

Raili Hildén, University of Helsinki

Sauli Takala, University of Jyväskylä

Relating Descriptors of the Finnish School Scale to the CEF Overall Scales for Communicative Activities

The aim of the study was to check the correspondence between two scales: the Finnish school scale (FINSS) created for national language syllabi and the CEF overall scales for the communicative activities (listening, speaking, reading and writing). Many FINSS descriptors derive from the CEF, but we also consulted Canadian Benchmarks to see if they would help in taking into account school-level curricular needs. For school learning purposes, a more fine-grained level division was used than in the CEF scales. The FINSS was subjected to a phase of social moderation and its internal consistency was studied across a group of users. In this study we set out to investigate the degree to which we can claim the correspondence of the FINSS with the CEF scales.

Two research problems were addressed:

1. What is the level of agreement between judges on the CEFR level of the FSS descriptors?
2. How are the Finnish syllabus descriptors related to CEFR levels?

To empirically establish the relationship between the two scales the descriptors of the FINSS were cut into atomistic propositions and

jumbled for the judges (expert language teachers, about 20 in number) to sort them back onto successive CEF levels (+levels included) for the global scale of respective skills. 20 expert language teachers with solid prior knowledge about the CEF and the FINSS were invited to cooperate. The data was collected using an online questionnaire.

The results indicate a good agreement between raters, as confirmed by several types of indicators. The compatibility between the two scales also proved to be satisfactory, since 65% of the FSS descriptors were assigned to the original CEF levels. For the rest of the descriptors, some tendency of overestimation was observed.

Context of the study

Finnish language syllabi have been influenced by European trends during more than three decades. Clear links were established with the Council of Europe work since the mid- 1970s when the basically functional-notional syllabus was introduced for the language teaching of the Finnish comprehensive school. The core curricula have since then been revised and reformed with some ten years' intervals. This study relates to the latest cycle of curriculum work initiated at the turn of the millennium.

When the current work on new curricula started in 2001, it was decided to try to adopt Common European Framework reference scales and adapt them to the national context, as part of the curriculum. This decision was based partly on ten years' of positive experience using proficiency scales in adult education and on policy decisions by ministers of education.

Its outcome, core curricula for compulsory basic education and the Upper Secondary School, became effective in 2005 at the latest.

The latest curriculum reform for the second national language and foreign languages in compulsory basic education (at the age of 7 – 16 years) and the Upper Secondary School (16 – 19 years) takes place in the following context:

- Language core curricula from 1994 were designed to be **open** to local applications and consequently written in considerably **general terms**. Therefore they may **not** have given **enough support to the teacher** in his/her work. (National Core Curriculum for the comprehensive school 1994; National Core Curriculum for the Upper Secondary Level 1994.)
- There is also an increasing need of **coherence** of curricular outcomes among different types of educational institutions at home and abroad. There should be greater **transparency** in the syllabuses to guarantee that all the stakeholders (teachers, administrators, students and their parents) understand the goals and required activities and operations in a sufficiently similar way. (Principles and guidelines 2000; Sheils, 1999.)
- An intelligibly worded connection needs to be established among **teaching, learning and assessment**, the latter covering all significant stakes from a simple school test up to the matriculation examination. Transparency of goals is a primary condition for our ability to keep track of whether teachers and test writers assess what is taught at schools and if that is close to what students are learning. (Kohonen 1997, 2000; Principles and guidelines 2000.)
- The need for transparency in an easy-to-understand language is also connected with the idea of fostering learners towards **autonomy**, as the learners only can take responsibility for

something they understand and know how to put in practice (Huttunen 1996; Little 1991).

- During a decade of local adaptation of the frame curriculum from 1994, a growing concern for **educational equality** was voiced in Finland. As a consequence, the present cycle of curricular reform is characterized by a demand for being more **normative** so as to guarantee as **equal outcomes** as possible, irrespective of such factors as region, county, school and teacher. The cornerstone of Finnish educational policy, educational equality, is receiving stronger attention than in the recent past.
- The Finnish language curricula traditionally fail to show a clear **progression** of language proficiency from one stage to another in logically coherent (not to speak of empirically verified) formulations.
- The goal settings were further specified by a supplement in 1999, presenting the criteria for a good grade in a school subject at the end of the compulsory basic school. (This supplement was based on an earlier version of the CEF and it has been consulted in the course of the work on the new versions)

To meet the challenges mentioned above, the language curriculum group in fall 2001 agreed on developing a proficiency level scale for Finnish basic schools and upper secondary schools (FINSS). The scale – as well as the entire process of writing new language curricula – would follow the principles of Common European Framework of Reference. The majority of the scales consulted were applied selectively so that the final impact of the CEF scales as a source for FINSS was determined by its relevance for language studies in the Finnish comprehensive education.

