

# Developing Proficiency Scales: the Common European Framework of Reference for Languages in the Escoles Oficials d'Idiomes in Catalonia\*

*Neus Figueras*

*This paper is an initial account of the procedures followed to develop empirically validated, context-specific proficiency scales related to the CEF levels. It describes the six-level scales developed for Listening, Reading, Speaking, Writing, Grammar and Vocabulary. These scales will be the point of reference for describing curricula objectives and reporting learner achievement and exam levels in the EOI in Catalonia.*

Key words: CEF, EOI Catalunya, DIALANG, language certification, assessment.

## I. BACKGROUND

The situation of the Escuelas Oficiales de Idiomas across Spain is not homogeneous academically or administratively, as local governments have taken different approaches to their organisation and funding. This brief foreword provides some background context on the Escoles Oficials d'Idiomes (EOIs) in Catalonia, a rather unique context in Spain, to help understand the scope and aims of the work they have embarked on to relate their levels to the levels of the Common European Framework of Reference for Languages (*CEF*) and to stress the importance of the *CEF*-related proficiency scales developed, which are now published for consultation and use.

The publication of the *Decreto 1523/1989* by the Spanish Ministry of Education fixed content level specifications for curricula and examinations in all the Escuelas Oficiales de Idiomas in Spain. This represented a major step in the standardisation of curricula and examinations of the Escoles Oficials d'Idiomes governed by the Department of Education of the Generalitat de Catalunya. Different teams of teachers and experts in different languages were commissioned to explore the implications of the contents of the official documentation and to develop curricula and examinations

that were relevant in the Catalan context. The work of these teams, which mainly concentrated on testing and curricular issues, materialised over time in:

- the gradual development of centralised, standardised certificate examinations at two levels (*Certificat de Cicle Elemental* and *Certificat d'Aptitud*) for English, German, Italian, Russian, Spanish, Catalan and French, which for most of these languages have now been in operation for over ten years,
- the publication in 1998 of curricula for the 13 languages taught in Catalonia in the *Decreto 312/1997*, which incorporated the philosophy and the principles of Draft 2 of the *CEF*, circulated in 1996,
- the creation of a Unit within the Department of Education to deal with academic and administrative agendas related to the implementation of the above.

Achieving consensus in the development of centralised tests and unified curricula was not easy, and required a combination of top-down and bottom-up approaches. Whereas it was necessary to incorporate a diversity of existing expertise and established traditions across individuals, schools and languages, it was also felt that decisions had to be informed by the need to have in place an operational system that was perceived by all parties (teachers, students, employers, educational administrators, ...) as a useful, adequate, valid and reliable response to the growing demands of the Catalan society in relation to foreign language learning and certification.

In 2001, the publication in book form of the *CEF* and the celebration of the European Year of Languages, when the centralised, standardised exams were almost ten years old, brought to the foreground the need to revisit both the exams and the curricula, not only in terms of content but also in terms of levels, as it was clear then that the six *CEF* levels would become the new learning and testing currency in Europe. This process entailed working with teachers, who also acted as item writers and examiners, and with external experts during in-house seminars held annually.

Two main fields of study were prioritised. One was the relation between the *Escoles Oficials d'Idiomes* in Catalunya levels and the *CEF*-levels. In order to be able to establish a first link between the *EOI* certificates and the *CEF* levels, the contents of the two *EOI* certificate exams in the different languages and the *CEF* level descriptors were studied and analysed in the item writing seminars held periodically. *CEF* familiarisation exercises following the procedures developed by the EU-funded *DIALANG* project ([www.dialang.org](http://www.dialang.org)) were used to help teachers develop a common understanding of the levels. Once this was done, the certificate exams were mapped against the *CEF* levels skill by skill in terms of content. The results of these exercises pointed at the *CEF* level B1 as the band within which the *Cicle Elemental* certificate was placed and at B2 as the band for the *Aptitud* certificate. For languages such as Italian and French, close to Catalan and Spanish in vocabulary and structure, it was observed that the *Aptitud* certificate could be within the C1 band in Reading. These results were reported internally and included in the examination documentation made public.

The other field of study was the comparability of the certificates. There were a number of empirical questions that needed attention: was examination difficulty equivalent across time? Was the lower certificate consistently easier than the higher

certificate? How comparable were the different certificates in the different languages? Data had been gathered through the years on the pass rates of the exams for all languages and levels, but comparing the difficulty across versions and sessions was difficult because the varied nature and distribution of the candidature (*“official” students* – those attending courses – *“lliure” students* – external candidates only sitting exams – and *That’s English students* – those following distance learning courses) made valid comparisons hard to carry out. Careful item-by-item and task-by-task analyses also showed low reliability in some of the skills tested in the exams and pointed to the need to further scrutinise and improve item writing and piloting procedures. This was considered to be a crucial issue to be tackled in parallel with the linking process, because, as following Alderson (2002):

If an examination result is not reliable, then the examination cannot be compared to the *CEF*.

