French edition :

Cadre européen commun de référence pour les langues:
Apprendre, enseigner; évaluer.
Etudes de cas

ISBN 92-871-

Council of Europe Publishing
F-67075 Strasbourg Cedex

ISBN 92-871-

Common European Framework of Reference for Languages: Learning, Teaching, Assessment

CASE STUDIES

CONTENT

Foreword

Using the Common European Framework in language teaching and assessment

An Introduction to the Common European Framework and to this volume

The Council of Europe, through its language programmes in Strasbourg, has developed over 30 years a series of monographs on language learning, language use and language proficiency. Probably the best known of these are the different "levels" of language achievement within language teaching and learning "schemes" - the Threshold Level (1975, 1990, 1998), which has been enormously influential throughout Europe and beyond, and later the Waystage and Vantage Levels. In the mid 1990s, the Council of Europe began the challenging task of putting these different levels and their associated guidelines together into one coherent Framework. This Framework, which is referred to as the Common European Framework of reference for language learning, teaching and assessment (CEF for short) has been developed by Dr John Trim, Dr Brian North, Professor Daniel Coste and Mr Joseph Sheils. It has been informally published in two versions so far: an initial version in 1996, and after consultation, a revised version in 1998, known as Draft 2. Extensive feedback and discussions have since led to the production of the current version of the Framework, which was published in English by Cambridge University Press and in French by Didier, to coincide with the European Year of Languages, 2001.

What is the Common European Framework?

The following introduction is taken from the 2001 version of the CEF:

"The Common European Framework provides a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe. It describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively. The description also covers the cultural context in which language is set. The Framework also defines levels of proficiency which allow learners' progress to be measured at each stage of learning and on a life-long basis.

The Common European Framework is intended to overcome the barriers to communication among professionals working in the field of modern languages arising from the different educational systems in Europe. It provides the means for educational administrators, course designers, teachers, examining bodies, etc., to reflect on their current practice, with a view to situating and co-ordinating their efforts and to ensuring that they meet the real needs of the learners for whom they are responsible.

By providing a common basis for the explicit description of objectives, content and methods, the Framework will enhance the transparency of courses, syllabuses and qualifications, thus promoting international co-operation in the field of modern languages. The provision of objective criteria for describing language proficiency will facilitate the mutual recognition of qualifications gained in different learning contexts, and accordingly will aid European mobility." (CEF, 2001, p12)

The Framework itself is a formidable document, describing as it does the complexity of language learning, language use and language proficiency, and drawing the implications for language pedagogy. Not surprisingly, a "Guide for Users"of the Framework has been developed alongside the Framework, to help users to interpret and adapt the Framework for their own uses.

However, even as the Framework was being drafted in its original form, interest throughout Europe was such that efforts were made in many countries and institutions to implement aspects of the emerging Framework in a variety of domains - in teacher training, in syllabus design, in materials development and, latterly in particular, in assessment, not only in formal tests and scales of assessment but also in portfolios of achievement in modern European languages.

Partly as a result, a wealth of experience has developed throughout member states of the Council of Europe in applying and developing the Framework. On the occasion of the European Year of Languages 2001, the revised Framework was published, and this sister volume was also produced, containing accounts of the use of the Framework across Europe, both in order to publicise the case studies in their own right, but also in order to encourage others to use the Framework in a variety of different ways and contexts.

This volume thus is addressed at potential users of the Framework, ranging from syllabus designers, materials developers, publishers, teacher trainers, students and teachers on pre-service and in-service courses to language testers, ministry officials and applied linguists in general.

How was the Framework developed?

The 2001 version of the Common European Framework represents the latest stage in a process which has been actively pursued since 1971 and owes much to the collaboration of many members of the teaching profession across Europe and beyond. In particular, many colleagues and institutions across Europe responded, often with great care and in concrete detail to the request for feedback on the First Draft, circulated in the autumn of 1995. Well over 1000 copies of the First Draft were sent out for comment and more than 200 questionnaires were returned. A Second Draft was issued in 1996 and submitted for scrutiny and approval to the Final Conference of the Modern Language Project in April 1997, which recommended that the Framework should be trialled in use in the next stage of development.

