Attribuer un niveau à chaque descripteur proposé (CECR)
- Reclassez les descripteurs du tableau 3 du CECR.

16.15-16.45 PAUSE (Salle café)

16.45 - 18.30 (Salle de conférences)

FORMATION ET DISCUSSION

Première séquence :

Margarida (brésilienne) Mariana (colombienne)

Evaluation critère par critère et phase par phase

production 1 / production 2 / interaction

- Discussion en groupes et mise en commun

Deuxième séquence :

Valérie (belge) Sophie (anglaise)

Evaluation critère par critère et phase par phase

18.30-19.30 APERITIF (Salle Bleue)

19.30 DINER (Salle à manger)

Vendredi 3 décembre

(Salle de conférences)

9.00-11.00 FORMATION ET DISCUSSION

Troisième séquence :

Deborah (brésilienne) Iryna (ukrainienne)

Evaluation critère par critère et phase par phase

11.00-11.30 PAUSE

11.30-13.00 EVALUATION DE 2 SEQUENCES

Nataliya (ukrainienne) Xi (chinoise)

Evelyne (suisse) Andrea (suisse)

13.00-14.30 DEJEUNER

14.30-16.30 EVALUATION DE 3 SEQUENCES

Ambrogio (italien) Silvia (allemande)

Aleksandar (Serbie-Monténégro) Luis (péruvien)

Aleksandar (Serbie-Monténégro) format DELF DALF

16.30-17.00 PAUSE

17.00- 18.30 EVALUATION DE 2 SEQUENCES

Rachel (américaine) Josue (péruvien)

Sally (américaine) Suzanne (américaine)

19.15 DINER AU SELF OU EN VILLE

Samedi 4 décembre

(Salle de conférences)

9.00-11.00 EVALUATION DE 2 SEQUENCES

Jessica (américaine) Viggo (suédois)

Katell (chinoise) Sun Ying (chinoise)

11.00-11.30 PAUSE

11.30-12.00 EVALUATION D'UNE SEQUENCE

Gu Jung (chinois) Aamer (émirien)

12.00-12.30 SYNTHÈSE ET CLOTURE

Evaluation du séminaire

Discussion

Appendix 2

List of participants

1.	Reidun Oanes Andersen Responsable de l'évaluation pour l'entrée à l'université de Bergen	Adult Education Association Norsk språktest, University of Bergen, Olaf Ryes vei 11 c N-5007 Bergen Norvège	reidun.andersen@nor.uib.no Tél: + 47 55 58 26 96
2.	Arlette Aubier Assessora Tècnica Docent	Departament d'Ensenyament de la Generalitat de Catalunya Via Augusta 202-226 08021 - Barcelona Espagne	aaubier@pie.xtec.es Tél. : + 34 93 400 69 16 poste 3907
3.	Denise Besch enseignante de français et de luxembourgeois, chef du Comité de Pilotage pour l'adaptation des examens du Centre de Langues au Cadre Européen	Centre de Langues, Luxembourg Bvd de la Foire L-1528 Luxembourg	denise.besch@education.lu Tél: + 352 369066
4.	Sybille Bolton Responsable du développement des nouveaux examens	Goethe-Institut Zentrale Bereich 326 Dachauer Str. 122 80637 München	bolton@goethe.de Tél.: +49 89 15921-382
5.	Daniel Bonnardel Chef de Cabinet du Recteur Président du jury académique du DELF	Rectorat de l'Académie de Nice 53, avenue Cap de Croix 06 181 - Nice cédex 2	Daniel.Bonnardel@ac-nice.fr Tél : + 33. 5.93.53.72.35
6.	Gilles Breton Responsable du Test de Connaissance du Français	CIEP 1, avenue Léon Journault 92318 – Sèvres Cédex France	breton@ciep.fr Tél: + 33.1. 45 07 60 59
7.	Marie Bretonnier Directrice pédagogique	Collège International de Cannes 1 av Docteur Pascal 06400 Cannes France	pedagogie.cic@french-in-cannes.com Tél : + 33.4.93.90.55.77
8.	Nathalie Carvin Professeur de français langue étrangère	International School 9 cours des Arts et Métiers 13100 Aix-en-Provence – France	liselotte@is-aix.com info@is-aix.com Tel : + 33. 4 42 93 47 90
9.	Dominique Chevallier Chargée de programme DELF DALF	CIEP 1, avenue Léon Journault 92318 – Sèvres Cédex France	chevallier@ciep.fr Tél. : + 33 . 1 45 07 60 74

10.	Yasmine Daàs Attachée de coopération pour le français	Institut français de Prague Štěpánská 35 111 21 Praha 1 République Tchèque	yasmine.daas@ifp.cz Tél : + 42 0 221 40 10 64
11.	Dorothée Dupleix Chargée de programme Test de Connaissance du Français	CIEP 1, avenue Léon Journault 92318 – Sèvres Cédex France	dupleix@ciep.fr Tél : + 33.1 45 07 69 37
12.	Béatrice Dupoux Chargée de programme DELFF DALF	CIEP. 1, avenue Léon Journault 92318 – Sèvres Cédex France	dupoux@ciep.fr Tél: + 33. 1 45 07 60 00 poste 6189
13.	Annie Fazelly Professeur de FLE	Institut de Touraine Place du 14 juillet 37000 – Tours France	chalabi.pedagogie@wanadoo.fr Tél: + 33.2. 47 05 76 83 47 37 91 71
14.	Jean-Daniel Garabedian Professeur de FLE	Eurocentres Paris 13 Passage Dauphine, 75006 – Paris France	par-info@eurocentres.com Tél: + 33.1. 40 46 72 00 53 28 15 33
15.	Barbara Glowacka directrice adjointe Chaire de Néophilologie	Uniwersytet w Białymstoku, Faculté Philologique Pologne	barbara@piasta.pl glowacka@fll.uwb.edu.pl Tél./fax.: + 48 (85) 7457516
16.	Kerstin Häggström Chercheuse adjointe	Enheten för språk och litteratur Institutionen för pedagogik och didaktik Box 300 SE-405 30 Göteborg universitet Suède	Kerstin.Haggstrom@ped.gu.se Tel: +46. 31. 773 23 22
17.	Anne-Marie Havard Chargée de programme DELFF DALF	CIEP. 1, avenue Léon Journault 92318 – Sèvres Cédex France	havard@ciep.fr Tél: + 33.1.45.07.60.00 poste 62.73
18.	Martine Havet Attachée de direction	Centre de Langues, Luxembourg 21, boulevard de la Foire L-1528 Luxembourg	attache-direction@cfl.lu Tél: + 352 26 44 30 88
19.	Catherine Houssa Chargée de programme Test de Connaissance du Français	CIEP 1, avenue Léon Journault 92318 – Sèvres Cédex France	houssa@ciep.fr Tél : + 33. 1 45 07 60 60
20.	Gareth Hughes Chef de projets Office de coordination des Ecoles-clubs Fédération des coopératives Migros	Ecoles clubs Migros, BP 1766 Habsburgstrasse 9 8031 Zurich Suisse	Gareth.Hughes@mgb.ch Tél: + 41. 1. 277 2035