If each time a new form of an examination is produced it varies according to content and difficulty, it is very difficult to compare the examination to the *CEF* since the examination does not present a stable standard.

If examination results vary by teacher and professor, by school or university, they cannot be matched to the *CEF*.

In the summer of 2003, the Department of Education invited two experts in the fields of assessment and psychometrics to visit Barcelona and help analyse the situation. These experts were Dr Felianka Kaftantdjieva from the University of Sofia and Professor Sauli Takala from the University of Jyväskylä, who reviewed in detail exam specifications, the exam development system, exam results and item analyses, and developed, together with the Unit in the Department of Education, a workplan which would:

- allow us to relate the Catalan EOI exam certificates to the *CEF* levels, the results of which would confirm or challenge the approximate link established through empirical work with teachers, and
- provide the basis for an item-banked system that would guarantee the comparability of the certificates.

This paper contains a first account of the work carried out so far. It reports on the procedures followed to develop empirically validated, context-specific proficiency scales related to the *CEF* levels, and contains the six-level scales developed for Listening, Reading, Speaking, Writing, Grammar and Vocabulary (see Appendix). These scales will be the point of reference from now on not only for describing curricula objectives but also for reporting learner achievement and exam levels in the EOI in Catalonia.

It is important to point out, however, that these scales do not correspond to years of study in an Escola Oficial d’Idiomes. The Catalan EOI scales presented in this document are broad, “common and conventional” in *CEF* terms (2001: 22), and need to be broken into “narrower, local, pedagogic” levels to reflect learners’ progress in terms of years of study in an EOI. Work along these lines has already started and it is possible to say that there is some evidence in the data gathered so far that a first

year in an Escola Oficial d'Idiomes in Catalonia (130 hours of tuition) takes learners beyond the *CEF* level A1 whereas learners having passed the *Certificat d'Aptitud* exam after five years of tuition (at 130 hours per year in Catalonia)<sup>1</sup> could be placed in between *CEF* levels B2 and C1 depending on their level of achievement or on the grades obtained in the exam. But matching *CEF* levels and EOI years is a complicated issue, which presupposes not only further empirical research but also entails a number of administrative and political issues which need to be decided upon and which are unclear at the moment of sending this text to press.

The work carried out so far and the work currently in progress is particularly relevant in the wider context of all the Escuelas Oficiales de Idiomas across Spain, as the Spanish Ministry of Education is currently developing the implementation of a new Law of Education which regulates the levels of the Escuelas Oficiales de Idiomas in terms of the *CEF* levels. It is to be hoped that the insights gained in carrying out the different procedures to develop the Catalan EOI scales can provide an empirical grounding for the meaning of *CEF* levels upon which to base political decisions and publish content specifications for curricula and exams which can be implemented in real life.

## 2. WHY DEVELOP NEW SCALES?

“A scale, like a test, has validity in relation to contexts in which it has been shown to work” *CEF* 2001: 22.

The Common Reference Levels in the *CEF* have been rapidly and widely adopted and adapted across Europe in teaching and testing contexts. In some cases, local curricula have been rewritten to mirror the *CEF*, and in some others the *CEF* scales used as objectives have replaced them. While in some cases this has been done following reflection and strict procedures of analysis (Alderson 2002, Morrow 2004), in other cases this has been done taking over the *CEF* levels and scales without analysing their relevance to the local context and – against the advice given in the *CEF* itself – without carrying out any validation.

The Catalan EOIs have taken the advice in the *CEF* seriously and, following the methodology described in Appendix A of the *CEF*, have carried out a research project similar to the one carried out in Switzerland and described in Appendix B of the *CEF*. In this respect, the Catalan EOIs have followed other institutions in Europe, which have already developed their own scales and benefited from their work, such as the YKI scales in Finland, and others described in Alderson (2002) or Morrow (2004).

The process of developing new scales has provided benefits that go far beyond the production of the scales themselves. The project has helped disseminate the *CEF* principles and levels amongst EOI teachers (over 50% of the more than 400 teachers in the EOIs were involved in one way or another in the process), who agreed enthusiastically to come to the Catalan Department of Education to participate in a project which made them

reflect on the teaching and learning process and paved the way for future changes in the curricula and in the certificate exams content and format. And the project will also affect other levels of education, as its methodology is seen as a way forward in the development of *CEF*-related descriptors and scales for public Primary and Secondary schools.

