



CENTRO VALUTAZIONE
CERTIFICAZIONE LINGUISTICA
Università per Stranieri Perugia



COUNCIL OF EUROPE CONSEIL DE L'EUROPE

January 2006

DGIV/EDU/LANG (2006) 6

**Seminar to calibrate examples of spoken performance
in Italian L2**
to the scales of the
Common European Framework of Reference for Languages

UNIVERSITA' PER STRANIERI DI PERUGIA - CVCL
(Centro per la Valutazione e la Certificazione Linguistica)

Perugia, 16th-17th December 2005

REPORT

By

Giuliana GREGO BOLLI

Language Policy Division

1. Background of the Seminar

The Language Policy Division of the Council of Europe asked its partners to produce a set of tools to give examples, in different languages and for different language abilities, useful in illustrating the levels of the *Common European Framework of Reference*(CEFR).

Regarding the spoken performance, two video cassettes have already been produced, in the nineties, by the Eurocentres Foundation and the Ecoles-club Migros (within the Swiss Research Project) and by Cambridge ESOL in 2004, for English¹. In 2005 a DVD with examples of spoken performances in French was also produced by the Foundation Eurocentres and the CIEP (International Centre for Pedagogical Studies). The Language Policy Division of the Council of Europe and the CVCL (Centre for Assessment and Language Certification) of the 'Università per Stranieri' in Perugia agreed that a similar DVD was also needed for Italian. Italian, as a mother language, is one of the five most spoken languages in the European Union.

The general purpose of the videos is to disseminate examples of oral performance in different languages calibrated on the CEFR scale, following well defined procedures and according to the judgement of experts.

2. Aims of the Seminar in Perugia

The CVCL (Centre for Assessment and Language Certification) organised the seminar in order to produce (jointly with the Language Policy Division of the Council of Europe) a DVD² with examples of oral performances in Italian that could offer a concrete representation of the CEFR language descriptors. The seminar was part of the empirical work done so far by the CVCL in piloting the Manual: *Relating Language Examinations to the CEFR*³.

The seminar organised in Perugia also intended to give a contribution, within the context of Italian L2, to a more consistent interpretation of the CEFR levels as well as demonstrating the possibility of using the CEFR scale descriptors and concretely working with them.

Similar seminars had previously been organised, for French and German, by the CIEP in Sèvres (December 2004) and by the Goëthe Institut in Munich (October 2005), following the procedures listed in Chapter 5 of the Manual in order to standardise judgements.

The methodology for running such a piloted seminar in Sèvres first, and in Munich afterwards, was fundamental in order to give practical input and guide in organising the seminar in Perugia.

3. The procedures and the Seminar's organisation

The procedures listed in Chapter 5 of the Manual recommend that participants 'are made familiar with CEFR, analyse and assess oral performances and reach consensus in terms of assigning them to a CEFR level'. Appropriate documents and directions should be provided during the seminar by the coordinator.

The seminar in Perugia, as the ones in Sèvres and Munich, was organised according to these procedures. It lasted for two days (the programme is reported in Appendix 1), starting the 16th (9.00 a.m.) and ending the 17th (6.30 p.m.). Four familiarisation activities were run; 24 oral performances were analysed by 25 participants and all the nine levels (A1, A2, A2+, B1, B1+, B2, B2+, C1, C2) were assigned.

¹ Both videos have been recently copied to a DVD format

² The DVD will be part of a 'kit' in different languages aimed at providing a concrete interpretation of the capacities described in the CEFR scales.

³ The CVCL is currently in the process of relating its main-suite examinations to the levels of the CEFR. A CD Rom with examples of reading and listening items in Italian provided by the CVCL was published by the Council of Europe in spring 2005. Examples of written production provided by several ALTE partners, among which the CVCL, will be published soon.

The seminar had two coordinators: prof. G. Grego Bolli and dr. Francesca Pelliccia⁴. Before the familiarisation activities began, the seminar coordinators explained the purpose of the seminar (which had already been explained in the letter of invitation that each participant had received) and what the participants were asked to do in relation to the rating.