21.	Neil Jones Senior Research and Validation Coordinator	Cambridge University ESOL University of Cambridge Local Examinations Syndicate 1, Hills Road Cambridge CB1 2EU Royaume-Uni	jones.n@ucles.org.uk Tél: +44. 1223 553739
22.	Tarja Jukkala Inspectrice principale Evaluation et certification en langues étrangères	Direction nationale de l'Enseignement BP 380 00531 Helsinki, Finlande	tarja.jukkala@oph.fi Tél. + 358 9 7747 7709
23.	Bożenna Jurkiewicz expert en baccalauréat (langues romanes)	Commission Centrale des Examens/Centralna Komisja Egzaminacyjna ul.Lucka 11 00 842 Warszawa Pologne	bozennajur@gazeta.pl Tél: 22 656 37 24 (bureau) 601 195 726 (portable).
24.	Florence Kentzel Professeur de FLE	Eurocentres La Rochelle Avenue Marillac 17000 – La Rochelle France	lar-info@eurocentres.com Tél: + 33. 5. 46 50 57 33 46 42 03 77
25.	Françoise Kerrien Responsable des études Français Langue Etrangère	Centre International d'Etudes des Langues Rue du Gué Fleuri BP 35 29480 – Le Relecq-Kerhuon France	francoise.kerrien@ciel.fr Tél : +33 2 98 30 45 79
26.	Christophe Lagorce Professeur de FLE	Eurocentres Amboise 9 Mail St Thomas 37400 – Amboise France	amb-info@eurocentres.com Tél: + 33. 2. 47 23 10 60 47 57 10 91
27.	Sylvie Lepage Chargée de programme DELFDALF	CIEP 1, avenue Léon Journault 92318 – Sèvres Cédex France	lepage@ciep.fr Tél : 33 1 45 07 60 00 poste 6233
28.	Céline Savin. Professeur de FLE	Institut Français de Madrid calle Maqués de la Ensenada, 12 28004 Madrid Espagne	celine.savin@ifmadrid.com Tél: + 34 91 700 48 25
29.	Marie-Claude Moyer Directrice Eurocentres France	Eurocentres France 13 Passage Dauphine, 75006 – Paris France	mcmoyer@eurocentres.com Tél: + 33.1. 40 46 72 00

30.	Fiona Murray State Examinations Commission responsable de l'organisation et la gestion des examens nationaux de français	State Examinations Commission 8, Orchard Close Ballincurrig Park Douglas Cork Irlande	murrayft@eircom.net Tél: + 353-21-4291282
31.	Isabelle Normand Chef de service pédagogie et certifications Alliance française de Paris	Alliance Française 101 bd Raspail 75006 Paris France	inormand@allianfr.org Tél.: + 33. 1 42 84 90 08
32.	Brian North Head of Academic Development, Eurocentres Foundation	Fondation Eurocentres Seestrasse 247, CH-8038 Zurich Suisse	bjnorth@eurocentres.com Tél: +41.1.485 52 45
33.	Johanna Panthier Administratrice	Conseil de l'Europe Division des Politiques Linguistiques DG IV F- 67075 Strasbourg France	Johanna.Panthier@coe.int Tél. +33. 3 88 41 23 84
34.	Alba Pardina Iñiguez Instituto Cervantes de Paris Responsable D.E.L.E.	Instituto Cervantes de Paris 8, rue Quentin Bauchart 75008 – Paris France	alba.pardina@wanadoo.fr Tél : + 33. 1 40 70 92 92
35.	Anne-Chantal Petter Chef de projets Office de Coordination des Ecoles-club	Ecoles clubs Migros, PO Box 1766 Habsburgstrasse 9 8031 Zurich Suisse	anne-chantal.petter@mgb.ch Tél: + 41. 1 277 25 62
36.	Veselina Popova Ministère de l'Education et de la Science Expert en chef de langues Direction de la politique dans l'éducation générale	Ministère de l'Education et de la Science boul. Kniaz Dondukov 2A Sofia 1000 Bulgarie	v.popova@minedu.government.bg Tél. : + 35929217452
37.	Béatrix Sampsonis Coordnatrice des travaux ALTE/Bulats	Alliance Française 101 boulevard Raspail 75006 Paris France	Bsampsonis@alliancefr.org Tél: + 33.1. 42 84 90 00 poste 2237
38.	Christine Tagliante Responsable du Pôle évaluation et certifications	CIEP 1, avenue Léon Journault 92318 – Sèvres Cédex France	tagliante@ciep.fr Tél: + 33. 1. 45 07 60 42

Appendix 3

Performances evaluated

15 mn au total C2	
A gauche : Rachel (américaine) C2	A droite : Josue (péruvien) C2
Production : Une expérience personnelle (3mn)	Production : un film (3mn)
Interaction : (9mn) Tous les hommes devraient savoir cuisiner. Le sport n'a-t-il que des vertus ? Les enfants d'aujourd'hui grandissent-ils plus vite ?	

16mn	
Aleksandar (Serbie Monténégro) : Les effets positifs et négatifs du tourisme dans votre pays. C1	<i>Luis (péruvien) : même sujet</i> <i>(non évalué)</i>
Interaction : Les enfants doivent commencer l'apprentissage des langues étrangères dès l'école primaire.	

15mn C1	
Aleksandar (Serbie Monténégro) Exposé : La malnutrition	FORMAT DALF

16mn C1/B2+	
A gauche : Ambrogio (italien) C1	Silvia (allemande) B2+
Production d'Ambrogio : Raconter un film : Le Terminal (4mn30) Production de Silvia : Raconter un film : Depuis que Ottar est parti (4mn)	
Interaction : (5 à 7mn) Le sport n'est pas bon pour le couple. (2mn) Faut-il condamner la chirurgie esthétique ? (3mn)	

(21mn au total) B2/B2+	
A gauche : Nataliya (ukrainienne) B2	A droite : Xi (chinoise) B2+

Production de Nataliya : Le port de l'uniforme doit-il être obligatoire ? (6mn)
 Production de Xi : Le rôle des femmes a-t-il changé dans votre pays dans les 20 dernières années ? (8mn)
 Interaction : La télévision est une catastrophe du point de vue social. (7mn)

6.30mn

A gauche : Valérie (belge) 1.30mn
 B1

A droite : Sophie (anglaise) 2mn
 B1/B1+

Production : Mon film préféré
 Interaction : Paris est la plus belle ville du monde. 3.00mn

15mn

Discussion générale : 4mn

Gu Jung (chinois) : un film (3.30mn A gauche)
 B1+

Aamer Al Aamri (Emirats) : une bonne expérience (3.30mn A droite)
 A2+

Interaction : 4.00mn
 Quelles vacances préférez-vous ? Pourquoi ?
 Votre vie en France

12mn au total

A gauche : Margarida (brésilienne) 4.20mn
 B1

A droite : Mariana (colombienne) 1.25mn
 B1

Production : Mon pays
 Interaction : Apprendre le français est inutile et sans intérêt. 5.10mn

12.30mn

<p>Evelyne Bösiger (suisse) : B1</p> <p>A gauche Le téléphone portable est-il indispensable ? 4.15mn</p>	<p>Andrea Amacher (suisse) : A2+/B1</p> <p>A droite Les enfants doivent commencer l'apprentissage des langues étrangères dès l'école primaire. 2.45mn</p>
---	--

Interaction : La télévision est une catastrophe du point de vue social. 2mn
 De nos jours les enfants grandissent plus vite que lorsque nous étions enfants.
 3.15mn

9mn au total

A gauche : Deborah (brésilienne) 0.45mn
A2+

A droite : Iryna (ukrainienne) 1.00mn
A2+

Production : Se présenter

Interaction : Les meilleures vacances 6.45mn

9mn au total

Katell (chinoise) : un bon souvenir 3.30mn
A2+

Sun Ying (chinoise) : une personne que vous aimez 1.30
A2

Interaction : Vos prochaines vacances 3.00mn

14 mn en tout

A gauche : Sally (américaine)
A1

A droite : Suzanne (américaine)
A1

Production

Sally : Une journée habituelle (1mn20s)

Suzanne : Ma famille (40 sec.) et ma ville (45 sec.)