### 3. DRAFTING AND SELECTING LEVEL DESCRIPTORS PER SKILL

To achieve the aim of the project, that is, to create overall proficiency scales for Reading, Listening, Writing, Speaking, Grammar and Vocabulary, the project team assembled a descriptor<sup>2</sup> pool based on the recommendations in the available literature (North & Schneider 1998: 224) and using the following sources:

- *CEF* and DIALANG scales
- Curriculum objectives for EOI at two levels: “Certificat de Cicle Elemental” and “Certificat d’Aptitud”
- Finnish scales (YKI)
- Tables 4.3, 4.4 and 4.5. from the Manual for Relating Language Examinations to the *CEF*.

The choice of descriptors from the different sources was based on two main criteria: the descriptors should generalise well across a variety of situations of use and they should provide precise measurement. The principles outlined in Appendix A of the *CEF* (positiveness, definiteness, clarity, brevity and independence) were taken into account in the formulation of the ‘can-do’ statements and applied whenever possible. The ultimate aim was to construct a useful, practical descriptive framework and, to this end, a great deal of drafting and subsequent sifting of textual elements was carried out to formulate descriptors. Thus, when similar sentences described the same thing in different statements, they were assimilated into one wording to eliminate repetition of ideas; if a descriptor was negatively formulated, it was re-worded positively; and when adjacent sentences were part of the same point, they were edited into a compound sentence. Contentwise, some statements describing socio-cultural or strategic competences were left in the pool following the principle that they were part of EOI curriculum objectives, despite the problems noticed by North & Schneider (1998: 230) that such descriptors “attempt to define aspects of communicative proficiency which are not purely linguistic” and are “a little nebulous and difficult to observe”. As will be shown in section 4, statistical analysis proved them unstable and misfitting, and they were therefore not included in the final scales.

Descriptors were discarded when:

- they were poorly defined – too vague or unclear – (an example for Listening was *Can understand films without too much effort*),
- they described strategic or cognitive competences rather than linguistic ones (an example of study skills for Reading was *Can use dictionaries and other reference materials effectively*),

- what they described was applicable at different levels of competence (an example for Speaking was *Can relate own contribution skilfully to those of other speakers*).

The descriptors that remained in the pool were selected with the expectation that empirical evidence would confirm that:

- they would be interpreted consistently with regard to their meaning and to the approximate level allocated to them,
- they were relevant to the potential end users of the scales (teachers, assessors and learners).

The resulting descriptor pool was read carefully by a team of teachers to locate possible gaps in describing the proficiency continuum and as a first intuitive expert assessment of the above.

The descriptors in the pool were originally written in English, and then translated into Catalan, the language shared by all teachers in the EOI system. This presented some difficulties, as there was no available Catalan translation of the YKI descriptors or of the *Manual Relating Exams to the CEF*. Terminological problems were tackled bearing in mind the experience in translating the *CEF* into Catalan, and paying close attention to homogeneity and consistency of terminology.

Table 1 shows the total number of surveyed descriptors and the number of descriptors selected for the descriptor pool per skill. The high number of speaking descriptors surveyed illustrates the many speaking-related *CEF* scales which had to be processed. However, many of these Speaking scales contained descriptors which were either too detailed or described strategic and socio-cultural competences and they were not perceived to be relevant in the overall reporting scales we were attempting to develop, hence the relatively small number of descriptors selected.

Table 1: Descriptors surveyed and descriptors selected for descriptor pool

Skill	N. surveyed descriptors	N. selected descriptors
Reading	54	40
Listening	54	46
Writing	74	53
Speaking	338	80
Grammar	47	34
Vocabulary	66	61

#### 4. MAPPING DESCRIPTORS ONTO LEVELS

The descriptors in the initial descriptor pool were then submitted to the scrutiny of judges in order to be able to assess their usefulness as proficiency descriptors at the different *CEF* levels. It was expected that the surviving descriptors for each level would in

fact constitute the text for each band scale. This was done by means of a sorting exercise carried out at working sessions during which judges had to assign *CEF* levels to all the descriptors in the pool, which had been organised randomly on a worksheet. Judges were teachers with at least five years' experience in teaching EOI levels. They came from different schools all over Catalonia voluntarily because of their interest in getting to know more about the *CEF*. There were three different sessions on three different days: session 1: Reading and Listening; session 2: Speaking and Writing; session 3: Vocabulary and Grammar. The sessions lasted five hours each and they started with *CEF* familiarisation activities for each skill, such as those described in Chapter 3 of the *Manual for Relating Examinations to the CEF*; they were followed by individual work that consisted in sorting the descriptors into *CEF* levels. As a follow-up, yet another sorting exercise was carried out where years of study at an Escola Oficial d'Idiomes were assigned to each descriptor in order to facilitate future cross comparisons.

