

SurveyLang: A European survey on language competences

Neil Jones

Cambridge ESOL

1 Hills Road

Cambridge Cambridge CB1 2EU

tel +44 (0)1223 553739

jones.n@ucles.org.uk

Multilingualism is at the core of EU language policy. Now with its very own Commissioner, multilingualism is promoted as a key determinant of the richness of European culture and society. More instrumentally, improving language skills in Europe has been made an important objective within the drive to improve skills and competences as part of the so-called Lisbon growth and jobs strategy. In March 2002 the European Council called for further action “...to improve the mastery of basic skills, in particular by teaching at least two foreign languages from a very early age.” They also called for the establishment of a linguistic competence indicator. The linguistic competence indicator is in fact one of sixteen core educational indicators aimed at monitoring progress towards the Lisbon objective of becoming the most competitive knowledge-based economy in the world by 2010. Specifying the requirements has taken longer than originally intended. Invitations to tender were finally issued in mid 2007 and SurveyLang¹, a consortium led by Cambridge ESOL, was finally confirmed as the successful bidder in February 2008. SurveyLang presented detailed project plans to the Commission’s advisory board in June 2008, and at the time of writing these have just been accepted. The project is thus about to get officially under way, with systems trialling scheduled for 2010 and the survey itself for 2011.

¹ The SurveyLang consortium members and the countries represented are as follows:

Centre international d’études pédagogiques (CIEP) – France, Gallup Europe, Goethe-Institut – Germany, Instituto Cervantes – Spain, National Institute for Educational Measurement (Cito) - The Netherlands, Universidad de Salamanca – Spain, Università per Stranieri di Perugia– Italy, University of Cambridge ESOL Examinations - UK

In this paper I will describe the project in outline, looking at particular challenges posed by the terms of reference. I will focus on the design of the language tests and in particular at how the requirement to measure outcomes against the Common European Framework of Reference (CEFR) can be addressed. The critical issues here are the validity of the tests themselves and the consistent, comparable interpretation of outcomes in terms of CEFR levels.

To measure and compare language proficiency levels in school settings across Europe is a challenging task. Languages are introduced at very different ages, taught with differing duration and intensity, and as compulsory or optional subjects. Exposure to languages outside school varies, as does the impact of the culture which the language represents. The range of achievement within a grade-based cohort will be very wide, which is a challenge for efficient test design. These considerations alone make it clear that summarizing outcomes in a simple league table of countries will not make much sense (although it will certainly be done). Additionally the terms of reference, which reflect the deliberations of the advisory board and pragmatic judgments of practicality, set further constraints on the scope of the survey. The major requirements are as follows:

- Two languages are tested in each country: the first and second most taught official European languages in that country, from a choice of English, French, German, Italian and Spanish.
- The sample should comprise pupils in education and training at the end of ISCED 2 level (lower secondary). If a second language is not taught at ISCED 2 level then the sample should comprise the second year of ISCED 3 (higher secondary).
- Test performance should be interpreted with reference to the CEFR.
- Three language skills only are to be assessed: Listening, Reading and Writing (as Speaking was considered impractical in the first iteration of the survey).
- Each sampled pupil will be tested in one language only.
- Tests should be in both computer-based (CB) and paper-based (PB) form, and adaptive if possible.

Several issues, evident or less evident, arise from these requirements. There are countries (Belgium a notable example) where what are mother-tongue languages in one region are learned as foreign languages in another. Such countries will require specific treatment. Other countries will have a first or second foreign language which is not an official European language, e.g. Russian in some Baltic countries. Such languages will not be catered for within the survey. The requirement to sample from one of two levels in a country, depending on when a language is taught, turns out on close examination to present specific sampling problems, so that countries will require case-by-case treatment. The mixed-mode administration (CB, PB) presents potential problems of comparability, particularly if adaptivity is introduced.

The major limitations are the restricted range of languages and the omission of Speaking. The Commission has already undertaken that future iterations of the survey will include all European languages and all four skills. This will in turn raise new challenges which will not be discussed here, but which will most likely require significant modifications to the large-scale survey approach taken to this first iteration. Critics have also pointed out that to test only pupils currently studying a language in full-time education, and only in one language, will fail to give a true picture of the extent of plurilingual competence in Europe. This is a valid criticism, although I will suggest ways it might be mitigated.

The Commission specifies the CEFR as the framework against which to measure language learning outcomes for this survey. Since its publication in 2001 the CEFR has exerted a considerable influence on language teaching in Europe, at least at the policy level. Thus for the survey the CEFR is both the explicit point of reference, but also one of the likely variables whose impact is to be determined. Moreover, the adoption by the European Union of the CEFR as an educational performance indicator can itself be expected to strengthen its influence on language policies, in which case the survey will not simply provide an objective snapshot of language teaching and learning in Europe, but also act as an instrument of change.