3.1 The Familiarisation

Familiarisation activities should ensure that the participants have a clear and common understanding of the CEFR both in terms of categories of description and common reference levels. Four activities were carried out at the beginning of the seminar using the following materials⁵ (see Appendix 2) under the constant guidance of the two coordinators:

- Table 1 of the CEFR. The participants were asked to read it quickly and think about the salient features distinguishing each level from the highest to the lower one. Then one of the two coordinators in plenary and using an OHP, where she had already highlighted the most important features, lead the discussion focusing on the characteristics representative of each level.
- Table 5.4 of the Manual. The participants worked in groups of four/five in order to detect the salient features of each level, what distinguishes it from the higher and the lower level, in relation to the Oral production. Each group reported the results that were discussed in plenary and written on a board by the coordinators.
- Two Scales of the CEFR: *Overall spoken Interaction* and *Sustained Monologue: Describing experience*. The participants received two envelopes with all the descriptors mixed up. They worked in groups of four/five to put them in the original order. The results were discussed in plenary, using OHP with the original scales. The coordinators lead the discussion focusing on how they manage to link the descriptors to the levels.
- Table 3 of the CEFR (Table 5.5 of the Manual) and a supplementary grid giving descriptors for the “plus” levels. The participants had to analyse them carefully detecting what they considered particularly important in the light of the consensus they reached about the levels during the previous activities. The participants exchanged opinions with their neighbours and with the coordinators who were walking around answering questions and trying to clarify possible doubts.

All the participants were asked to read and study, before coming to the seminar: Table 3 of the CEFR and the CEFR scales that would not be specifically treated during the Familiarisation activities in the seminar: Overall Spoken Production, Sustained Monologue: (putting a case), Conversation, Informal Discussion (with friends), Formal Discussion and Meetings.

3.2 The Performances

The performances were video recorded between June and November 2005 by a professional video production company at the ‘Università per Stranieri’ in Perugia’, after the CELI examinations. The recordings shown at the seminar had performances from adults (mainly young) from 17 countries: Poland, Japan, Russia, United States of America, Venezuela, Cameroon, Austria, The Netherlands, Cyprus, China, Korea, Brasil, Syria, Romania, Spain, Germany, United Kingdom. The possibility of filming students abroad, in the local CVCL Examination Centres, was abandoned because of the difficulties put forward by the Centres for organising the filming, without an economical support from Perugia. The CVCL budget for the seminar was not enough to cover such expenses.

⁴ They both work at the CVCL, are used to working together and with the CEFR and have quite a large experience as teacher trainers.

⁵ The materials were included in a folder given to each participant at the beginning of the seminar.

3.3 The students' selection

All the students filmed were selected in a systematic way and their language level in relation to the CEFR levels was estimated by class observation and teacher's evaluation, if they were students at the University; by exam results, if they took a CELI exam. All the students were also interviewed by the seminar's coordinators to have a clearer idea of their possible performance in relation to the CEFR levels, to the format and the tasks suggested for the videos.

Each student had signed, before being filmed, a form authorising inclusion and dissemination of their image on a standardised DVD and on internet (see Appendix 3).

3.4 The videos' format and 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 English video circulating in April 2004 and were recommended in the Council of Europe's Brief of Recording." (North, B & S. Lepage 2005. Seminar to calibrate examples of spoken performances in line with the scales of the *Common European Framework of Reference for Languages*, Strasbourg: Council of Europe: p. 5). The format set shows pairs of students.

Each recording shows two learners and consists of the following phases:

- A production phase by the first learner with a sustained monologue, which in many cases is followed by questions from the other learner.
- A similar production phase by the second learner.
- A spontaneous interaction between the two learners.

At the beginning of the videos, shown at the Perugia seminar, the two learners in turn shortly introduce themselves. The reason for this was to reduce the stress due to the filming (that in some cases was quite high at the beginning) as well as to give the learners the opportunity of knowing each other better and consequently to create a more familiar and relaxed atmosphere.

This same format was used at all levels. The production phase was "semi prepared". The learners chose a topic from a list and had about 10 minutes to prepare themselves. The interaction phase was spontaneous: the learners together selected a topic from a list and immediately started to talk about it. By getting the learners involved in the choice of the topic (see Appendix 4 for the list of the topics) in a collaborative way, allowed them to discard topics that did not interest them.