Sally : Mes prochaines vacances (1mn15sec.)

Interaction :

Quelles sont vos activités préférées ? (1mn)

Est-ce que vous aimez regarder la télévision ? (3mn)

Quelles vacances préférez-vous ? (1mn30)

Présentez-vous l'une à l'autre et faites connaissance. (3mn)

13 mn en tout A1 et < A1

A gauche : Jessica (américaine)
Inférieur à A1

A droite : Viggo (suédois)
A1

Production

Jessica : ma famille (50sec.)

Mes vacances (30 sec) et ma maison (15 sec)

Interaction

Présentez-vous (1mn)

Quelles vacances préférez-vous ? (1mn)

Quelles sont vos activités préférées ? (1mn10)

Est-ce que vous aimez la télévision ? (2mn)

Appendix 4

Oral Assessment Criteria Grid (CEFR Table 3)

	RANGE	ACCURACY	FLUENCY	INTERACTION	COHERENCE
C2	Shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis, to differentiate and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.	Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).	Can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it.	Can interact with ease and skill, picking up and using non-verbal and intonational cues apparently effortlessly. Can interweave his/her contribution into the joint discourse with fully natural turntaking, referencing, allusion making etc.	Can create coherent and cohesive discourse making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.
C1	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say.	Consistently maintains a high degree of grammatical accuracy; errors are rare, difficult to spot and generally corrected when they do occur.	Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.	Can select a suitable phrase from a readily available range of discourse functions to preface his remarks in order to get or to keep the floor and to relate his/her own contributions skilfully to those of other speakers.	Can produce clear, smoothly-flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.
B2+					
B2	Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, without much conspicuous searching for words, using some complex sentence forms to do so.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstanding, and can correct most of his/her mistakes.	Can produce stretches of language with a fairly even tempo; although he/she can be hesitant as he or she searches for patterns and expressions, there are few noticeably long pauses.	Can initiate discourse, take his/her turn when appropriate and end conversation when he / she needs to, though he /she may not always do this elegantly. Can help the discussion along on familiar ground confirming comprehension, inviting others in, etc.	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some "jumpiness" in a long contribution.
B1+					
B1	Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Uses reasonably accurately a repertoire of frequently used "routines" and patterns associated with more predictable situations.	Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.	Can initiate, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest. Can repeat back part of what someone has said to confirm mutual understanding.	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
A2+					
A2	Uses basic sentence patterns with memorised 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.	Can make him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident.	Can answer questions and respond to simple statements. Can indicate when he/she is following but is rarely able to understand enough to keep conversation going of his/her own accord.	Can link groups of words with simple connectors like "and," "but" and "because".
A1	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorised repertoire.	Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words, and to repair communication.	Can ask and answer questions about personal details. Can interact in a simple way but communication is totally dependent on repetition, rephrasing and repair.	Can link words or groups of words with very basic linear connectors like "and" or "then".

Appendix 5

Supplementary Grid with Criteria for A2+, B1+, B2+

	RANGE	ACCURACY	FLUENCY	INTERACTION	COHERENCE
C2					
C1					
B2+	Can express him/herself clearly and without much sign of having to restrict what he/she wants to say.	Shows good grammatical control; occasional "slips" or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.	Can communicate spontaneously, often showing remarkable fluency and ease of expression in even longer complex stretches of speech. Can use circumlocution and paraphrase to cover gaps in vocabulary and structure.	165-166- 94/261 Can intervene appropriately in discussion, exploiting a variety of suitable language to do so, and relating his/her own contribution to those of other speakers.	162 Can use a variety of linking words efficiently to mark clearly the relationships between ideas.
B2					
B1+	Has a sufficient range of language to describe unpredictable situations, explain the main points in an idea or problem with reasonable precision and express thoughts on abstract or cultural topics such as music and films.	Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influences.	Can express him/herself with relative ease. Despite some problems with formulation resulting in pauses and "cul-de-sacs", he/she is able to keep going effectively without help.	94/111 & 130 Can exploit a basic repertoire of strategies to keep a conversation or discussion going. Can give brief comments on others views during discussion. Can intervene to check and confirm detailed information.	<i>No descriptor available</i>
B1					
A2+	Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics, though he/she will generally have to compromise the message and search for words.	<i>No descriptor available</i>	Can adapt rehearsed memorised simple phrases to particular situations with sufficient ease to handle short routine exchanges without undue effort, despite very noticeable hesitation and false starts.	Can initiate, maintain and close simple, restricted face-to-face conversation, asking and answering questions on topics of interest, pastimes and past activities. Can interact with reasonable ease in structured situations, given some help, but participation in open discussion is fairly restricted.	Can use the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points.
A2					
A1					

Appendix 6

Analytic Rating Form: Draft Manual for "Relating language examinations to CEFR" - Form B2

Nom de l'étudiant:

Numéro de boîtier : _____

Niveaux : A1 A2 A2+ B1 B1+ B2 B2+ C1 C2

Si vous avez changé d'avis après discussion, veuillez indiquer pourquoi.

Analyse détaillée / Estimation à l'aide de la grille

ETENDUE	CORRECTION	AISANCE	INTERACTION	COHERENCE

Classement final

*Fiche analytique d'évaluation d'après celle de :
Eurocentres (North 1991/1992)/Projet suisse (Schneider and North 2000)*

Appendix 7

Supplementary CEFR Scales

PRODUCTION ORALE GÉNÉRALE	
C2	<i>Peut produire un discours élaboré, limpide et fluide, avec une structure logique efficace qui aide le destinataire à remarquer les points importants et à s'en souvenir.</i>
C1	<i>Peut faire une présentation ou une description d'un sujet complexe en intégrant des arguments secondaires et en développant des points particuliers pour parvenir à une conclusion appropriée.</i>
B2	<i>Peut méthodiquement développer une présentation ou une description soulignant les points importants et les détails pertinents.</i>
	<i>Peut faire une description et une présentation détaillées sur une gamme étendue de sujets relatifs à son domaine d'intérêt en développant et justifiant les idées par des points secondaires et des exemples pertinents.</i>
B1	<i>Peut assez aisément mener à bien une description directe et non compliquée de sujets variés dans son domaine en la présentant comme une succession linéaire de points.</i>
A2	<i>Peut décrire ou présenter simplement des gens, des conditions de vie, des activités quotidiennes, ce qu'on aime ou pas, par de courtes séries d'expressions ou de phrases non articulées.</i>
A1	<i>Peut produire des expressions simples isolées sur les gens et les choses.</i>

