Realizations of the CEFR requirements in the ECL examinations

László Háry PhD – Judit Huszti dr.

Key words: cultural diversity, language testing, CEFR, alignment, ECL examinations

Our modern age can witness the growing importance of language competence both on the level of the individual and that of the communities as well. Language knowledge is the prerequisite of communication on the global level. Within the frames of lifelong learning special emphasis is put on language learning. A language examination is an effective motivating drive to assess the achieved knowledge and certify it with a help of a language certificate. The different EU organizations are also aware of the importance of maintaining cultural and linguistic diversity. In the unifying Europe the common goals, the joint activities and the mobility of labour require the transparency of qualifications and among them the transparency and comparability of language certificates. Language learners and language teachers are familiar with several internationally recognised language examination systems. The picture is rather diverse. There are examinations which can be taken only in one language, some offer two or more languages. There are some with two levels some with up to twelve levels, in some of them there are only written parts while in the others there are only oral tasks. They differ according to their philosophy, the assessed skills, the method of assessment etc. We need a tool which is transparent, coherent and helps learning, teaching and assessment as well. Recognizing the need of comparability a long preparatory work started in Rüschlikon in 1991 with a symposium "Transparency and Coherence in Language Learning in Europe: Objectives, Evaluation, Certification". The milestones of this process are 1997 when the Common European Framework of Reference was compiled, 2003 with the Manual for Relating Language Examinations to CEFR and *Reference Supplement* to the preliminary pilot version of the Manual for Relating Language examinations to the CEFR, the Manual Project involving 12 languages, 20 countries, 40 institutions and the end of 2008 when the revised version will be published with a series of case studies. There are reference materials like calibrated sample oral performances and sample tasks in writing, calibrated sample items and tasks for reading and listening. And we should not forget about a new professional

organization the European Association for Language Testing and Assessment. In the mission statement it claims among others to improve language testing and assessment systems and practice in Europe and to make expertise in language testing and assessment readily available for the solution of assessment problems

In the Hungarian context there is a significant increase in the number of candidates at language exams. The cause of this popularity can be explained by the legislation. A successful language exam may mean a privilege. Bonus points are allocated to university students at the entrance examination, university graduates have to present a language certificate on receiving their degree, at workplaces employees can get a bonus if they have a language certificate. All these measures were meant to raise the language competence of the Hungarian population. To assure quality control and promote good practice in language testing the Ministry of Education founded the Accreditation Centre for Foreign Language Examinations and obliged with a ministerial decree the language examination providers to link and to prove this linkage of their examination levels with the CEFR. In the meantime the Hungarian Association of Language Examiners and Measurement Specialists came into being to help this work.

In case of assessment it is a central issue that how objectively the levels and contents of the language competence can be defined. The traditional Hungarian measurement practice tried to use the longstanding beginner, intermediate and advanced categories to assess the achieved knowledge. However these levels came into life on a basis of a consensus and not as real definitions excluding international comparability and acceptance.

Below we give an overview of the linking process of the international language examination system, the ECL (European Consortium for the Certificate of Attainment in Modern Languages). After many years of professional preparatory work the Member States of the European Union established a *consortium* in 1992, with *London* as the centre. With the help of the ERASMUS and later the LINGUA programmes its task was to develop a *uniform language* test in the languages of the Member States of the European Union. In accordance with the EU unification policy the Member States that formed the consortium wanted to ensure equivalency and recognition of the certificates in each language *without* having to be validated nationally. In the second phase the centre extended the already existing ECL standards to all the official languages of the EU (EU grant: PREPARATORY VISITS (3) LINGUA D 1998, LINGUA D 1999-2002). The Foreign Language Centre of the University of Pécs submitted a LINGUA-D project proposal called ***"Equal Rights for the Less Widely Used and Less Taught Languages of the EU"*** in July 1999. The project proposal was approved by the commission in charge in Brussels and a significant grant was awarded for the

implementation of the project. The ECL examinations assess language competence in four skills on four levels. The construct of the exam is based on the communicative competence. At the time of the publishing the CEFR it was clear that test development should follow the European trend and test constructors should go in line the with the CEFR level descriptions. The ECL examinations joined the Manual project and provided detailed report on the works done. At a conference *Reflection on the use of the preliminary pilot version of the Manual for "Relating Language Examinations to the CEFR"* in Cambridge on 6 December 2007 the academic co-ordinator gave an account of the steps of this work.