The approach to language testing which we have proposed for the survey sets out to reflect the CEFR's action-oriented, functional model of language use, while ensuring relevance for 15-year-olds studying language in a school setting. We are aware of the potential educational impact of the survey, and also that the CEFR has been criticized

by some language testers for having more to say about the outcomes of learning than the inputs to, or the process of, learning itself. This partly reflects the unintended prominence which the can-do descriptor scales have achieved relative to the actual text of the CEFR. The reader who focuses on the scales will be more aware of the social dimension of language in use than the cognitive dimension of language as a developing set of competences, skills and knowledge, this despite the CEFR's intention to provide a 'comprehensive, transparent and coherent framework for language learning and teaching'.

We start from the definition of language use offered by the CEFR:

Language use, embracing language learning, comprises the actions performed by persons who as individuals and as social agents develop a range of **competences**, both **general** and in particular **communicative language competences**. They draw on the competences at their disposal in various contexts under various **conditions** and under various **constraints** to engage in **language activities** involving **language processes** to produce and/or receive **texts** in relation to **themes** in specific **domains**, activating those **strategies** which seem most appropriate for carrying out the **tasks** to be accomplished. The monitoring of these actions by the participants leads to the reinforcement or modification of their competences (Council of Europe 2001: 9, emphasis in original).

In the learning context of the 14–16 year-olds who are the objects of the Survey the above-mentioned *conditions*, *constraints*, *texts*, *themes*, *domains* and *tasks* must be understood not as emerging from the daily exigencies of life, but rather as parameters which are carefully selected and manipulated in order to provide a supportive context for learning, the major determinant of progression being, naturally, linguistic. The testing approach must reflect this. Our approach is to focus on those parameters of the above model which relate most strongly to level: the communicative tasks to be accomplished; the range of topics, in the sense of a progression from the immediate and personal through routine and familiar to increasingly unfamiliar and abstract at the highest levels; and the language activities, cognitive processes and strategies which these engage. The remaining contextual parameters should be sampled and

manipulated so as to achieve relevance, content coverage, and a linguistic challenge appropriate to the general level.

We develop a socio-cognitive model (Weir 2005, Jones and Saville 2007), which gives due weight to these two dimensions. The constructs of Reading, Listening and Writing are then defined at each CEFR level covered by the survey, in three ways: indicative CEFR descriptors describe aspects of language in use; posited cognitive processes and knowledge relate to the socio-cognitive model; a list of contextual parameters (range, themes and texts) indicate appropriate contexts for testing. On the basis of this description a set of abilities or ‘subskills’ are identified for each skill. There is no claim that these are all in any sense discrete or empirically distinguishable; together however they assure adequate coverage of the defining features. Some of these refer more explicitly to formal language knowledge, or what might be called enabling skills (e.g. understanding ‘general existential, spatial or relational notions’) while others are more clearly functional (e.g. ‘understanding signs, notices and announcements’). These two aspects of competence are reflected over all the skills and levels tested. This differentiation of testing focus is in fact evident in the exams offered by the five language partners, and reflects a practical consensus regarding validity and relevance to classroom practice. Defined in this way the construct should be as neutral as possible with respect to curriculum or methodology, while remaining faithful to the survey’s remit to measure against the CEFR.

Having defined these testable abilities at each proficiency level, we map them to specific task types, drawing chiefly on types which have been used by the consortium’s language partners in their exams. In this way we hope to ensure a rigorous implementation of the construct that ensures comparability across languages. Nearly all test material will be written specifically for the survey, and the proposed collaborative development process is also critical to ensuring quality and comparability across languages. Several major multilingual testing projects have been compromised by failure to achieve adequate consistency in this respect (the EBAFLS project, supported by the European Commission under the Socrates programme, is perhaps a recent example, although no final report has yet been published).

Another major challenge for the Survey is standard setting: reporting learners' performance in relation to CEFR levels in a way which is defensible and demonstrably comparable across languages. The approach to language test design described above should ensure that test tasks will measure the same trait across languages, that this trait can be related to the CEFR, and that tasks at each level will be broadly comparable in terms of the challenge they present. However, assigning students to CEFR levels on the basis of their test performance is an additional interpretative step. This is a current active area of research in the language testing community, much of it focussed on the pilot *manual* for aligning exams to the CEFR (Council of Europe 2003, 2004). While the details are yet to be established, a combination of approaches will clearly be necessary. For Writing, standard setting will be based on samples of written performance. For Reading and Listening tests a test centred standard setting method will be used. That is, the approach described in the manual will essentially be followed. However, the pilot version of the manual deals with the case of linking a single language and test to the CEFR. It says nothing about the explicit comparison across languages which is clearly essential for the Survey. One can in fact argue that setting standards language by language is not strictly possible: as the CEFR's frame of reference takes in all European languages then the correctness of a standard set for any language can only be evaluated by comparison with other languages. Before making absolute judgments of level we should rather be thinking in terms of comparative judgments: rankings rather than ratings.