We also consulted other sources: among these may be mentioned Brian North's work on language proficiency descriptors, the Canadian benchmarks 2000 and the extensive Finnish experience with scales in adult education. The Finnish piloting work on ELP (European Language Portfolio) has shown promising prospects of applying the CEF scale at lower stages of language proficiency in a school context. The level descriptions are common to all languages, and the final specifications are left to different language sections and detailed in teacher training materials.

The principles of CEF are also in evidence in the language specific choice of descriptive categories of the external context of use (CEF 2001, 48 – 49).

As more than a half of Finnish youngsters attend upper secondary education, it seems most natural that the goals and the scale reflecting the goals are drawn up in close cooperation by experts on both the comprehensive school and the upper secondary school. Three check points of progress are defined: one at the end of grade 6, the second at the end of grade 9 of the comprehensive education and the third, finally, at the end of the upper secondary school, when there is an additional connection to the level of the matriculation examination.

For the purposes of the basic school and the upper secondary school, the CEF levels are unfortunately too broad to register small – but real - advances in language proficiency and the need for splitting up the original levels was soon recognized. On the other hand, the highest levels of mastery are rarely attained in the course of regular school studies. As the goals prescribed in the curriculum should reflect the proficiency of an

average student at each stage, the CEF descriptions above Strong vantage were omitted from the FINSS as irrelevant for the context as a whole. The conceptual links between CEF and the FINNS are shown below:

CEF level	FINNS level
A1 Breakthrough	A1.1. First stage of elementary proficiency
	A1.2. Developing elementary proficiency
	A1.3. Functional elementary proficiency
A2 Waystage	A2.1. First stage of basic proficiency
A2+ Strong Waystage	A2.2. Developing basic proficiency
B1 Threshold	B1.1. Functional basic proficiency
B1+ Strong Threshold	B1.2. Fluent basic proficiency
B2 Vantage	B2.1. First stage of independent proficiency
B2+ Strong Vantage	B2.2. Functional independent proficiency

At each level band, four performance statements are specified for following communicative activities and the relevant aspect of them:

Listening comprehension

- Themes, text and tasks (1)
- Conditions and constraints (2)

Reading comprehension

- Themes, texts and tasks (2)
- Conditions and constraints (1)

Speaking

- Themes, texts and tasks (monologue and interaction)
- Fluency
- Pronunciation
- Linguistic range
- Linguistic control

Writing

- Themes, texts and tasks (2)
- Linguistic range
- Linguistic control

The first empirical stage of the scale validation examined how consistently the expert teachers involved in the curriculum process shared the meaning of the descriptors to be scaled (Hildén & Takala 2002). In the course of this process certain conditions of approval were set for every single descriptor in order to be included in the FINSS. Whenever the range of suggested placements turned out to be too broad the descriptor was reworded and subjected to a new circle of descriptor sorting. However, establishing the empirical link to the CEF itself remained a topic for further exploration. This is the focal point of the article at hand.

Research questions

The main objective of this study was to establish the link the descriptors of the FINSS and descriptors picked out from related CEF scales, in terms of consistency between the ratings of expert teachers. Two research questions were formulated to guide the analyses:

RQ1. What is the level of agreement between judges on the CEF level of the FSS descriptors?

RQ2. How are the Finnish syllabus descriptors related to CEF levels?

Design

To conduct the study, the FSS level descriptors were split up into 184 statements. We ended up with 38 descriptors for listening, 66 for speaking, 31 for reading and 49 statements illuminating writing ability. The statements were coded and grouped in terms of communicative activities (ranging from S1 to W184). Criterion scales used in the rating of FSS descriptors were selected from among relevant CEF scales. A sample of 40 Finnish language teaching experts was contacted by an e-mail questionnaire. A randomised selection of statements referring to each of the four communicative activities was e-mailed to the raters (in Finnish translation by Huttunen & Jaakkola 2003).

The procedure might be demonstrated by following example:
The respondent received the CEF scale for Spoken fluency and a set of FINSS descriptors designed to illustrate fluency aspects of speaking arranged as a table with FINNS statements in rows and CEF level codes in columns. For each of the FINNS statements the respondent was to choose the best matching CEF level by drawing a cross in respective column.