MONOLOGUE SUIVI : décrire l'expérience	
C2	<i>Peut faire des descriptions limpides et courantes, élaborées et souvent mémorables.</i>
C1	<i>Peut faire une description claire et détaillée de sujets complexes.</i>
	<i>Peut faire une description ou une narration élaborées, en y intégrant des thèmes secondaires, en développant certains points et en terminant par une conclusion appropriée.</i>
B2	<i>Peut faire une description claire et détaillée d'une gamme étendue de sujets en relation avec son domaine d'intérêt.</i>
B1	<i>Peut faire une description directe et simple de sujets familiers variés dans le cadre de son domaine d'intérêt.</i>
	<i>Peut rapporter assez couramment une narration ou une description simples sous forme d'une suite de points.</i>
	<i>Peut relater en détail ses expériences en décrivant ses sentiments et ses réactions.</i>
	<i>Peut relater les détails essentiels d'un événement fortuit tel un accident.</i>
	<i>Peut raconter l'intrigue d'un livre ou d'un film et décrire ses propres réactions.</i>
A2	<i>Peut décrire un rêve, un espoir ou une ambition.</i>
	<i>Peut décrire un événement, réel ou imaginaire.</i>
	<i>Peut raconter une histoire.</i>
	<i>Peut raconter une histoire ou décrire quelque chose par une simple liste de points. Peut décrire les aspects de son environnement quotidien tels que les gens, les lieux, l'expérience professionnelle ou scolaire.</i>
	<i>Peut faire une description brève et élémentaire d'un événement ou d'une activité.</i>
A1	<i>Peut décrire des projets et préparatifs, des habitudes et occupations journalières, des activités passées et des expériences personnelles.</i>
	<i>Peut décrire et comparer brièvement, dans une langue simple, des objets et choses lui appartenant.</i>
	<i>Peut expliquer en quoi une chose lui plaît ou lui déplaît.</i>
A1	<i>Peut décrire sa famille, ses conditions de vie, sa formation, son travail actuel ou le dernier en date.</i>
	<i>Peut décrire les gens, lieux et choses en termes simples.</i>
A1	<i>Peut se décrire, décrire ce qu'il/elle fait, ainsi que son lieu d'habitation.</i>

MONOLOGUE SUIVI : argumenter (par exemple, lors d'un débat)	
C2	<i>Pas de descripteur disponible</i>
C1	<i>Pas de descripteur disponible</i>
B2	<i>Peut développer méthodiquement une argumentation en mettant en évidence les points significatifs et les éléments pertinents.</i>
	<i>Peut développer une argumentation claire, en élargissant et confirmant ses points de vue par des arguments secondaires et des exemples pertinents.</i> <i>Peut enchaîner des arguments avec logique.</i>
	<i>Peut expliquer un point de vue sur un problème en donnant les avantages et les inconvénients d'options diverses.</i>
B1	<i>Peut développer une argumentation suffisamment bien pour être compris sans difficulté la plupart du temps.</i>
	<i>Peut donner brièvement raisons et explications relatives à des opinions, projets et actions.</i>
A2	<i>Pas de descripteur disponible</i>
A1	<i>Pas de descripteur disponible</i>

ANNONCES PUBLIQUES	
C2	<i>Pas de descripteur disponible</i>
C1	<i>Peut faire une annonce avec aisance, presque sans effort, avec l'accent et l'intonation qui transmettent des nuances fines de sens.</i>
B2	<i>Peut faire des annonces sur la plupart des sujets généraux avec un degré de clarté, d'aisance et de spontanéité qui ne procurent à l'auditeur ni tension ni inconfort.</i>
B1	<i>Peut faire de brèves annonces préparées sur un sujet proche des faits quotidiens dans son domaine, éventuellement même avec un accent et une intonation étrangères qui n'empêchent pas d'être clairement intelligible.</i>
A2	<i>Peut faire de très brèves annonces préparées avec un contenu prévisible et appris de telle sorte qu'elles soient intelligibles pour des auditeurs attentifs.</i>
A1	<i>Pas de descripteur disponible</i>

Note: Les descripteurs de cette sous-échelle n'ont pas été calibrés empiriquement.

S'ADRESSER A UN AUDITOIRE	
C2	<i>Peut présenter un sujet complexe bien construit avec assurance à un auditoire pour qui il n'est pas familier, en structurant et adaptant l'exposé avec souplesse pour répondre aux besoins de cet auditoire.</i> <i>Peut gérer un questionnement difficile, voire hostile.</i>
C1	<i>Peut faire un exposé clair et bien structuré sur un sujet complexe, développant et confirmant ses points de vue assez longuement à l'aide de points secondaires, de justifications et d'exemples pertinents.</i> <i>Peut gérer les objections convenablement, y répondant avec spontanéité et presque sans effort.</i>
B2	<i>Peut développer un exposé de manière claire et méthodique en soulignant les points significatifs et les éléments pertinents.</i>
	<i>Peut s'écarter spontanément d'un texte préparé pour suivre les points intéressants soulevés par des auditeurs en faisant souvent preuve d'une aisance et d'une facilité d'expression remarquables.</i>
	<i>Peut faire un exposé clair, préparé, en avançant des raisons pour ou contre un point de vue particulier et en présentant les avantages et les inconvénients d'options diverses.</i> <i>Peut prendre en charge une série de questions après l'exposé avec un degré d'aisance et de spontanéité qui ne cause pas de tension à l'auditoire ou à lui/elle-même.</i>
B1	<i>Peut faire un exposé simple et direct, préparé, sur un sujet familier dans son domaine qui soit assez clair pour être suivi sans difficulté la plupart du temps et dans lequel les points importants soient expliqués avec assez de précision.</i> <i>Peut gérer les questions qui suivent mais peut devoir faire répéter si le débit était rapide.</i>
A2	<i>Peut faire un bref exposé préparé sur un sujet relatif à sa vie quotidienne, donner brièvement des justifications et des explications pour ses opinions, ses projets et ses actes.</i> <i>Peut faire face à un nombre limité de questions simples et directes</i>
	<i>Peut faire un bref exposé élémentaire, répété, sur un sujet familier.</i> <i>Peut répondre aux questions qui suivent si elles sont simples et directes et à condition de pouvoir faire répéter et se faire aider pour formuler une réponse.</i>
A1	<i>Peut lire un texte très bref et répété, par exemple pour présenter un conférencier, proposer un toast.</i>