A mother language speaker was present during the filming not as an interlocutor, but just to introduce the topic chosen by the learners helping them only in case of serious breakdown in the communication.

The average duration of each recording, shown during the seminar, was around 15 minutes. All the recordings were shown integrally during the seminar.

Some of the students in the videos were also filmed during the oral production in the CELI exams. Possible differences in their performance due to different tasks and different performance conditions will be analysed later on and will become part of the training programme of the examiners.

All the students filmed in both versions were asked to fulfil a questionnaire specifying which format (in relation to their performance and other aspects) they preferred and why.

3.5 Assigning the levels: rating instruments and procedures

The main rating instruments were: Table 5.4 (see Appendix 5) translated into Italian from the Manual and Table 3 from the CEFR Italian version (see Appendix 6).

A supplementary grid, translated into Italian, (see Appendix 7) was also furnished, giving descriptors for the “plus” levels: A2+, B1+ and B2+ already mentioned. The addition of these three “plus” levels provided 9 levels overall. “This 9 levels reflect the linear scale produced in the Swiss research project.” (North & Lepage 2005: 5). The participants were suggested to consult, before rating and voting, also other scales: Overall Spoken Production, Overall Spoken Interaction, Sustained Monologue: Describing experience. These were included in a second folder with a set of materials to be used during the rating. The rating material was distributed to all the participants after the Familiarisation activities.

Rating was done first on paper adapting Form B2 (analytic rating form) from the Manual in two versions with two different colours: pink and green; the first one to be used before the discussion while the latter one after the discussion (see Appendix 8). Votes were then recorded electronically with CEFR levels corresponding to buttons on the keypad (see Appendix 9).

Accordingly to the conclusions made in the Report of the first seminar in Sèvres (North & Lepage 2005) and following the procedures adopted in Munich, it was decided to reverse the original procedure, going from a global assessment (using Table 5.4) to an analytic one (using Table 3 and the “plus” levels grid).

The same basic pattern of rating was followed:

1. Watch the sequence (Production: Learner A; Production: Learner B; Interaction between Learner A and B).
2. Consult Table 5.4 and other scales provided, reflect and rate individually the Global Level for both learners on Form B2 (pink version)
3. Consult criteria grids, reflect and record on Form B2 (pink version) the rating for both learners for the 5 criteria of the grids: Range (Estensione), Accuracy (Accuratezza), Fluency (Fluenza), Interaction (Interazione) and Coherence (Coerenza).
4. Electronic voting: individual votes: Range (Estensione), Accuracy (Accuratezza), Fluency (Fluenza), Interaction (Interazione) and Coherence (Coerenza), Global Level (Livello Globale).
5. View Histogram of the individual global judgement
6. Plenary discussion
7. View of the Excel table of the individual votes for each of the five criteria for learner A. Plenary discussion. View of the Excel table of the individual votes for each of the five criteria for learner B. Plenary discussion⁶.
8. Reflect and rate individually the final Global Level for learner A and for Learner B on Form B2 (green version).
9. Electronic Voting: Global (Livello Globale)
10. View Histogram of global judgements after discussion.

Only the most problematic recordings were shown twice.

⁶ For most of the sequences.

The decision to not split into small groups for the discussion, was taken principally for time constraints. In addition to that, it was clear, since the conclusion of the Familiarisation activities, that the group functioned very well as a whole: the participants seemed to share quite a common understanding of the levels. The concentration was high as was the sense of responsibility. The atmosphere was very collaborative. Each participant was willing to give his/her personal contribution and share opinions with the other colleagues.

The results of the rating will be statistically analysed as it was in Sèvres and Munich and the results will be commented and published.

3.6 Training in the rating procedures

The first three sequences were intended to represent the three CEFR broad levels: B, C, A (in the order they were presented). Even if only the first evaluation sequence was really used as training, more time was spent for each of the three finding out according to which descriptors the participants assigned the level, commenting the analysis of the Excel table shown after the first voting, reading descriptors and discussing about difficulties they had met and doubts they had had. The intention was to use the first three sequences as ‘broad’ points of reference for the following rating.

The first evaluation sequence was a form of training of the participants in the rating and voting procedures. This task aimed, as North and Lepage pointed out: “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.” (North & Lepage 2005: 7).