20 respondents out of the 40 originally contacted returned the questionnaire.

Results

1. Range of rater agreement

The **range** is the difference between the highest and lowest level setting that appeared among the responses, and therefore indicates how homogeneous the rating is for a certain descriptor. The acceptable range was set at 2 (or less) which would mean that all ratings are in the interval of 3 consecutive sub-levels. In our case the maximum range detected was 6 and more than 50% of the descriptors had a range of 2 or less, which can be considered as a satisfactory result.

The range was also calculated and scrutinized for single descriptors. We combined the range information with variance and checked how close in wording the particular FINNS descriptor was to its counterpart in the CEF scale. This process revealed a cluster of descriptors that could be labeled as acceptable or unacceptable. Fortunately, in only 16 cases out of 184,

the variance exceeded 2,0, whereas 49 descriptors counted a variance lower than 0,5.

Examples of both types of descriptors are presented in Tables 1 and 2.

Table 1. A descriptor with low level of agreement between raters

RQ 1.
 Agreement with the CEFR levels
 Example of a descriptor with a low level of agreement between raters.
 "Can identify the writer's bias and the purpose of the text and locate and integrate several specific pieces of information in a longer text. Can quickly identify the content and relevance of new items deciding whether closer study is worthwhile"

R12					
	Level	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4,00	1	5,0	5,0	5,0
	6,00	7	35,0	35,0	40,0
	8,00	5	25,0	25,0	65,0
	9,00	7	35,0	35,0	100,0
	Total	20	100,0	100,0	

variance=2,37 FINSS=B21 CEF
 level = B2

Table 2. A descriptor with high level of agreement between raters

RQ 1.

Agreement with the CEFR levels Example of a descriptor with a high level of agreement between raters.

“Can write the alphabet of the language and all numbers and numerals.
Can write down basic personal identification information and write a small number of familiar words and simple phrases.”

W42

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	11	55,0	100,0	100,0
Missing	System	9	45,0		
Total		20	100,0		

variance=0 FINSS=A11 CEF level
= A1

Slightly unexpectedly, though, the closeness in wording with the original CEF formulations did not seem to play a crucial role for the consistency among expert ratings. No clear trend could be distinguished.

To conclude from these findings a number of features of a good descriptor can be stated, among others:

- The statement is concretely worded.
- The number of propositions is limited.
- The statement is brief or the meaning units of a longer formulation have the same reference frame.
- Tähän kai jotain lisää vielä? riittänee

The match between ratings was explored at the level of communicative activities as well. This was done by combining the disaggregated FINSS statements back into their respective skill areas: listening, speaking, reading and writing. No clear differences were detected, but the raters tended to be most unanimous in regard of listening, where 50% of the cases (ratings submitted by the experts) lie within a range of two CEF levels as compared with a range of three CEF levels for the rest of the skills. This is probably due to the fact that listening descriptors are more parsimonious both in number and length than the statements written to illustrate speaking and writing.

Generally, the inter-rater consistency was quite high. The average inter-rater correlation was 0.87 and the average correlation with the aggregated rating (ratings in average) was 0.92. The correlations with the 45 descriptors, where there was no missing values, point to even higher inter-rater consistency. The homogeneity of the raters is further established by the fact that one main factor explains 85% of the variation found among the same 45 descriptors.

The homogeneity of the experts is a benefit in terms of reliability and it also supports the claim for the importance of familiarization prior to the phase when the level scales are actively used in the school context. Our judges were well prepared to work with the scales, but the need for further training among field working teachers is obvious.

2. Relation of the Finnish syllabus descriptors to CEF levels

A closer look at the agreement between the FINSS level (syllabus level) and the CEF level of a descriptor revealed a good match. 118 out of 184 (64%) descriptors were assigned by judges at the original CEF levels (A1 – C2) as can be seen in table 3.