Table 2 is an overview of the work done in descriptor sorting, showing the figures corresponding to the number of descriptors sorted per skill, the teachers involved as judges, and the different languages they taught.

Table 2: Overview of sorting exercises

SKILL	DESCRIPTORS	TEACHERS	LANGUAGES*
Reading	40	103	10
Listening	46	99	10
Speaking	80	92	12
Writing	53	89	12
Grammar	34	81	12
Vocabulary	61	73	12

\* Arabic, Basque, Catalan, Dutch, English (>40), French (>20), German (>15), Italian, Portuguese, Spanish

The analysis of the data collected during the sorting exercises included calculating the mean, the median, the mode and the standard deviation of the ratings, plus the range of levels allocated to descriptors, the discrepancies across judges and the factor score. Table 3 is a section of one of the tables developed (here for Writing) and shows the descriptors upon which judges agreed most during the sorting exercises and which were to be included in the final scales. The descriptors in the table are organised in decreasing degree of difficulty (shown in the Factor Score column). This table also shows the tags used for each descriptor for ease of reference plus the most relevant statistical information. From left to right, DESCR was the code used during the rating sessions and includes the skill initial plus a randomised number (W33 = descriptor number 33 for Writing). ID corresponds to the internal code each descriptor had in the pool (EE-A1-5 = *Expressió escrita* (Writing in Catalan)-estimated *CEF* level-the number of the descriptor in the pool). The DESCRIPTOR column includes the text of each descriptor. The Md column lists the median of ratings (A1 corresponding to 1, A2 to 2 and so on). The Range column illustrates the discrepancies across judges by list-

ing the number of levels they used when rating the descriptors, and the Factor Score column shows the level of difficulty perceived by judges. Although most descriptors “survived” in the pool in the light of statistical evidence some did not. Table 3 illustrates why descriptors W07 and W10 were dropped. Descriptor W07 (*Can express his ideas in writing but with errors and not always adequately*) did not result in agreement (four different levels were assigned by 90% of the judges) probably because its vague formulation was likely to describe different levels of competence, and because it is a

Table 3: Writing Descriptors organised according to sorting exercises’ results

DESCR	Id	DESCRIPTOR	Md	Range	Factor Score	“Survived” in the pool
W33	EE-A1-5	Can complete simple forms with personal information (name, nationality, address)	1	1	-1,964	Yes
W31	EE-A1-2	Can write very short notes using very basic language	1	1	-1,839	Yes
W42	EE-A1-4	Can write some simple sentences about him/herself and his/her most immediate surroundings	1	1	-1,794	Yes
W39	EE-A1-3	Can write a short simple greeting text, such as a holiday postcard	1	2	-1,743	Yes
W16	EE-A1-1	Can write very short texts that may not be totally comprehensible because they are little or poorly structured and content is poor	1	2	-1,642	Yes
W22	EE-A1-6	Can link expressions and simple sentences using very basic cohesive devices	2	1	-1,517	Yes
W24	EE-A2-3	Can write short simple messages related to his/her everyday needs or to familiar subjects	2	2	-1,449	Yes
W35	EE-A2-6	Can link simple sentences with simple connectors	2	1	-1,413	Yes
W17	EE-A2-2	Can write short texts, but has difficulty in structuring and adequacy of language, which may render texts not totally coherent or comprehensible	2	2	-1,344	Yes
W30	EE-A2-4	Can write texts using a repertoire of simple sentences and structures to communicate in everyday situations	2	2	-1,141	Yes
W07	EE-A2-1	Can express his ideas in writing but with errors and not always adequately	2	3	-,946	No
W09	EE-B1-1	Can write intelligible texts but with some inaccurate expressions that may make reading difficult	3	2	-,914	Yes
W15	EE-B1-5	Can write short personal letters that deal with a variety of subjects	3	2	-,788	Yes
W01	EE-A2-5	Can write texts where he/she combines familiar words to make more complex ones	3	2	-,742	Yes
W34	EE-B1-3	Can convey his/her message in writing using some strategies to overcome his/her linguistic limitations	3	2	-,571	Yes
W06	EE-B2-11	Can narrate a sequence of events coherently	3	2	-,556	Yes
W10	EE-A2-7	Can recognise the basic differences between the spoken and written registers	3	3	-,550	No



complex statement whose interpretation varied depending on which part of the statement was given more attention by different judges. Likewise, descriptor W10 (*Can recognise the basic differences between oral and written register*) failed to result in agreement, possibly due to the fact that this descriptor illustrated the problem of defining aspects of communicative proficiency that are not purely linguistic and can be present at different levels of language proficiency.