INTERACTION ORALE GÉNÉRALE	
C2	<i>Possède une bonne maîtrise d'expressions idiomatiques et de tournures courantes, avec une conscience du sens connotatif. Peut exprimer avec précision des nuances fines de signification, en utilisant assez correctement une gamme étendue de modalités. Peut revenir sur une difficulté et la restructurer de manière si habile que l'interlocuteur s'en rende à peine compte.</i>
C1	<i>Peut s'exprimer avec aisance et spontanéité, presque sans effort.. Possède une bonne maîtrise d'un vaste répertoire lexical lui permettant de surmonter facilement des lacunes par des périphrases avec apparemment peu de recherche d'expressions ou de stratégies d'évitement. Seul un sujet conceptuellement difficile est susceptible de gêner le flot naturel et fluide du discours.</i>
B2	<i>Peut utiliser la langue avec aisance, correction et efficacité dans une gamme étendue de sujets d'ordre général., éducationnel, professionnel et concernant les loisirs, en indiquant clairement les relations entre les idées. Peut communiquer spontanément avec un bon contrôle grammatical sans donner l'impression d'avoir à restreindre ce qu'il/elle souhaite dire et avec le degré de formalisme adapté à la circonstance</i>
	<i>Peut communiquer avec un niveau d'aisance et de spontanéité tel qu'une interaction soutenue avec des locuteurs natifs soit tout à fait possible sans entraîner de tension d'une part ni de autre. Peut mettre en valeur la signification personnelle de faits et d'expériences, exposer ses opinions et les défendre avec pertinence en fournissant explications et arguments.</i>
B1	<i>Peut communiquer avec une certaine assurance sur des sujets familiers habituels ou non en relation avec ses intérêts et son domaine professionnel. Peut échanger, vérifier et confirmer des informations, faire face à des situations moins courantes et expliquer pourquoi il y a une difficulté. Peut exprimer sa pensée sur un sujet abstrait ou culturel comme un film, des livres, de la musique, etc.</i>
	<i>Peut exploiter avec souplesse une gamme étendue de langue simple pour faire face à la plupart des situations susceptibles de se produire au cours d'un voyage. Peut aborder sans préparation une conversation sur un sujet familier, exprimer des opinions personnelles et échanger de l'information sur des sujets familiers, d'intérêt personnel ou pertinents pour la vie quotidienne (par exemple, la famille, les loisirs, le travail, les voyages et les faits divers)</i>
A2	<i>Peut interagir avec une aisance raisonnable dans des situations bien structurées et de courtes conversations à condition que l'interlocuteur apporte de l'aide le cas échéant.. Peut faire face à des échanges courants simples sans effort excessif ; peut poser des questions, répondre à des questions et échanger des idées et des renseignements sur des sujets familiers dans des situations familières prévisibles de la vie quotidienne.</i>
	<i>Peut communiquer dans le cadre d'une tâche simple et courante ne demandant qu'un échange d'information simple et direct sur des sujets familiers relatifs au travail et aux loisirs. Peut gérer des échanges de type social très courts mais est rarement capable de comprendre suffisamment pour alimenter volontairement la conversation..</i>
A1	<i>Peut interagir de façon simple, mais la communication dépend totalement de la répétition avec un débit plus lent, de la reformulation et des corrections. Peut répondre à des questions simples et en poser, réagir à des affirmations simples et en émettre dans le domaine des besoins immédiats ou sur des sujets très familiers.</i>

COMPRENDRE UN LOCUTEUR NATIF	
C2	<i>Peut comprendre tout locuteur natif, même sur des sujets spécialisés, abstraits ou complexes et hors de son domaine, à condition d'avoir l'occasion de s'habituer à une langue non standard ou à un accent.</i>
C1	<i>Peut comprendre en détail une intervention sur des sujets spécialisés abstraits ou complexes, même hors de son domaine, mais peut avoir besoin de faire confirmer quelques détails, notamment si l'accent n'est pas familier.</i>
B2	<i>Peut comprendre en détail ce qu'on lui dit en langue standard, même dans un environnement bruyant.</i>
B1	<i>Peut suivre un discours clairement articulé et qui lui est destiné dans une conversation courante, mais devra quelquefois faire répéter certains mots ou expressions.</i>
A2	<i>Peut comprendre suffisamment pour gérer un échange simple et courant sans effort excessif. Peut généralement comprendre un discours qui lui est adressé dans une langue standard clairement articulée sur un sujet familier, à condition de pouvoir demander de répéter ou reformuler de temps à autre.</i>
	<i>Peut comprendre ce qui lui est dit clairement, lentement et directement dans une conversation quotidienne simple à condition que l'interlocuteur prenne la peine de l'aider à comprendre.</i>
A1	<i>Peut comprendre des expressions quotidiennes pour satisfaire des besoins simples de type concret si elles sont répétées, formulées directement, lentement et clairement par un interlocuteur compréhensif. Peut comprendre des questions et des instructions qui lui sont adressées lentement et avec soin et suivre des consignes simples et brèves.</i>

CONVERSATION	
C2	<i>Peut converser de façon confortable et appropriée sans qu'aucune limite linguistique ne vienne empêcher la conduite d'une vie personnelle et sociale accomplie.</i>
C1	<i>Peut utiliser la langue en société avec souplesse et efficacité, y compris dans un registre affectif, allusif ou humoristique.</i>
B2	<i>Peut s'impliquer dans une conversation d'une certaine longueur sur la plupart des sujets d'intérêt général en y participant réellement, et ce même dans un environnement bruyant. Peut maintenir des relations avec des locuteurs natifs sans les amuser ou les irriter involontairement ou les obliger à se comporter autrement qu'ils ne le feraient avec un interlocuteur natif. Peut transmettre différents degrés d'émotion et souligner ce qui est important pour lui/elle dans un événement ou une expérience.</i>
B1	<i>Peut aborder sans préparation une conversation sur un sujet familier. Peut suivre une conversation quotidienne si l'interlocuteur s'exprime clairement, bien qu'il lui soit parfois nécessaire de faire répéter certains mots ou expression. Peut soutenir une conversation ou une discussion mais risque d'être quelquefois difficile à suivre lorsqu'il/elle essaie de formuler exactement ce qu'il/elle aimerait dire. Peut réagir à des sentiments tels que la surprise, la joie, la tristesse, la curiosité et l'indifférence et peut les exprimer.</i>
A2	<i>Peut établir un contact social : salutations et congé ; présentations ; remerciements. Peut généralement comprendre un discours standard clair, qui lui est adressé, sur un sujet familier, à condition de pouvoir faire répéter ou reformuler de temps à autre. Peut participer à de courtes conversations dans des contextes habituels sur des sujets généraux. Peut dire en termes simples comment il/elle va et remercier. Peut gérer de très courts échanges sociaux mais peut rarement soutenir une conversation de son propre chef bien qu'on puisse l'aider à comprendre si l'interlocuteur en prend la peine. Peut utiliser des formules de politesse simples et courantes pour s'adresser à quelqu'un ou le saluer. Peut faire et accepter une offre, une invitation et des excuses. Peut dire ce qu'il/elle aime ou non.</i>
A1	<i>Peut présenter quelqu'un et utiliser des expressions élémentaires de salutation et de congé . Peut demander à quelqu'un de se nouvelles et y réagir. Peut comprendre des expressions quotidiennes pour satisfaire à des besoins simples de type concret si elles sont répétées, formulées directement, clairement et lentement par un interlocuteur compréhensif.</i>