Feedback on the use of the Framework in practice was gathered from a representative sample of specific users, and in addition, other users were invited to communicate to the Council of Europe the results of their experience in using it for their particular purposes. The feedback received was taken into account in revising the Framework and User Guides prior to their adoption on a Europe-wide basis, and the publication of the Framework by Cambridge University Press and Didier in 2001.

A detailed account of the early origins of the Framework itself is contained in Brian North's first chapter in this volume, entitled *Developing Descriptor Scales of Language Proficiency for the CEF Common Reference Level* (see Chapter Seven in this volume). In this chapter, North outlines the history of the project, and the development of the levels and scales which accompany and operationalise the Framework.

What is the point of the Framework?

The Framework itself was written with the following two main aims:

1. "to encourage practitioners of all kinds in the language field, including language learners themselves, to reflect on such questions as:

    - what do we actually do when we speak (or write) to each other?
    - what enables us to act in this way?
    - how much of this do we need to learn when we try to use a new language?
    - how do we set our objectives and mark our progress along the path from total ignorance to effective mastery?
    - how does language learning take place?
    - what can we do to help ourselves and other people to learn a language better?

2. to make it easier for practitioners to tell each other and their clientèle what they wish to help learners to achieve, and how they attempt to do so." (CEF, 2001, p7)

The authors of the Framework are emphatic that they do not wish to tell practitioners what to do, or how to do it. Rather "we are raising questions, not answering them. It is not the function of the Common European Framework to lay down the objectives that users should pursue or the methods they should employ" (loc cit).

The intention in publishing the Framework is to encourage all concerned with the organisation of language learning to base their work on the needs, motivations, characteristics and resources of learners.

"This means answering questions such as:

- what will learners need to do with the language?
- what do they need to learn in order to be able to use the language to achieve those ends?
- what makes them want to learn?
- what sort of people are they (age, sex, social and educational background, etc.)
- what knowledge, skills and experiences do their teachers possess?
- what access do they have to course books, works of reference (dictionaries, grammars, etc.), audio-visual aids, computer hard - and software, etc.?
- how much time can they afford (or are willing, or able) to spend?" (op cit, p 7/8)

What does the Framework look like?

The contents of the CEF are as follows:

Readers of this volume, as well as users of the Framework, will find the following summary of the various chapters useful:

Chapter 1 defines the aims, objectives and functions of the proposed Framework in the light of the overall language policy of the Council of Europe and in particular the promotion of plurilingualism in response to European linguistic and cultural diversity. It then sets out the criteria which the Framework should satisfy

Chapter 2 explains the approach adopted. The descriptive scheme is based on an analysis of language use in terms of the strategies used by learners to activate general and communicative competences. These strategies and competences are engaged to carry out the activities and processes involved in the production and reception of texts and the construction of discourse dealing with particular themes, and they enable learners to fulfil the tasks facing them under the given conditions and constraints in the situations which arise in the various domains of social existence.

Chapter 3 introduces the common reference levels. Progress in language learning with regard to the parameters of the descriptive scheme can be calibrated in terms of a flexible series of levels of attainment defined by appropriate descriptors. This apparatus should be rich enough to accommodate the full range of learner needs and thus the objectives pursued by different providers, or required of candidates for language qualifications.

Chapter 4 establishes in some detail the categories needed for the description of language use and the language user/learner according to the parameters identified, covering in turn: the domains and situations providing the context for language use; the themes, tasks, purposes and strategies of communication; communicative activities and processes, and text; especially in relation to activities and media.

Chapter 5 categorises in detail the user/learner's general and communicative competences and strategies.

Chapter 6 considers the processes of language learning and teaching, dealing with the relation between acquisition and learning and with the nature and development of plurilingual competence, as well as with methodological options of a general or more specific kind, in relation to the categories set out in Chapters 3 & 4.

Chapter 7 examines in greater detail the role of tasks in language learning and teaching.

Chapter 8 is concerned with the implications of linguistic diversification for curriculum design and considers such issues as: multilingualism and multiculturalism; differentiated learning objectives; principles of curriculum design; curricular scenarios; life-long language learning; modularity and partial competences.

Chapter 9 discusses the various purposes of assessment and corresponding assessment types, in the light of the need to reconcile the competing criteria of comprehensiveness, precision and operational feasibility.

