# SCALE VALIDATION STUDY

F. Kaftandjieva & S. Takala

The movement toward international integration in Europe leads to increased requirement of common standards in all areas, including language teaching and learning. This need of a common yardstick led to the development of A Common European Framework of Reference for language learning, teaching and assessment whose major function is

> '*... to allow all the different partners in the language teaching and learning process to inform others as transparently as possible of their objectives, primarily in terms of what they wish learners to achieve, the methods they use and the results actually achieved. This information will be of very great value in an interactive Europe to allow the different partners to provide learners with coherent provision and also to facilitate educational, vocational and professional as well as personal mobility across a continent in which artificial barriers to movement, communication and co-operation are being progressively removed.* '

Modern Languages: Learning, Teaching, Assessment: A Common European Framework of reference, Strasbourg, 1996, Chapter 2: Aims and Functions of the Framework [Available on line: http://culture.coe.fr/lang/eng/eedu2.4b.htm)

According to the Common European Framework, there are 6 common reference levels: **Breakthrough** (A1), **Waystage** (A2), **Threshold** (B1), **Vantage** (B2), **Effective-proficiency** (C1), and **Mastery** (C2). The first two levels (A1&A2) identify the **Basic user**, the next two (B1&B2) identify the **Independent user,** and the last two (C1&C2) identify the **Proficient user**. A number of six-point scales of language proficiency have already been developed for different language skills and purposes. This is, in fact, one of the main advantages of the Common European Framework – that it is an open and flexible system, allowing further development according to the purposes of their potential users.

The implementation of the Common European Framework for language learning, teaching and assessment in Finnish Polytechnics will provide common standards for language assessment and will facilitate the further co-operation and co-ordination between different educational institutions as well as the comparability and the mutual recognition of the language qualification of their students.

Despite the benefits, the implementation of the Common European Framework is not straightforward and unproblematic.

Some of the main problems concern the validity of the scales, proposed by the Common European Framework, and can be presented as a number of questions that have to be answered:

- Do the level descriptors cover the whole range of the process of language acquisition?
- Do the level descriptors represent the stages of language acquisition in a consecutive order?
- Do all independent units of the level description represent the same level of language development?





- Do the descriptor units constitute a scale, corresponding to the original language proficiency scale?
- Do language experts with different background interpret the level descriptors and descriptor units in a similar way?
- Can language experts differentiate between the level descriptors?
- Are the scales of language proficiency comparable across the languages?
- How can the Council of Europe (CoE) Scales of language proficiency be linked to the scales developed in the Framework of the Finnish National Foreign Language Certificate?
- How can the CoE Scales of language proficiency be improved in order to serve better the purposes of language learning, teaching, and assessment in Polytechnics?

The current workshop will use the **sorting** of descriptor units and **pair comparison** to collect date, whose further statistical analysis will provide answers to those questions. This will result in six six-point scales of language proficiency covering the four basic language skills (Speaking, Writing, Reading and Listening comprehension) as well as Vocabulary and Grammar/Structures, which closely correspond to CoE Scales, but are of better quality and more suitable for the purposes of Finnish Polytechnics. These scales will be linked to the nine-point scales used in the Finnish National Certificate, and evidence of their validity and scalability[1] will be provided.

With the increasing use of scales of language proficiency in all language testing, the question of scale quality gains ever more importance. However, a study like the current one is not easy to carry out, since it requires many resources. Therefore, for your everyday teaching practice it will be useful to have a set of criteria for the evaluation of scale of language proficiency, recommended for use by others or developed by yourself for some concrete specific assessment purposes. The following set of criteria is a modified version of those proposed by Chicago Board of Education (Available on line: http://intranet.cps.k12.il.us/Assessments/Ideas_and_Scales/Intro_Scoring/Eval_Scales/eval_scales.html):

- Does the scale relate to the outcome(s) being measured? The scale should address all aspects of the outcome(s) being measured and it should not address anything extraneous. For example, spelling and grammar might be considered extraneous on a science assessment, unless it is measuring an outcome that deals specifically with communication.
- Does it cover important dimensions of student performance?
- Do the criteria reflect current conceptions of excellence in the field?
- Does the scale reflect what you emphasize in your teaching?
- Does the highest scale level represent a truly exemplary performance or product? When you evaluate scales, you need not be concerned about having a certain number of students scored at every point on the scale. It may be that no student will attain the

[1] See Fig.1 where the calibrated descriptor units and level descriptors for the CoE scale for listening comprehension are presented. The scale values are based on a previous more limited validation study.

highest scale level. Nonetheless, it still may be worthwhile to have that level on the scale as a standard of excellence for which students should strive.