DISCUSSION INFORMELLE (ENTRE AMIS)	
C2	<i>Comme C1</i>
C1	<i>Peut suivre facilement des échanges entre partenaires extérieurs dans une discussion de groupe et un débat et y participer, même sur des sujets abstraits, complexes et non familiers.</i>
B2	<i>Peut suivre facilement une conversation animée entre locuteurs natifs. Peut exprimer ses idées et ses opinions avec précision. et argumenter avec conviction sur des sujets complexes et réagir de même aux arguments d'autrui.</i>
	<i>Peut participer activement à une discussion informelle dans un contexte familial, en faisant des commentaires, en exposant un point de vue clairement, en évaluant d'autres propositions, ainsi qu'en émettant et en réagissant à des hypothèses. Peut suivre, avec quelque effort, l'essentiel de ce qui se dit dans une conversation à laquelle il/elle ne participe pas mais peut éprouver des difficultés à participer effectivement à une conversation avec plusieurs locuteurs natifs qui ne modifient en rien leur mode d'expression. Peut exprimer et exposer ses opinions dans une discussion et les défendre avec pertinence en fournissant explications, arguments et commentaires.</i>
B1	<i>Peut suivre l'essentiel de ce qui se dit autour de lui sur des thèmes généraux, à condition que les interlocuteurs évitent l'usage d'expressions trop idiomatiques et articulent clairement. Peut exprimer sa pensée sur un sujet abstrait ou culturel comme un film ou de la musique. Peut expliquer pourquoi quelque chose pose problème. Peut commenter brièvement le point de vue d'autrui. Peut comparer et opposer des alternatives en discutant de ce qu'il faut faire, où il faut aller, qui désigner, qui ou quoi choisir, etc..</i>
	<i>Peut, en règle générale, suivre les points principaux d'une discussion d'une certaine longueur se déroulant en sa présence à condition qu'elle ait lieu en langue standard clairement articulée. Peut émettre ou solliciter un point de vue personnel ou une opinion sur des points d'intérêt général. Peut faire comprendre ses opinions et réactions pour trouver une solution à un problème ou à des questions pratiques relatives à où aller, que faire, comment organiser (une sortie, p. ex.) Peut exprimer poliment ses convictions, opinions, accord et désaccord.</i>
A2	<i>Peut généralement reconnaître le sujet d'une discussion extérieure si elle se déroule lentement et clairement. Peut discuter du programme de la soirée ou du week-end. Peut faire des suggestions et réagir à des propositions. Peut exprimer son accord ou son désaccord à autrui.</i>
	<i>Peut discuter simplement de questions quotidiennes si l'on s'adresse directement à lui/elle, clairement et simplement. Peut discuter de l'organisation d'une rencontre et de ses préparatifs.</i>
A1	<i>Pas de descripteur disponible</i>

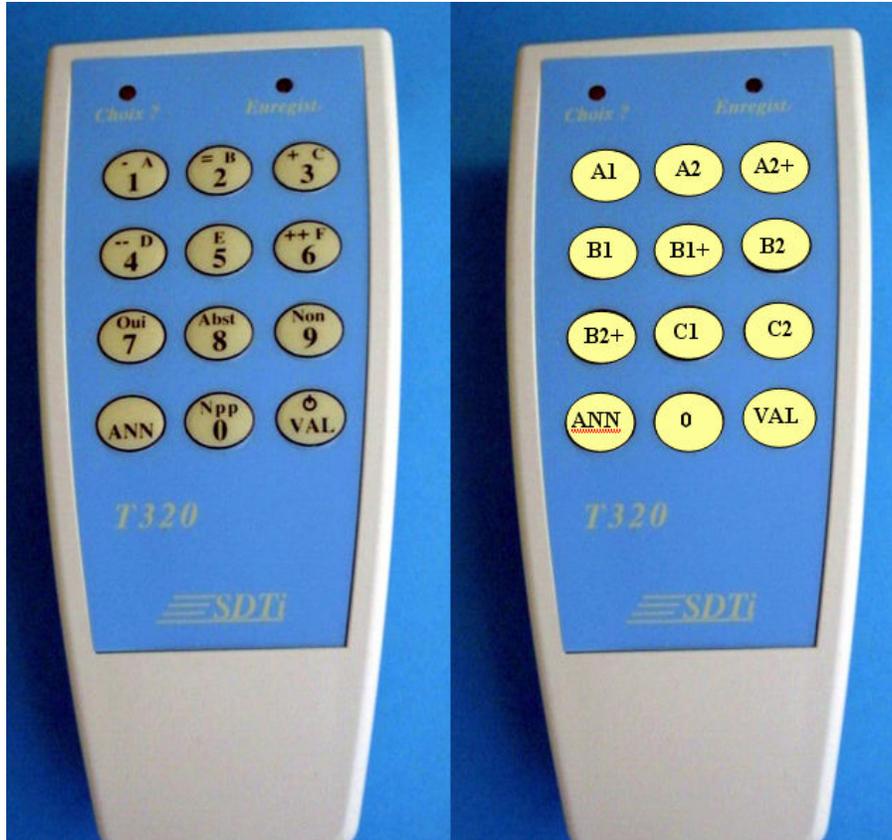
ÉCHANGE D'INFORMATION	
C2	<i>Comme B2</i>
C1	<i>Comme B2</i>
B2	<p><i>Peut comprendre et échanger une information complexe et des avis sur une gamme étendue de sujets relatifs à son rôle professionnel.</i></p> <p><i>Peut transmettre avec sûreté une information détaillée.</i></p> <p><i>Peut faire la description claire et détaillée d'une démarche.</i></p> <p><i>Peut faire la synthèse d'informations et d'arguments issus de sources différentes et en rendre compte.</i></p>
B1	<p><i>Peut échanger avec une certaine assurance un grand nombre d'informations factuelles sur des sujets courants ou non, familiers à son domaine.</i></p> <p><i>Peut expliquer comment faire quelque chose en donnant des instructions détaillées.</i></p> <p><i>Peut résumer - en donnant son opinion - un bref récit, un article, un exposé, une discussion, une interview ou un documentaire et répondre à d'éventuelles questions complémentaires de détail.</i></p> <p><i>Peut trouver et transmettre une information simple et directe.</i></p> <p><i>Peut demander et suivre des directives détaillées.</i></p> <p><i>Peut obtenir plus de renseignements</i></p>
A2	<p><i>Peut comprendre suffisamment pour communiquer sur des sujets familiers et simples sans effort excessif.</i></p> <p><i>Peut se débrouiller avec les demandes directes de la vie quotidienne : trouver une information factuelle et la transmettre.</i></p> <p><i>Peut répondre à des questions et en poser sur les habitudes et les activités journalières.</i></p> <p><i>Peut répondre à des questions sur les loisirs et les activités passées et en poser.</i></p> <p><i>Peut donner et suivre des directives et des instructions simples comme, par exemple, comment aller quelque part.</i></p> <p><i>Peut communiquer dans le cadre d'une tâche simple et routinière ne demandant qu'un échange d'information simple et direct.</i></p> <p><i>Peut échanger une information limitée sur des sujets familiers et des opérations courantes</i></p> <p><i>Peut poser des questions et y répondre sur le travail et le temps libre.</i></p> <p><i>Peut demander et expliquer son chemin à l'aide d'une carte ou d'un plan.</i></p> <p><i>Peut demander et fournir des renseignements personnels.</i></p>
A1	<p><i>Peut comprendre des questions et des instructions qui lui sont adressées lentement et avec soin et suivre des instructions simples et brèves.</i></p> <p><i>Peut répondre à des questions simples et en poser ; peut réagir à des déclarations simples et en faire, dans des cas de nécessité immédiate ou sur des sujets très familiers.</i></p> <p><i>Peut poser des questions personnelles, par exemple sur le lieu d'habitation, les personnes fréquentées et les biens, et répondre au même type de questions.</i></p> <p><i>Peut parler du temps avec des expressions telles que la semaine prochaine, vendredi dernier, en novembre, à 3 heures.</i></p>