The appendices deal with some further aspects of scaling. Appendix A deals with general and theoretical concerns for the benefit of users who wish to develop scales specifically geared to a particular population of learners. It discusses the establishment of scaling and levels. Methods of and criteria for, scaling and the requirements for formulating descriptors are explained.

Appendices B to D give information concerning scales developed by other agencies, such as the Swiss Language Portfolio Project, DIALANG (a project funded by the European Commission to develop diagnostic tests in 14 European languages for delivery over the Internet) and the "Can-do" scales of ALTE (Association of Language Testers in Europe).

The content of this book

Each chapter in this book is written by authors who have considerable experience of using the Common European Framework in their professional context, and they exemplify in some detail their varied experience. Geographically, they range from Ireland to Poland, from Spain to Finland, and in educational settings, the contexts range from language learning in primary education through to placement on regular university courses, to learning in an Open University context, and to independent adult learning.

The book is organized in such a way as to allow readers to develop their understanding of the Framework as well as of the various ways in which it has been used, in a cumulative fashion. Thus, it is recommended that readers who are not too familiar with the Framework should work their way systematically through the volume, starting at Chapter 1, since the accounts of the use of the Framework in educational settings in Poland and Catalonia, followed by a detailed description of the use of the Framework for the development of syllabuses and materials for teaching Spanish at the UK Open University, cover a wide range of possible uses of the Framework. Subsequent chapters build on this, and address particular aspects of the language education curriculum or the use of the Framework in different sectors of education.

However, other readers will already be more familiar with the Framework, and may prefer to concentrate on particular aspects such as assessment and scale development, syllabus development, or action research in classrooms. For the benefit of such readers, a brief summary is provided below, but each chapter begins with an introduction to the topic(s) it addresses, for the guidance of readers wishing to concentrate on particular areas.

The first two chapters present an overview of the use of the Common European Framework (henceforth CEF) in two very different educational and geographical settings - one in Central Europe, emerging from decades of central control, where in Poland at least the CEF was seen to offer choices in language education rather than the continuation of centralised prescriptivism, and its role in promoting diversity in various educational reforms was greatly valued. In Southern Europe, on the other hand, the autonomous Spanish province of Catalonia has a long tradition of innovation and reform in language education, and the CEF and its predecessors have had considerable influence. There the use of the CEF is seen as the culmination of a long process, rather than a revolutionary new venture. In contrast with Poland, perhaps, in Catalonia the CEF is seen as offering the possibility of a "common basis" and equivalences across the different languages taught in the region.

The first chapter, by Hanna Komorowska of Warsaw University, entitled The Common European Framework in Poland outlines how the work of the Council of Europe contributed to the Polish educational reform of 1999, specifically in four areas: syllabus design for primary and secondary education; upgrading curricula for teacher education; ongoing evaluation and self-evaluation of learners; and assessment of language skills in adult education and in-service training. . Komorowska also makes interesting comments on the dissemination and promotion of the CEF, and on the accessibility of the documentation.

Chapter 2, by Neus Figueras, of the Department d' Ensenyament, Generalitat de Catalunya, and Joan Melcion, from the Universitat Autonoma de Barcelona. is entitled The Common European Framework in Catalonia. It also provides a description of the implementation of the CEF across the language education curriculum. In a broad overview of developments in Catalonia, it illustrates the use of the CEF both as a point of reference and as a "tool for reflection", as the authors call it, for teachers and administrators. Representing the State of the Art, in the opinion of the authors, the CEF provides an opportunity to challenge previous concepts of language teaching, and to update teachers and testers in the latest thinking in the field. The chapter also looks forward to future developments in the region in the use of the CEF, but closes with a number of important questions addressed at administrators, and arguably, at the Council of Europe itself. The authors ask who will be responsible for ensuring that syllabuses, tests, assessment portfolios and the like, that claim to be based on and related to the CEF, are indeed actually so based and do indeed reflect the philosophy and even the levels of the Framework that are detailed in its pages?

Chapter 3, by Cecilia Garrido and Tita Beaven, is entitled *The Common European Framework of reference: The UK Open University experience*. It presents a detailed account of how the Common European Framework was consulted in order to develop Spanish courses at tertiary level, and in particular both the syllabus and accompanying audio-visual materials. The chapter goes into useful detail describing how each chapter of the CEF proved to be relevant, and how the various ideas contained in the CEF were used.