The robustness of the results of the ratings by each individual judge and of the degree of agreement amongst all judges was checked using different procedures, and proved to be satisfactory, as shown in Table 4, where N stands for the number of judges that took part in the sorting exercises.

Table 4: Judges' agreement in the sorting exercises

	READING N=99	LISTENING N=90	SPEAKING N=85	WRITING N=86	VOCABULARY N=65	GRAMMAR N=72
W	.819	.781	.813	.832	.840	.854
ICC	.803	.744	.769	.779	.824	.834
Rmean	.821	.773	.793	.803	.843	.856
Alpha	.998	.997	.997	.997	.997	.998

The procedures used are described briefly below:

**W:** Kendall's coefficient of concordance, most appropriate for ordinal data. It is a measure of concordance for more than two raters. When there are only two it is similar to Spearman rank correlation.

**ICC:** Intraclass correlation: shows agreement between raters. The value shows how the average rater agreed with all the others.

**Rmean:** Average of the correlation between different pairs of raters.

**Alpha:** reliability of rating in terms of rater consistency.

To analyse the effect of the language background on rater consistency the teachers acting as judges were classified into four groups: groups 1, 2 and 3 were the groups with most teachers (English, French and German respectively), whereas group 4 was the group including Arabic, Basque, Catalan, Dutch, Greek, Italian and Spanish. Very few differences across language groups were observed, and correlations between the ratings of different language group judges was above .971, the lowest being for Reading and the highest for Grammar.

The high level of correlation between the ratings of judges representing different languages led us to conclude that the resulting scales were language independent; in other words, they were perceived in the same way by teachers of different language backgrounds. However, a detailed analysis of the differences between judges highlighted some interesting aspects which can be said to be culturally dependent and which common knowledge has identified as closely related to one language or culture in particular. One anecdotal example that confirms this was descriptor R34 for Reading (*Can understand clearly written, straightforward instructions for a piece of equipment*), which German teachers considered to belong to a higher level than teach-

ers of English and French. This was probably so because of the importance given in the German culture and education to fully understanding and applying instructions accurately. Similarly, French teachers considered descriptor L5 for Listening (*Can use knowledge of socio-cultural aspects of countries where target language is spoken to increase comprehension*) to belong to a higher level than English or German teachers. In this respect it must be borne in mind that French is considered a language closer to the L1 (either Spanish or Catalan) of foreign language students in Catalonia and that use of strategic knowledge is not perceived therefore to be as relevant or necessary at the lower levels as is the case in foreign languages less similar to learners' L1.

## 5. FROM DESCRIPTORS TO SCALE DEVELOPMENT

The scales were drafted using descriptors with close scale values and the smallest discrepancy of ratings, as shown in Table 3. At the same time, the descriptor units for two successive band descriptors in the scale were chosen in such way as to establish a clear difference between their scale values.

In drafting the scales the recommendations from Appendix A in the *CEF* were applied, thus the wording in each band responds to the principles of positiveness, definiteness, clarity, brevity and independence. But the project team also paid close attention to the scale as such, and checked it for:

- regular progression from band to band.
- overall coherence and consistency,
- conditions and limitations.

Carrying out the recommendations was not easy or straightforward, although members of the project team had wide expertise in curriculum design and development of marking criteria and could also draw on insights gained in testing-related projects such as DIALANG and the *Dutch CEF Construct Group* (Alderson et al. 2004). The issues raised during the drafting of the scales pointed to some problems in the *CEF* scales themselves, which did not systematically address aspects such as limitations (it seemed to the team, for example, that having a C2 descriptor for overall Listening with no limitations whatsoever was not a true description of a C2 language proficiency) or were not perceived to evenly describe proficiency progression.

Systematic group discussion within the project team unveiled issues which will surely need to be addressed in future editions of the *CEF*. The work reported in this paper, together with the work carried out by other institutions going through the same process and developing *CEF* related proficiency scales to link their exams to the *CEF* levels and/or writing Case Studies in accordance with the guidelines of the *Manual for Relating Examinations to the CEF* (2003), will be very useful feedback in this respect.

## 6. SCALE VALIDATION: DATA GATHERING, ANALYSIS AND RESULTS

Once the scales were developed, it was necessary to:

- validate the consecutive order of band descriptors (the vertical dimension) and the degree of separation between them, and
- establish the link between context specific scales and *CEF* scales (the horizontal dimension).