According to the Manual there are four stages of the alignment: familiarization, specification, standardization, empirical validation. The aim of the first stage was to familiarize all the test constructors, raters, markers, item writers with the CEFR descriptors and arrive at a common understanding of its terms and levels and every participant had to internalize the level descriptors. There were familiarization trainings two times. The first time at the very beginning of the procedure before specification and the second time before standardization. The participants were the core team of test developers who whose responsibility was to rewrite the specifications. The team studied the CEFR individually, then they used the methods recommended by the CEFR (self assessment of own language level in a foreign language, sorting into piles by level the descriptors, reconstructing CEFR tables, sorting DIALANG scales etc). At the second familiarization training the participants were those who standardized the test tasks. The experts had different background in assessment. Among them there were experts of the exam centre, experienced language teachers from different levels of education and non expert language users, native speakers and non natives.

In the next stage, the harmonization of the ECL test specifications the test developers analyzed the ECL test content, the cycle of the test construction, marking procedure, analyzing test results, the general and detailed descriptions and compared and contrasted them with the CEFR and modified them where it was necessary. The new test specification became the basic document of the ongoing test construction. It facilitates the construction of the CEFR linked tasks.

During the standardization of judgements achieving the adequate understanding of the CEFR levels participants had to relate local test tasks and oral and written test performances to the CEFR levels. Difference had to be made between receptive and productive skills. Receptive skills – listening and reading comprehension can be assessed by objective, discrete point tasks. In case of the productive skills the team had to evaluate the task and the oral and written performance as well. It is important to say that different communicative goals can be achieved

with performance at different levels. We have to note that an operating test system should standardize its tasks continuously for the sake of the stability of its levels, stable standards but in this process the emphasis was put on other aspects of standardization. The team members solved the listening and reading tasks themselves. They filled in the rating sheet giving answer to the question that the competence of which level is needed for giving the right answer to the give item. The item characteristic of a task was mapped in this way. In case of the productive skills both the sample performances and the local performances were assessed. Then here again the recommended forms of activity were used – individual, pair rating of performance samples, group discussions, standard setting with setting of the cut-off score. The judgement data, degree of agreement between raters, inter-rater and intra-rater reliability were analyzed statistically. The experts used the reference materials (calibrated sample oral perfomances on CD and DVD, calibrated sample items and tasks for reading and listening, calibrated sample tasks in writing) provided by the Language Policy Department of the EC.

As the last stage the empirical validation phase was carried out. On the one hand the internal empirical validation can prove that the tests are reliable and valid measuring tools. The classical test analysis can show such data as true scores, standard error of measurement, standard deviation, variance, reliability of test scores, p-values for item difficulty, item discrimination, the Spearman-Brown formula. Item Response Theory (IRT) can give a picture of the test taker's degree of language proficiency in correlation with his/her probability of responding correctly to each item. This test analysis must be carried out as a natural part of the test construction cycle. On the other hand it must be demonstrated that the new tasks and items are really linked to the CEFR. In this work anchor tests are required. Two tasks has to be administered to a carefully selected group of the target population, one calibrated to the CEFR (an anchor test) and one the linkage of which is to be demonstrated. Both classical test theory (correlation analysis) and IRT (comparative analysis of item difficulty logits) methods can be used. In case of the ECL examinations the calibrated sample anchor task and an ECL task was put in the same test set and was tested on a fairly big target population of secondary school students in a national language competition. The results were analyzed and compared. In this way we could prove empirically that the expert moderated items are really in harmony with the CEFR levels. As there are always new tasks the process of the external empirical validation cannot be closed down.

References

Common European Framework of Reference for Languages (2001), Cambridge University Press

Manual for relating language examinations to the Common European Framework of Reference for Languages. A preliminary pilot version.

http://www.coe.int/T/DG4/Portfolio/?L=E&M=/documents_intro/Manual.html

Reference Supplement to the preliminary pilot version of the Manual for Relating Language examinations to the CEFR

http://www.coe.int/T/DG4/Portfolio/documents/CEF%20reference%20supplement%20version%203.pdf

Gábor Szabó (2008), Applying Item Response Theory in Language Test Item Bank Building in Language Testing and Evaluation, Volume 10. Peter Lang Internationaler Verlag der Wissenschaften