



Users' Guide to Item CD

The aim of this CD of pilot sample reading and listening items is to help users relate locally relevant test items to the CEF levels, and to gain insights into developing test items that can eventually claim to be related to CEF levels. Thus the users of this CD are expected to be

- those professionals who wish to pilot the Preliminary Version of the Manual and/or who wish to produce a Case Study to document the process of validating the linking of a particular examination to the CEF levels,
- test developers who wish to produce reading and listening test items which can be related to the CEF, both in terms of the Descriptive Scheme and in terms of the levels and the illustrative scales,
- examination bodies who wish to relate existing tests and test tasks to the CEF,
- teacher trainers who wish to help language teachers understand better the CEF and its levels for reading and listening,
- language teachers who wish to become more familiar with the CEF in terms of reading and listening abilities and who would like their own classroom assessment to reflect those levels,
- policy makers who wish to understand what the CEF levels mean in terms of the measurement of reading and listening abilities, in order to decide in a more informed manner what level might be appropriate for particular language uses and users.

A CD and a short Users' Guide cannot, on their own, satisfy the needs of such stakeholder groups, nor can it alone meet the general aims set out above. It is essential, therefore, that users consult fully the CEF itself, the Preliminary Pilot Version of the Manual for Relating Language Examinations to the CEF, and associated materials, including in particular the Final Report of the Dutch CEF Construct Project and the associated CEF Grid (see below).

As emphasised time and again in the Manual, it is only through thorough familiarisation with relevant documents, participation in related workshops and seminars, and the running of training events for language education professionals that the aims of this CD and the Manual more generally can be achieved.

Nevertheless, a Users' Guide can be of assistance in helping users engage with that process and achieve those aims. Detailed guidelines on how to run familiarisation and standardisation sessions, and on how to relate judged levels to empirical data on test and item performance, can be found in the Manual, in particular sections 5.5.2 to 5.7.4. In order for item writers to select texts and tasks and produce test items targeted at a specific CEF level, for test developers to judge the CEF level of existing test items, and for teachers and policy makers to understand the CEF levels for reading and listening, training in assessing test tasks is needed. Training with sample items with known difficulty values and CEF levels should take



place and, in the case of item writers and test developers in particular, then be followed by a process of analysing locally produced items with known difficulty values. This CD provides pilot samples of items and tasks whose difficulty is known, and which can be used in this process.

Training with calibrated items includes:

1. Becoming fully aware of the range of CEF subscales of descriptors for specific areas that are available in the CEF (see Manual, Chapter 4).
2. Identifying the content of the items analysed in terms of their coverage of CEF levels, scales and constructs (see the description of the Dutch CEF Grid, below).
3. Estimating the level the item represents in terms of the relevant CEF descriptors (also using the Dutch CEF Grid).
4. Discussing the possible reasons for discrepancies between estimated and empirically established levels.
5. Confirming the level of difficulty against empirical data.

It is wise to start with the skill of reading, since it is easier to organise group work and to read and re-read texts and items in print than it is to work on listening texts and listening items. In addition, it is not generally considered good practice to investigate the nature of listening texts and items by reading transcripts. Rather, it is essential for analysts to listen to the text in the same way as test takers will listen.

Once the process of assessing items for reading has been completed, organizing the session for the skill of listening and working with listening texts will be easier, as users will already be familiar with the task to be done.

Even if users have already attended a general Familiarisation session as described in the Manual, Chapter 3, a sorting exercise with descriptors for the skill concerned before starting difficulty estimation and standard setting is a necessary training exercise. Full details are given in the Manual, Chapter 5, sections 5.5, 5.6 and 5.7, on how this might be done.

The **standardisation process** follows a number of phases (see also 5.5.2 and 5.7.2):

- **Illustration:** A first assessment of the level of one text and its corresponding tasks and items. This preliminary activity will help users tune into the CEF levels for the skill being assessed. Any of the samples on this CD can be used, selecting those that are closest to one's own context.

It is essential to consider both the **level of the source text** and the **difficulty of the individual item(s)** associated with it. A text does not have a "level." It is the competence of the test takers as demonstrated by their responses to the items that can be related to a CEF level.