Chapter 4, by Hanna Jaakkola from the Department of Teacher Education, University of Helsinki, and colleagues in teacher training institutes and schools in Finland, continues the detailed description of the use of the CEF in particular aspects of language education, dealing with action research in language classrooms. Entitled *How to promote learning to learn in first foreign language classes*, it concentrates on that aspect of the CEF that describes and discusses learning how to learn, and reports on the value of the Framework in guiding the work of a number of teachers in Finnish schools – at primary, lower secondary and upper

secondary levels. The authors demonstrate convincingly that those features of the CEF that deal with existential competence, study skills, heuristic skills and metacognitive reflection (CEF, Chapter 5) can indeed be applied to the various settings, and that pupils and their teachers can both be shown to have benefited and developed in the course of the action research.

The theme of learner reflection and awareness is continued in Chapter 5 by David Little, Barbara Lazenby Simpson and Fiona O'Connor of the Centre for Language and Communication Studies, Trinity College Dublin. This chapter, entitled *Meeting the English Language Needs of Refugees in Ireland*, also introduces a major new theme, itself inspired by the work of the Council of Europe and closely linked to the Framework, namely the European Language Portfolio, or ELP. The authors report on a project that has used the CEF, as well as the ELP, to develop English language provision for refugees admitted to Ireland. The chapter shows how the CEF was used to develop English language proficiency benchmarks for the curriculum, and illustrates the use of the ELP as a pedagogical tool to encourage reflective learning and learner autonomy.

Chapter 6, by Peter Lenz and Günter Schneider of the University of Fribourg, goes into much more detail on the background to, and the nature of the European Language Portfolio and its development. Entitled *Developing the Swiss version of the European Language Portfolio* this chapter is an excellent introduction and guide to the work of the Council of Europe in fomenting the development of alternative assessment methods to traditional language examinations. The authors clearly illustrate the synergy between the CEF and the ELP, present an account of how the Swiss Portfolio was developed, and reflect on future developments in portfolio work.

Brian North of the Eurocentres Foundation was closely involved in the development of the scales that are included in the CEF, and indeed in the development of the CEF itself. In the first of two chapters by North in this volume, Chapter 7, entitled *Developing Descriptor Scales of Language Proficiency for the CEF Common Reference Levels* is an excellent overview of the history of the development of the CEF scales which earlier chapters have already shown to be highly useful and influential. The CEF descriptors are illustrated in detail and their value in the devising of assessment systems as well as curricular levels is emphasized by this contribution.

The next chapter, *Council of Europe Scales of Language Proficiency: a Validation Study* by Felianka Kaftandjieva and Sauli Takala, of the Centre for Applied Language Studies, University of Jyvaskyla, Finland, describes how a European-Union-funded project, DIALANG, took the CEF scales and validated them for its own purposes. Although somewhat technical, a careful reading of this chapter will be very rewarding, as it emphasizes the importance of empirical testing of any scales that are developed. Readers will be convinced of the value of the CEF scales as a result, and, by reading this chapter in conjunction with the previous one by North, will hopefully become aware of the methodologies by which valid scales can be developed and validated.

Ari Huhta and colleagues also from the DIALANG project, present a detailed description of the project in Chapter 9. The DIALANG project is the first directly to apply the CEF concepts as well as scales to diagnostic assessment. Entitled *DIALANG - A Diagnostic Language Assessment System for Learner,* the chapter illustrates how the CEF influenced DIALANG's Assessment Framework, the Detailed Assessment Specifications and the range of Self-

Assessment statements developed by the project, as well as the reporting scales and even the feedback given to learners on their test performance. The CEF lies at the very heart of DIALANG's development, and the project can be seen in part as an extended exploration of the validity and practicability of the CEF.

Chapter 10, the second by Brian North, also describes the development of a self-assessment instrument, this time a procedure used at the University of Basle to place students into appropriate classes when they begin their studies. North links this project to the work of the European Language Portfolio (ELP), and vividly illustrates the benefits to learners, teachers and administrators of a self-assessment instrument. He also shows how such an instrument can be validated, and how it can be used to validate aspects of the CEF itself.