INTERVIEWER ET ÊTRE INTERVIEWÉ (l'entretien)	
C2	<i>Peut tenir sa part du dialogue extrêmement bien, en structurant le discours et en échangeant avec autorité et une complète aisance, que ce soit comme interviewer ou comme interviewé, de la même manière qu'un locuteur natif.</i>
C1	<i>Peut participer complètement à un entretien comme interviewer ou comme interviewé, en développant et mettant en valeur le point discuté, couramment et sans aucune aide, et en utilisant les interjections convenablement.</i>
B2	<i>Peut conduire un entretien avec efficacité et aisance, en s'écartant spontanément des questions préparées et en exploitant et relançant les réponses intéressantes.</i>
	<i>Peut prendre des initiatives dans un entretien, élargir et développer ses idées, sans grande aide ni stimulation de la part de l'interlocuteur.</i>
B1	<i>Peut fournir des renseignements concrets exigés dans un entretien ou une consultation (par exemple, décrire des symptômes à un médecin) mais le fait avec une précision limitée.</i>
	<i>Peut conduire un entretien préparé, vérifier et confirmer les informations, bien qu'il lui soit parfois nécessaire de demander de répéter si la réponse de l'interlocuteur est trop rapide ou trop développée.</i>
	<i>Peut prendre certaines initiatives dans une consultation ou un entretien (par exemple introduire un sujet nouveau) mais reste très dépendant de l'interviewer dans l'interaction.</i> <i>Peut utiliser un questionnaire préparé pour conduire un entretien structuré, avec quelques questions spontanées complémentaires.</i>
A2	<i>Peut se faire comprendre dans un entretien et communiquer des idées et de l'information sur des sujets familiers à condition de pouvoir faire clarifier à l'occasion et d'être aidé pour exprimer ce qu'il/elle veut.</i>
	<i>Peut répondre à des questions simples et réagir à des déclarations simples dans un entretien.</i>
A1	<i>Peut répondre dans un entretien à des questions personnelles posées très lentement et clairement dans une langue directe et non idiomatique.</i>

MAÎTRISE DU SYSTÈME PHONOLOGIQUE	
C2	<i>Comme C1</i>
C1	<i>Peut varier l'intonation et placer l'accent phrastique correctement afin d'exprimer de fines nuances de sens.</i>
B2	<i>A acquis une prononciation et une intonation claires et naturelles.</i>
B1	<i>La prononciation est clairement intelligible même si un accent étranger est quelquefois perceptible et si des erreurs de prononciation proviennent occasionnellement.</i>
A2	<i>La prononciation est en général suffisamment claire pour être comprise malgré un net accent étranger mais l'interlocuteur devra parfois faire répéter.</i>
A1	<i>La prononciation d'un répertoire très limité d'expressions et de mots mémorisés est compréhensible avec quelque effort pour un locuteur natif habitué aux locuteurs du groupe linguistique de l'apprenant/utilisateur.</i>

Appendix 8

Electronic Keypad and CEFR Levels



Appendix 9

Familiarisation Exercise: identify level of descriptors

Nom : _____

Descripteurs de la production orale et de l'interaction orale Attribuer un niveau à chaque descripteur proposé

1.	Peut interagir avec une aisance raisonnable dans des situations bien structurées et de courtes conversations à condition que l'interlocuteur apporte de l'aide le cas échéant. Peut faire face à des échanges courants simples sans effort excessif ; peut poser des questions, répondre à des questions et échanger des idées et des renseignements sur des sujets familiers dans des situations familières prévisibles de la vie quotidienne.	A2+
2.	Peut assez aisément mener à bien une description directe et non compliquée de sujets variés dans son domaine en la présentant comme une succession linéaire de points.	B1
3.	Peut méthodiquement développer e présentation ou une description soulignant les points importants et les détails pertinents.	B2+
4.	Peut interagir de façon simple, mais la communication dépend totalement de la répétition avec un débit plus lent, de la reformulation et des corrections. Peut répondre à des questions simples et en poser, réagir à des affirmations simples et en émettre dans le domaine des besoins immédiats ou sur des sujets très familiers.	A1
5.	Peut produire des expressions simples isolées sur les gens et les choses.	A1
6.	Peut communiquer avec une certaine assurance sur des sujets familiers habituels ou non en relation avec ses intérêts et son domaine professionnel. Peut échanger, vérifier et confirmer des informations, faire face à des situations moins courantes et expliquer pourquoi il y a une difficulté. Peut exprimer sa pensée sur un sujet abstrait ou culturel comme un film, des livres, de la musique, etc.	B1+
7.	Peut faire une présentation ou une description d'un sujet complexe en intégrant des arguments secondaires et en développant des points particuliers pour parvenir à une conclusion appropriée.	C1
8.	Peut décrire ou présenter simplement des gens, des conditions de vie, des activités quotidiennes, ce qu'on aime ou pas, par de courtes séries d'expressions ou de phrases non articulées.	A2

9.	Peut faire une description et une présentation détaillées sur une gamme étendue de sujets relatifs à son domaine d'intérêt en développant et justifiant les idées par des points secondaires et des exemples pertinents.	B2
10.	Peut communiquer avec un niveau d'aisance et de spontanéité tel qu'une interaction soutenue avec des locuteurs natifs soit tout à fait possible sans entraîner de tension d'une part ni d'autre. Peut mettre en valeur la signification personnelle de faits et d'expériences, exposer ses opinions et les défendre avec pertinence en fournissant explications et arguments.	B2
11.	Peut produire un discours élaboré, limpide et fluide, avec une structure logique efficace qui aide le destinataire à remarquer les points importants et à s'en souvenir.	C2
12.	Peut communiquer dans le cadre d'une tâche simple et courante ne demandant qu'un échange d'information simple et direct sur des sujets familiers relatifs au travail et aux loisirs. Peut gérer des échanges de type social très courts mais est rarement capable de comprendre suffisamment pour alimenter volontairement la conversation.	A2
13.	Possède une bonne maîtrise d'expressions idiomatiques et de tournures courantes, avec une conscience du sens connotatif. Peut exprimer avec précision des nuances fines de signification, en utilisant assez correctement une gamme étendue de modalités. Peut revenir sur une difficulté et la restructurer de manière si habile que l'interlocuteur s'en rende à peine compte.	C2
14.	Peut exploiter avec souplesse une gamme étendue de langue simple pour faire face à la plupart des situations susceptibles de se produire au cours d'un voyage. Peut aborder sans préparation une conversation sur un sujet familier, exprimer des opinions personnelles et échanger de l'information sur des sujets familiers, d'intérêt personnel ou pertinents pour la vie quotidienne (par exemple, la famille, les loisirs, le travail, les voyages et les faits divers).	B1
15.	Peut s'exprimer avec aisance et spontanéité, presque sans effort. Possède une bonne maîtrise d'un vaste répertoire lexical lui permettant de surmonter facilement des lacunes par des périphrases avec apparemment peu de recherche d'expressions ou de stratégies d'évitement. Seul un sujet conceptuellement difficile est susceptible de gêner le flot naturel et fluide du discours.	C1
16.	Peut utiliser la langue avec aisance, correction et efficacité dans une gamme étendue de sujets d'ordre général, éducationnel, professionnel et concernant les loisirs, en indiquant clairement les relations entre les idées. Peut communiquer spontanément avec un bon contrôle grammatical sans donner l'impression d'avoir à restreindre ce qu'il/elle souhaite dire et avec le degré de formalisme adapté à la circonstance.	B2+