Participants were asked to watch a few minutes of the production phase of two learners: Marta and Megumi (considered to be B1) and decide first, as a group, the broad level (A, B, C) of each learner. When they agreed on that, the procedures were followed as described in point 3.5.

During the discussion participants were asked, by the seminar coordinators, to read the descriptors they focussed on in order to link the performances to the level, both in relation to the scales and to the criteria in the grids. The descriptors were read, commented and discussed in relation to each performance; as well as the descriptors for the level above and below. The possibility of using, as training, only some criteria was abandoned so as not to run the risk for the participants of getting used only to those criteria.

In order to clarify, as far as possible, both the procedures and criteria, the training phase was quite fruitful. It took more than one hour just for the first sequence.

4. The participants

Twenty four (24)⁷ of the twenty five (25) voting participants (see Appendix 9) to the calibration seminar in Perugia are or have been teachers of Italian L2, in different school and university contexts, in Italy and abroad. All have quite a high both theoretical and practical competence in the field of teaching learning Italian as L2. They differ in experience and age. In particular:

Four (4) are curriculum planners and language coordinators as well.

Three (3) are experts in the CEFR levels working in the Piloting project and in the Reference Levels Description Project.

⁷ Dr José Pascoal has represented the University of Lisbon that is going to organise a benchmarking seminar this year.

Some of them (14) have also been examiners of the Oral Production in the CELI examinations (as well as in other Italian certification systems), some are going to be examiners of the CELI exams in the near future.

Two (2) are item writers and examiner trainers at the CVCL.

The great majority (17) has already been trained to use the CEFR descriptors during the seminars that the CVCL organised regularly in Perugia and abroad.

In addition to the voting participants there were also other categories (see Appendix 11) represented at the Perugia seminar:

one representative of the Italian Ministry of Education (who joined the seminar only the second day);

one representative of the CoE (who joined the seminar only the second day);

one expert on statistical analysis from Cambridge ESOL;

the colleagues at the CVCL of the Università per Stranieri di Perugia.

After the seminar's conclusion, Saturday afternoon, the participants were asked to answer a questionnaire (the same submitted in Sèvres) on the seminar's organisation whose results will be analysed later on and published.

The CVCL is very grateful to all the participants who spent, in many cases, four days in order to participate in the seminar paying, on their own, travelling and accommodation expenses.

Conclusion and Remarks

The seminar was undoubtedly an interesting and fruitful experience for all those involved, as any kind of event which implies a collaborative exchange of competences and experiences among people working in the same field in different contexts.

The two different levels of aims of an international benchmarking seminar, that North and Lepage pointed out:

“1. **Calibrate and document performances**

2. **Establish a consensus in interpreting the CEFR levels in the pedagogic culture(s) concerned**” (North & Lepage 2005: 3)

imply starting to do some empirical work with the CEFR descriptors, which was really needed in the context of Italian L2 where most of the teachers, curriculum planners, examiners have only a very general idea of the CEFR and only a probably vague, theoretical and superficial knowledge of the descriptors. All the participants in the seminar had this important opportunity and enjoyed the experience, hopefully they will reproduce it in their own contexts.

The major problem met during the seminar was the lack of time. More time would have been needed for each phase, but realistically it would have been impossible to ask participants to stay more in Perugia at their own expenses.

The results of the statistical analysis will tell us about the participants' 'behaviour' in assigning levels to the performances, as well as their answers to the questionnaires will tell us about their actual difficulties; nevertheless at least two difficulties can be already highlighted because raised during the seminar: the first in relation to the management of the criteria in assigning a level to a performance. Even if the concept of profiling skills was quite clear some participants were wondering if different criteria (particularly accuracy) could have a different 'weight' according to

the level and to what extent that should be influential in assigning the level. The second in relation to the influence of the culture and experiential world in the performances of some adult learners compared to the performances of young learners particularly in relation to the capacity to handle complex topics.

The CVCL is very grateful to Sylvie Lepage and Sibylle Bolton for their precious suggestions, to Waldemar Martyniuk, Johanna Panthier and Philia Thalgott for their collaboration and support in organising the seminar.