The use of rank-ordered data to do test equating is a relatively recent development. See Bramley (2005) for an example and discussion. A recent study involving explicit cross-language comparison is a multilingual benchmarking conference organised by CIEP at Sevres in June 2008. This concerned Speaking samples at all levels of the CEFR in five languages (English, French, German, Italian, Spanish). Rating data against the CEFR was elicited at the conference, while ranking data were collected from the same judges in an online study organised prior to that.

This valuable event points the way forward for cross-language alignment, and lessons learned here will be fed into the methodology developed for the survey. Logically the alignment of languages to a single scale, which can be achieved elegantly by ranking, should precede standard setting, because it ensures that the standards set are

automatically comparable across languages, which is, finally, the fundamentally important issue.

Several approaches to validating the standards set will be explored. A can-do study using self-report by sampled students is planned. There has also been discussion of developing a calibrated exemplar bank of texts (Listening and Reading) with different degrees of comprehension simply described – that is, an attempt to provide an intuitive and concrete description of progression that could be used both to agree and then to communicate the meaning of the standards set. Such a bank would make better intuitive sense to many stakeholders than the bank of test tasks itself, where item difficulty is a complicating factor alongside text difficulty. However, performance on the two banks could be empirically aligned in order to validate a standard setting.

The second major instrument of the Survey is the set of contextual questionnaires which will be administered to individual students, foreign language teachers, school principals and the national research coordinators (that is, the body implementing the Survey in each country). It is the questionnaire data which will allow us to interpret the language test outcomes, and to detect context factors that are related to foreign language achievement. As noted above, the context of foreign language learning differs widely between nations, and these differences provide us with an opportunity to assess how system-wide factors may relate to achievement. Major themes that might be explored include:

- demographic, affective and experiential characteristics of individual learners and teachers
- features of curricular and instructional practices;
- economic and specific language policy issues, which can be explored at the school or national level.

Developing the questionnaire will involve agreeing with countries the particular themes to be included, given the limited time available (30 minutes) for the student questionnaire.

There are many other aspects to conducting the Survey that I will not discuss here, such as sampling issues, or the comparison of computer-based and paper-based tests,

and the logistical difficulty of delivering both formats. All in all this will be an exceptionally challenging project, and a sensitive one too: few countries seem confident of performing well with respect to the European multilingual target. At the same time it is an important project for the CEFR and the nature of its impact on teaching and learning. Surveys into the use of the CEFR conducted by the Council of Europe in 2005 and 2006, and the Council of Europe Intergovernmental Policy Forum 2007, showed that despite the CEFR's considerable impact at a policy level, it is not yet found widely relevant to the teaching profession at school level. According to these sources, teachers still find the CEFR complex and inaccessible, and need better support in working with it. Simplification and more accessible presentation is needed. There is a lack of awareness of the richness of the framework as a point of reference for teaching and learning; attention tends to focus in a reductive way on the reference levels and the descriptive scales. Making the CEFR the benchmark against which progress towards a multilingual Europe is measured certainly carries the risk of further emphasizing the assessment of outcomes over the creative discussion of teaching and learning. At the same time the Survey provides an opportunity, both in the way it is implemented and its results presented, to link classroom practice and curricular objectives more closely to the CEFR in ways that build on best practice in each country, and thus exert a positive impact on language learning.

References

Council of Europe (2001): *Common European Framework of Reference for Languages*, Cambridge: Cambridge University Press

Council of Europe (2002): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Case Studies*. Strasbourg: Council of Europe Publishing

Council of Europe (2003) Relating language examinations to the CEFR. Manual; Preliminary Pilot Version. Retrieved from: [http://www.coe.int/T/E/Cultural Co-operation/education/Languages/Language Policy/Manual/default.asp](http://www.coe.int/T/E/Cultural_Co-operation/education/Languages/Language_Policy/Manual/default.asp)

Weir C (2005): *Language Testing and Validation*, Basingstoke: Palgrave Macmillan Ltd.

Jones N and Saville N (2007) Scales and Frameworks. In Spolsky B. and Hult F. M. *The Handbook of Educational Linguistics* pp 495-509. Malden MA: Blackwell Publishing

Bramley, T. (2005) A Rank-Ordering Method for Equating Tests by Expert Judgment. *Journal of Applied Measurement* 6(2), pp202-223