Annexe 10

CEFR Cut-offs on the underlying Statistical Scale

Table 7.2: Equal Interval Levels and Common Reference Levels

Common Reference Levels	Finer Level (Swiss)	Abbr	Cut-off	Range on Scale	
C2	Mastery	Mastery	M	3.90	
C1	Effective Operational Proficiency <i>(Vantage Plus)</i>	Full Effectiveness	EOP	2.80	1.10
B2	Vantage <i>(Threshold Plus)</i>	Full Independence	V	0.72	1.02
B1	Threshold <i>(Waystage Plus)</i>	Independence	T+	-0.26	0.98
		Threshold	T	-1.23	0.97
		Waystage Plus	W+	-2.21	0.98
A2	Waystage	Waystage	W	-3.23	1.02
A1	Breakthrough	Breakthrough	B	-4.29	1.06
	-----	Tourist	Tour	-5.39	1.10
	-----	Smattering			

Appendix 11

Summary of Questionnaire Feedback (completed by 35 participants)

1. Quels niveaux du Cadre européen commun de référence pour les langues connaissiez-vous le mieux avant ce séminaire ?

Très variable.

- 14 personnes disent très bien connaître les niveaux A1 et A2
- 21 personnes disent très bien connaître le niveau B1
- 14 personnes disent très bien connaître le niveau B2
- 7 personnes disent très bien connaître le niveau C1
- 1 personne dit très bien connaître le niveau C2
- 3 personnes disent très bien connaître les niveaux C
- 9 personnes disent très bien connaître tous les niveaux

2. Qu'avez-vous appris en ce qui concerne le Cadre européen commun de référence pour les langues ? Est-ce que ce séminaire a changé votre manière d'apprécier les niveaux du CECR ?

- Affiner l'appréciation des niveaux (9)
- L'utilisation des niveaux plus (6)
- Distinction entre niveau acquis et en cours d'acquisition (4)
- Appliquer la théorie à la pratique (4)
- Précision sur le contenu de certains descripteurs
- « Reciblage » sur les éléments pertinents des descripteurs. Réalisation que le niveau A1 est moins exigeant que prévu.
- Comprendre que l'internalisation des descripteurs chez les examinateurs est une condition pour arriver à un consensus
- Précisions sur ce que recouvrent les mots « étendue », « aisance »...

3. Qu'est-ce qui vous a semblé le plus utile dans ce séminaire ?

- Les discussions de groupes (20)
- La concertation pour arriver à un consensus
- Confronter sa propre évaluation à celle des autres
- La discussion en groupe et en plénière – et la justification de notes données.

- La familiarisation x
- Affiner l'utilisation des grilles
- Clarifier les mots-clés des critères
- La reprise systématique des critères
- L'obligation de se référer à des descripteurs et non pas à des pratiques. De justifier nos choix.
- Travailler avec les 3 documents (tableau 3, niveaux +, échelles de l'oral)

- Les explications de Brian du jeudi matin sur les enjeux
- Les précisions de Brian sur les grilles (4)
- Les interventions de Brian pour recadrer les discussions et harmoniser les échanges
- Les interventions de Brian et les échanges entre collègues
- Rencontrer d'autres personnes qui travaillent dans le même domaine
- La dimension internationale

4. Qu'est-ce qui vous a semblé le plus difficile à faire ?

- Evaluer les 1ères séquences
- A attribuer un niveau sans une 2^{ème} écoute (2)
- Saisir les nuances entre les descripteurs proches
- Interpréter de manière standardisée les descripteurs
- Objectiver l'analyse, ne pas se laisser trop influencer par des paramètres difficiles à analyser. Par ex dans le cas de AMMER.
- Evaluer les niveaux A1, A2, A2+
- Différencier les niveaux B1 et B1+/ B2 et B2+/ C1 et C2
- Comprendre la différence entre les niveaux et les niveaux +
- Commencer l'évaluation des performances par les cinq critères détaillés comme : Étendue, Interaction, Correction, etc.
- Distinguer entre les niveaux et les niveaux plus.
- Le manque de critères pour la prononciation. Pas acceptable !!!
- Décider d'un niveau par rapport au niveau acquis. Certaines séquences semblent intéressantes pour illustrer des niveaux en cours d'acquisition (2)
- Arriver à un consensus (2)
- La prise en compte ou non de la prononciation (2)
- Estimer un niveau lorsque la tâche ne permet pas de mettre à l'épreuve tous les critères (notamment argumentation)
- Evaluer 2 étudiants en même temps (2)
- Changer nos pratiques d'évaluation souvent négatives (2)
- Trop de flou dans le libellé de certains critères
- Rester objectif quand on a connu les étudiants et qu'ils ont progressé depuis

5. Si vous deviez organiser un séminaire semblable, que feriez-vous différemment ?

- Proposer une formation plus importante au début du séminaire (5)
- Eclaircissement en début de séminaire sur certains termes dans les descripteurs et sur les critères « étendue », « aisance », etc. (5)
- Plus d'information le premier jour
- Une demie-journée de mise en route
- Présenter de manière très claire les étapes de l'épreuve (préparées / non préparées) (3)
- Préciser les consignes en ce qui concerne les objectifs, surtout par rapport à l'évaluation en fonction d'un niveau acquis.

- Revoir les niveaux du Cadre afin d'être certain de parler un langage commun
- Plus de temps pour la familiarisation (3)
- Plus de temps pour les discussions (6)
- Plus de temps pour la relecture des descripteurs
- Ecouter les séquences 2 fois (5)
- Proposer moins de séquences (3)
- Avoir une journée supplémentaire pour le séminaire, plus de temps en général (4)
- Séparer les niveaux A B et C (2)
- Commencer par demander aux participant de mettre les candidats sur un des trois niveaux : A, B ou C – et leur demander de justifier leur évaluation. Ensuite distinguer entre A et A1, ou B et B1 ou C et C1. Puis travailler avec les critères détaillés pour décider si c'est un niveau plus ou non.
- Regrouper les niveaux + et le tableau 3 sur une même feuille (3)
- Réflexion sur les stratégies pour évaluer
- Organiser les groupes à l'avance
- Salle plus propice aux travaux de groupes avec des tables (4)
- Utiliser un tableau pour permettre de visualiser les arguments et contre-arguments
- Rien

6. Comment pensez-vous mettre à profit l'expérience acquise ici ? Pourrez-vous en faire bénéficier d'autres personnes lorsque le DVD produit pour le Conseil de l'Europe sera disponible ?

- Formation d'examineurs et de profs et de concepteurs (15)
- Partager avec les collègues (3)
- Poursuivre la réflexion sur les pratiques de l'évaluation (2)
- Mieux calibrer les programmes et les examens (2)
- Utiliser la vidéo plus systématiquement dans l'évaluation
- Organiser un séminaire du même type mais sur une journée (2)
- Organiser un séminaire du même type pour l'espagnol et l'allemand
- Prendre en compte l'expérience acquise pour la production des exemples de performances orales et pour l'organisation d'un séminaire du même type.
- Envisager d'organiser des séances de formation au CELV (Graz) à partir de 2008

Remarques supplémentaires (non sollicitées par une question) :

- Nombreux commentaires sur la qualité de l'organisation
- Nombreux remerciements