In order to do so, a pair comparison methodology was followed. Each descriptor for each level in each scale (those from *CEF* scales and those from draft EOI scales) and for each skill was paired to the descriptor of the same level in the other scale and to the rest of the descriptors of both scales. For level A1, for example, the expert judges were presented with the following pairs:

*CEFA1 – EOI A1* , *CEFA1 – EOI A2* , *CEFA1 – EOI B1*, *CEFA1 – EOI B2*, *CEFA1 – EOI C1*, *CEFA1 – EOI C2*, *CEFA1 – CEF A2*, *CEFA1 – CEF B1*, *CEFA1 – CEF B2*, *CEFA1 – CEF C1*, and *CEFA1 – CEF C2*.

The expert judges therefore had to follow the instruction below 66 times per skill, and decide in all cases which descriptor seemed higher:

“Read carefully each pair of descriptors and choose the one that describes a higher level of language proficiency.”

As for the descriptor sorting explained in section 4, all the judges doing the exercises were familiar with the *CEF* levels and had more than five years’ experience in teaching EOI levels. The number of judges participating ranged from 23 for Vocabulary to 30 for Reading, and the languages they taught encompassed up to 7 different languages.

### VALIDATION RESULTS

Tables 5 and 6 present the results of the pair comparison exercises for each skill (L=listening; R=reading; W=writing; S=speaking; G=grammar; V=vocabulary) in relation to the vertical dimension, i.e. the progression of each scale (*CEF* and the draft EOI scales) from A1 to C2. The levels shaded in grey illustrate the levels perceived by judges to follow a different order to the one expected. The levels framed are those for which judges’ ratings did not provide sufficiently distinctive values in terms of scaling. In the Grammar scales, for example, the ratings did not seem to distinguish clearly between C1 and C2 levels, and C2 was perceived as only slightly higher than C1. It is clear from the tables that judges identified problems in top levels of the draft EOI scales, and that they also perceived some problems in the *CEF* scales. In Read-

ing, for example, the draft EOI bands for C1 and C2 were defined in such a way that led judges to perceive them as quite the opposite of what they were expected to be, whereas in Listening, it was the *CEF* scales which presented problems.

Table 5: Vertical consecutive order for *CEF* band scales resulting from pair comparison exercises

L	R	W	S	G	V
CEF	CEF	CEF	CEF	CEF	CEF
A2	A1	A1	A1	A1	A1
A1	A2	A2	A2	A2	A2
B1	B1	B1	B1	B1	B1
B2	B2	B2	C1	B2	B2
C1	C1	C2	B2	C1	C1
C2	C2	C1	C2	C2	C2

Table 6: Vertical consecutive order for new EOI band scales resulting from pair comparison exercises

L	R	W	S	G	V
EOI	EOI	EOI	EOI	EOI	EOI
A1	A1	A1	A1	A1	A1
A2	A2	A2	A2	A2	A2
B1	B1	B1	B1	B1	B1
B2	B2	B2	B2	B2	B2
C1	C2	C1	C2	C1	C1
C2	C1	C2	C1	C2	C2

A number of issues are worth mentioning in relation to the results highlighted in Tables 5 and 6. The first one is that more detail in band descriptors seems to lead judges to perceive them as higher, a tendency also identified in the ALTE ‘*Can do*’ project as described by Jones (Alderson 2002), and also by Kaftandjieva & Takala (also in Alderson 2002), whereas more general band descriptors tend to be scored lower. Another issue which needs to be mentioned and which might account for the difficulties observed in the higher levels is that judges participating in the pair comparison were in fact not used to teaching at those levels, as the higher EOI levels taught at the moment are placed between the *CEF* B2 and C1 levels.

The results of the analysis showed that some of the bands needed to be revised in order to make them more distinct. This was done on the basis of the pair comparison results and also by revisiting the results of the descriptor-sorting exercises described in Section 4. The scale bands for Reading and Listening are available at <[www.xtec.es/eoi](http://www.xtec.es/eoi)><sup>3</sup> and show the differences between the *CEF* and the EOI scales and also to illustrate some of the revisions which had to be made to the EOI scales before carrying out yet another pair comparison exercise to confirm the adequacy of the changes for the final version.

## LINKING EOI SCALES WITH *CEF* SCALES

The analysis of the pair comparison exercises also allowed us to see how comparable the levels were in horizontal terms as described by the EOI scales and the *CEF* scales. As can be seen from the graph below (Fig. 1) there is not an exact match between *CEF* levels and EOI levels – in fact all EOI levels are higher than the corresponding *CEF* levels. The left-hand column of Figure 1 shows the rated *CEF* scale bands (EC2 to EA1), while the right-hand column shows the rated EOI scale bands for six levels (NC2 to NA1).