Table 3. Level assignments of the FINSS descriptors onto the CEF levels

	CEF level						Total
	A1	A2	B1	B2	C1	C2	
FINSS A1 level	28	18	1	0	0	0	47
A2	2	25	8	0	0	0	35
B1	0	2	26	6	0	0	34
B2	0	0	4	28	12	1	45
C1	0	0	0	11	11	1	23
Total	30	45	39	45	23	2	184

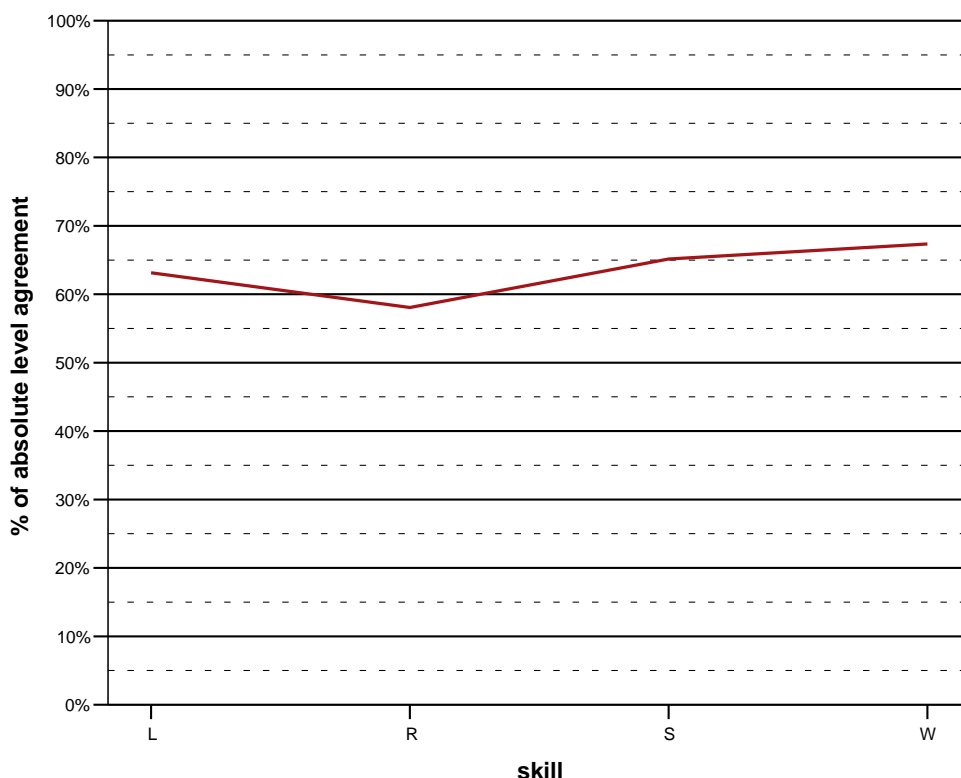
As can be seen from the table above, there is some tendency for an overestimation: 19 descriptors (10%) were assigned to a lower level while 46 (25%) were assigned to higher level. The same tendency is confirmed by other analyses (differentiation analysis and a paired t-test).

This tendency for overestimation is not, however, present for all skills. For writing the difference is not significant although still negative. The strongest tendency for overestimation is observed for reading. This means that Finnish experts tend to place FINNS descriptors to higher CEF levels than they were initially targeted to. The reasons are not clear, and they deserve further study. At the designing phase we already found that keeping apart the levels from intermediate to moderately advanced was

complicated. Possibly the descriptors at FINSS levels B12 – C11 are still not mutually distinct enough in their formulations.

As far as it concerns the absolute agreement between the initial levels and aggregated rating, Table 4 shows the percentage of absolute agreement per skill. The differences in the percentage of absolute agreement between skills are not significant.

Table 4. Absolute agreement between FINNS and CEF levels based on aggregated ratings



Discussion

We set out with the principle stated in the first pages of the CEF document: “The construction of a comprehensive, transparent and coherent framework for language learning and teaching does not imply the imposition of one single uniform system. On the contrary, the framework should be open and flexible, so that it can be applied, with such adaptations as prove necessary, to particular situations.” (CEF, 2001, 7) The study conducted provided us valuable insight into the quality of the Finnish syllabus scales as for their relation to the Common European standards.

The majority of the descriptors processed were of moderate or high quality, but there were a number of weak statements as well, whose clarity is either obscured by combining too many aspects of mastery into the same description or by usage of indefinite terminology that is not equally transparent to all users. Ideally, we would wish to replace the weak descriptors by new formulations or drop them out of the scale. Doing this is not a clear-cut matter, though, as the CEF was not the only point of reference we consulted. The link to the other sources should first be studied as well. This would mean running the same research procedure for the Canadian Benchmarks. Another calibration should be done against the supplement from 1999, which presents the criteria for a good grade in a school subject at the end of the compulsory basic school. The ultimate aim would be a firm empirical calibration of the two scales discussed. IN this work, we will also draw on qualitative feedback by the users of the FINSS and hope to continue the effort towards user-friendly and informative standard descriptions of the levels attained by Finnish school education.