— It is very important that users actually read the text and answer the item/s individually before they estimate the difficulty of the question concerned and the CEF level it best illustrates. Mere 'eyeballing' of a text and its associated items is NOT adequate.

— After responding to the item(s), users should be able to compare their own responses to the correct answer for the item(s) concerned. Discussion to ensure clear understanding of the answer key or scoring rubric should precede their estimation of the item difficulty.

— The instruction users are recommended to apply to each item individually is:

"At what CEF level can a test taker already answer the following item correctly?"

— It is recommended that if working in groups, users first note their ratings for the items on their own, and only then justify and discuss their decisions in pairs or small groups, before being informed of "the" level that the item(s) really are calibrated to.

- **Controlled Practice:** Once the above steps are complete, users can then assess different texts with their corresponding tasks and items, relating them to CEF levels and identifying the CEF descriptors operationalised by each item/task.
- **Individual Assessment** (Manual, page 86).

The process of Standardisation briefly described here is summarised in Table 5.11 in the Manual (page 98). That table provides a useful overview of the various steps involved.

Content Analysis

Content analysis of the test items and tasks, as well as completion of the Communicative Language Activities forms A9 and A10 in the Manual, may be facilitated by the use of the Grid developed by the Dutch CEF Construct Project.

The Dutch CEF Construct Project aimed to help test developers and other language educationists construct or relate test items to the CEF. Since the CEF is not directly relevant for the construction of test specifications, or the evaluation of test items, it was necessary to supplement guidance provided in the CEF itself with information from other sources on what reading and listening tests might contain. The Final Report of this Project has been included on this CD for ease of reference. Full descriptions of the constructs of reading and listening can be found in Alderson (2000) and Buck (2001).

One major outcome of the Project was an Internet-based Grid which can be used to help characterise reading and listening texts, items and tasks, and this Grid has been used in this way with the samples on this CD. Application of the Grid results in summary tables of the sort that accompany those samples. Readers of this Users' Guide can access the Grid at:

www.ling.lanccs.ac.uk/cefgrid

and use it in the content analysis of their own texts, items and tasks. It must be pointed out, however, that the grid now available at this address is slightly different to the grid initially used in the analysis of the items in this CD. The changes recommended by the analysts have now been incorporated, and hence the differences in some of the categories listed.

As part of the process of familiarisation with the Grid itself, users of this Guide might like to attempt to characterise one or more of the samples on this CD without making reference to the Summary pages, and then to compare their own Summary page (s) with those provided on this CD.

The Grid can be used by individuals, who should, of course, already have familiarised themselves with the CEF and its associated scales. However, in order to help them estimate the CEF levels of texts, items and tasks, relevant scales from the CEF are available within the Grid for reference purposes.

It is recommended that the Grid be completed by teams working together on the materials they wish to analyse, in order to arrive at a consensus analysis. It is well known that judges will differ in their characterisation of the content of most texts and tasks, and so it is important to attempt to reach agreement before inputting one's analysis to the Grid.

However, if so desired, individual team members can first input their own analyses to the Grid and these can then be compared by examining the results stored in the associated database. Reasons for discrepancies should then be discussed, and a common agreement should be reached. Such an exercise is also very useful in the process of familiarisation of teachers with the CEF, and in the training of item writers and test developers.

References

- Alderson, J. C.** (2000) *Assessing Reading*. Cambridge Language Assessment Series. Cambridge: Cambridge University Press
- Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S. and Tardieu, C.** (2004α) *Specifications for Item Development and Classification within the CEF: Reading and Listening (English, French and German): The Dutch CEF Construct Project*. Paper presented at a Symposium on Psycholinguistic and Psychometric Aspects of Language Assessment in the Common European Framework of Reference for Languages, University of Amsterdam, Feb 13th – 14th, 2004.
- Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S. and Tardieu, C.** (2004b) The 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 CEF Construct Project*. Unpublished manuscript, July 2004.
- Buck, G.** (2001) *Assessing Listening*. Cambridge Language Assessment Series. Cambridge: Cambridge University Press.
- Council of Europe** (2003) *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF). Manual: Preliminary Pilot Version*. Strasbourg: Language Policy Division, Council of Europe.