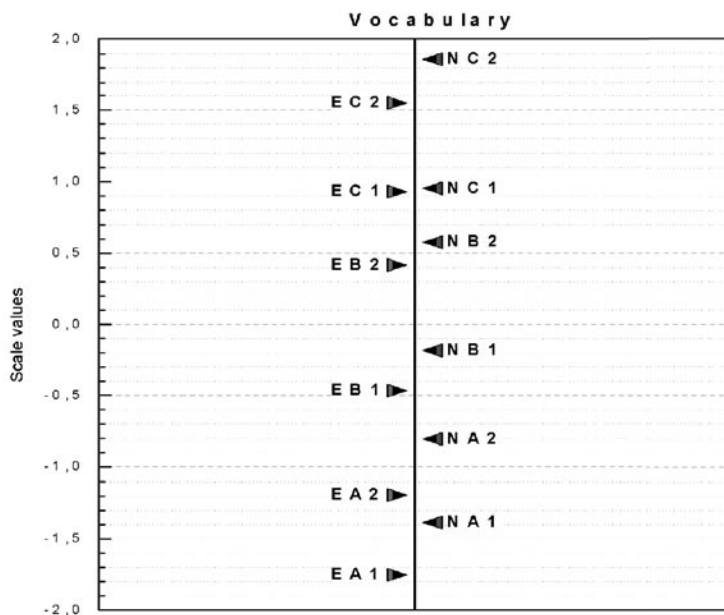


Fig. 1: Pair-Comparison Results – Vocabulary

As for the other scales developed (Listening, Speaking, Writing, Grammar and Reading), the picture is very similar in that the corresponding *CEF* and EOI levels are not an exact match. These results, however, need not necessarily be interpreted in a negative way. The aim of the whole process of writing the now validated EOI scales was to develop content specific scales whose descriptor bands were perceived in a similar way by users with different language backgrounds in the EOI context and which could be linked to the *CEF* scales. This aim can now be considered fulfilled. Linkage, however, does not mean equivalence – it means that the values of one scale can be converted into those of the other scale on the basis of empirical evidence.

On the other hand, the pair comparison analysis revealed that the *CEF* scales are not ideal either and certainly need improvement. The *CEF* is a good, comprehensive document, but understanding it thoroughly also means that it is more important to

know how close any new scales are to it than trying to identify possible gaps in it and attempting to change the scales. Of course, this will require further work not only in the direction of validating the revised EOI scales, but also in deeper analysis to ascertain the reasons why judges perceived some bands as higher than their counterparts, but the work done so far also needs to be assessed for what it is worth. We know now where the new EOI scales stand in relation to the *CEF* levels for the different scales, which is far more than we did know when we started the project of developing new scales and the corresponding validation procedures.

## 7. CONCLUSION

After this brief and somewhat panoramic description of the work carried out in the development of *CEF*-related scales for the Catalan *Escoles Oficials d'Idiomes*, we are in the position to say that the scales which are presented in this document are a useful and valid instrument to describe language proficiency for all the languages taught in the EOI in relation to the *CEF* levels. The scales will be used in the revision of present curricula to link their objectives to the *CEF* and also in reporting learner achievement and examination levels in relation to the *CEF*. In the interpretation of the scales, however, it should be borne in mind that the EOI scales are linked to the *CEF*, but not equivalent to the *CEF*. In other words, in order to convert the EOI scale bands or levels into the *CEF* levels the empirical results of the work carried out so far will need to be taken into account and reported when relevant.

The project of developing EOI scales, although far more arduous than the team expected, has been extremely rewarding, thanks to the huge amount of help we have received from our enthusiastic teachers all over Catalonia. It is worth highlighting at this point that the experience gained during the development and drafting of the scales and the various seminar sessions and exercises by the project team and by the participating teachers is invaluable and illustrates the force of the *CEF* in bringing together professionals and helping them reflect on how languages are spoken, taught and learnt. The experience gained in the development of these six-level scales will surely be very useful when drawing up course objectives, not only from the content but also from the descriptive and the methodological points of view.

The expertise gained in these three crucial aspects (content, descriptive and methodological) in the development of scales will also be extremely useful for teachers in Primary and Secondary education, who are now in the process of adapting course objectives to the *CEF* levels and in the process of ascertaining the usability of the European Language Portfolio. These teachers face the difficult task of having to adapt and adopt *CEF* descriptors initially developed for the adult learner to suit learners in Primary and Secondary schools.

The fact that most of the descriptors in the initial pool were in English and had to be translated into Catalan has also been a revealing learning experience in relating

words and meaning. The *traduttore-tradittore* dictum has been present throughout the project, and it is important to highlight that wording is crucial and can never be given sufficient attention.