- Are the scale levels well defined?
- Is it clear to everyone what each level mean?
- Is there overlapping between level descriptors?
- Is it clear exactly, what a student needs to do to get a score at each scale point?
- Can you easily differentiate between scale levels? An easily understood scale with clear definitions of each score level is the ideal. Conversely, it is usually best to avoid scales that are labelled only at the highest and lowest points.
- Can the scale be applied consistently by different raters? Inter-rater reliability depends on how well the scales and scale level are defined and the extent to which you and your colleagues can arrive at consensus about how performance should be measured and what constitutes good performance.
- Can the scale be understood by students and the other parties involved?
- Can it be explained without technical jargon and in terms that even non-experts in language teaching and learning can understand?
- Is the scale developmentally appropriate?
- Can the scale be applied to a variety of tasks? The most useful scales can be applied to more than one task.
- Is the scale fair and free from bias?
- Does it reflect teachable skills or does it address variables over which students and educators have no control, such as the student's culture, gender, or home resources?
- Does the scale reward or penalize students based on skills unrelated to the outcome being measured?
- Have all students had an equal opportunity to learn the content and skills addressed in the scale?
- Is the scale appropriate for the conditions under which the task was completed?
- Does the scale make sense to you?
- Will it provide the kind of information you need and can use effectively?
- Does the scale have a reasonable number of levels?
- Is the scale useful, feasible, manageable, and practical?

The choice of a proper scale does not mean that it will be applied in a proper way. The answer to the last criterion can be discovered only after a try-out of the scale that will give the information not only about how feasible, manageable, and practical the scale is, but also how consistent you and your colleagues are in the application of the scale.





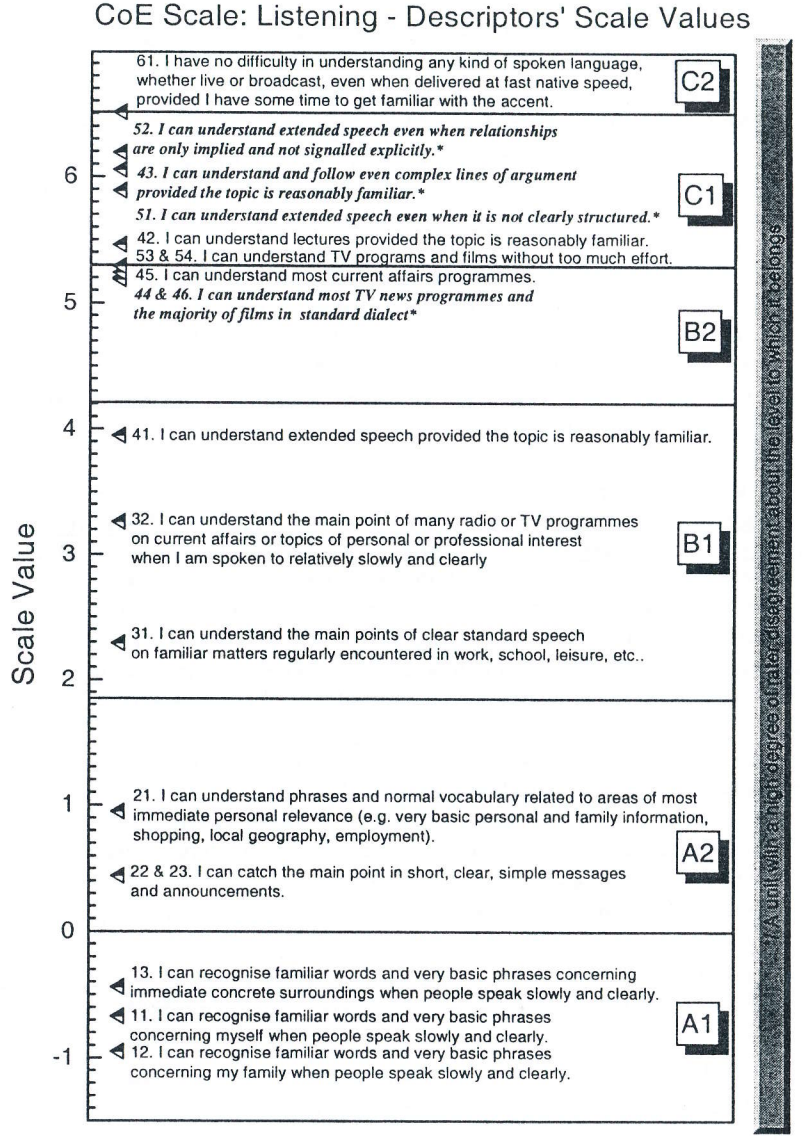


**Fig. 1**