To conclude, we would like to stress the fact that this paper is only a first account of the work still in progress. The process of implementation of the scales presented here will take time, but it will surely result in better quality tests and programmes and hopefully better learning and teaching in the EOI context. We have gained insights, raised questions and outlined issues which are still to be resolved, but the experience has made us better aware of what it means to use and apply the *CEF*.

## BIBLIOGRAPHY

*All web sites mentioned in the article were active 07/05.*

- Alderson J.C. (ed.). 2002. Common European Framework of Reference for Languages: Learning, teaching, assessment. Case Studies. Strasbourg: Council of Europe Publishing.
- Alderson J.C., Figueras N., Kuijper H., Nold G., Takala S. & Tardieu C. 2004. Development of Specifications for Item Development and Classification within The Common European Framework of Reference for Languages: Learning, teaching, assessment. Reading and Listening. Final Report of The Dutch Construct Project (available on request from the Project Coordinator, J. Charles Alderson, c.alderston@lancaster.ac.uk).
- Council of Europe 2001. Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: CUP. (available online at <[http://www.coe.int/T/E/Cultural\\_Co-operation/education/Languages/Language\\_Policy](http://www.coe.int/T/E/Cultural_Co-operation/education/Languages/Language_Policy)>).
- Council of Europe 2002. Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza, evaluación. Instituto Cervantes; Ministerio de Educación, Cultura y Deporte. Editorial Anaya.
- Council of Europe 2003. Marc europeu comú de referència per a les llengües: aprendre, ensenyar, avaluar. Generalitat de Catalunya; Govern d'Andorra; Govern de les Illes Balears.
- Council of Europe 2003. Manual for Relating Language Examinations to the CEF. A Preliminary Pilot Version. Strasbourg: Language Policy Division. DGIV/EDU/LANG. at <[http://www.coe.int/T/E/Cultural\\_Co-operation/education/Languages/Language\\_Policy](http://www.coe.int/T/E/Cultural_Co-operation/education/Languages/Language_Policy)>.
- Jones N. 2002. Relating the ALTE Framework to the Common European Framework of Reference. In Alderson J.C. (ed.) Common European Framework of Reference for Languages: Learning, teaching, assessment. Case Studies, 167-183. Strasbourg: Council of Europe Publishing.
- Kaftandjieva F. & Takala S. 2002. Council of Europe Scales of Language Proficiency: A Validation Study. In Alderson, J.C. (eds). Common European Framework of

- Reference for Languages: Learning, teaching, assessment. Case Studies, 106-129. Strasbourg: Council of Europe Publishing.
- Morrow K. (ed). 2004. Insights from the Common European Framework. Oxford: Oxford University Press.
- North B. 2004. The CEF levels & Descriptor Scales. at <[www.ealta.eu.org/conference/2004/ppt/north14may.ppt](http://www.ealta.eu.org/conference/2004/ppt/north14may.ppt)>.
- North B. & Schneider G. 1998: Scaling descriptors for language proficiency scales. *Language Testing* 15, 2: 217–262.

#### Acknowledgements

*Project Team:* Arlette Aubier, Birgit Ferran, Neus Figueras (*Project Coordinator*), Plàcida Mingarro, M.Cristina Riera, Sonia Sáez, Dolors Solé (*Directora del Centre de Recursos de Llengües estrangeres*).

*External experts:* Felianka Kaftandjieva from the University of Sofia (Bulgaria) and Sauli Takala from the University of Jyväskylä (Finland).

Neus Figueras holds a PhD in language testing from the University of Barcelona, Spain and works in the local Ministry of Education in Catalonia, where she is in charge of foreign language certificate exams to adult learners of seven different languages. She has been involved in a number of international research and development projects (Speakeasy, Dialang, Ceftrain) and collaborates regularly with the Council of Europe in the dissemination of the Common European Framework of Reference in relation to testing and assessment. She is the president of EALTA (European Association for Language Testing & Assessment).

- \* The contents of this article are published on the web at [http://www.xtec.net/crle/eoi/fulleto\\_2005.pdf](http://www.xtec.net/crle/eoi/fulleto_2005.pdf). We gratefully acknowledge permission to publish this article in print form.
- <sup>1</sup> Implementation of the programmes for languages such as Arabic, Chinese, Greek, Japanese or Russian requires a longer tuition period.
- <sup>2</sup> A descriptor being a stand-alone statement which describes an aspect of communicative proficiency and which is part of a band on a rating proficiency scale
- <sup>3</sup> Generalitat de Catalunya, Departament d'Educació-CEF/EOI scales article <[http://www.xtec.net/crle/eoi/fulleto\\_2005.pdf](http://www.xtec.net/crle/eoi/fulleto_2005.pdf)>.