In parallel to the chapters by Huhta et al and North, Chapter 11, by Neil Jones of the University of Cambridge Local Examinations Syndicate, reports how the Association of Language Testers in Europe (ALTE) developed its "Can do" statements of language competence, and related the ALTE Framework to the CEF.

The final chapter in this volume, by Lukas Wertenschlag, Martin Muller and Helen Schmitz, describes the European Level Descriptions for German as a Foreign Language (ENDaF) and their relationship to the CEF. The ENDaF descriptions were produced in order to expand and build on the CEF at the four lowest levels of the Framework, to illustrate and exemplify for German as a Foreign Language what in the Framework is necessarily language-independent, to fill in gaps in the CEF, and to provide all this in a user-friendly format on a CD -ROM. This final chapter thus pays tribute to the CEF by showing how it has inspired detailed further work in the implementation and application of the Framework in one particular language.

Indeed, this whole volume illustrates in rich detail how the Common European Framework has been disseminated, implemented after consultation and adapted. Clearly the influence of the Framework has been widespread and deep, impacting on curricula, syllabuses, teaching materials, tests and assessment systems and the development of scales of language proficiency geared to the six main levels of the CEF. There is no doubt that the influence of this Framework will grow over the coming decade, and it is hoped that the accounts of the use of the Framework in this volume will contribute in some small measure to encouraging the use of and experimentation with the Framework in language education throughout Europe.

J Charles Alderson, Lancaster University (editor)

CHAPTER 1: THE COMMON EUROPEAN FRAMEWORK IN POLAND

Hanna Komorowska
Warsaw University

Introduction

Over the last two decades Poland has had a rich experience in using the documentation of the Council of Europe, especially in two fields: curriculum construction and teacher education, but also recently in evaluation and assessment, and this chapter will give an account of this experience.

1. The work of the Council of Europe in the foreign language teaching and teacher education reform of 1990.

In the mid-eighties the Waystage and Threshold Level documents helped to upgrade syllabuses for secondary schools. The free choice of coursebooks proclaimed by the Ministry of Education clashed with a centralised, outdated syllabus for each of the foreign languages taught in the school system. Syllabuses were, therefore, modified and presented in the form of inventories to avoid strict progression of grammatical structures. Lists of functions and situations to be covered in the 4- year high school programme were also introduced alongside lists of structures and topics. The communicative approach became one of the recommended methods.

In the nineties a significant role was played by the documentation of the Council of Europe in the implementation of educational reforms after the fall of communism. All the foreign languages were granted equal status in the school system. English, German and French were introduced into upper primary education for children aged 11- 15, i.e. for the age group that had previously learnt Russian only. The teaching of two foreign languages was promoted throughout the whole of general secondary education and the teaching of at least one foreign language – throughout secondary vocational education for students aged 15-18\19 both in the so- called full secondary tracks leading to final Matura exams and in the so-called non-full ones leading to vocational qualifications only. Due to teacher shortages combined with low professional mobility, a new dense network of teacher training colleges leading to a BA with a 2 - year follow- up MA at major universities was established and soon more than 60 colleges of English, German, French and Spanish started training about 3 thousand students each year. The aim was to promote the communicative approach in college language teaching and guarantee a high level of teacher education, provided that the recommendations of the Council of Europe Modern Languages Project were followed as closely as was possible in the Polish context. In the process of designing the college programme models and solutions were used coming from materials which had been prepared within the framework of the Council of Europe Project no 4 " Modern Languages: improving and intensifying language learning as a factor making for European understanding, co-operation and mobility" and Project no 12 "Learning and teaching modern languages for communication". Workshop materials published within those projects were also used as supplementary texts for EFL methodology classes as the Council of Europe kindly supplied the first 30 colleges established in Poland in 1990 with full documentation of both projects.

The reform of 1990 has so far demonstrated that the Council of Europe documentation can be used successfully, although it became evident that a lot of work has to be put by experts into the selection of materials on which the educational strategy is based and then also by teacher trainers in the process of adapting workshop materials for teacher education purposes. This led to a generally optimistic approach of the Polish specialists to the Common European Framework and to the Language Portfolio. Attention given in Poland to the Common European Framework is, however, naturally influenced by the present needs in the field of FL teaching. These can be divided into four groups

- syllabus design for primary and secondary education
- upgrading curricula for teacher education
- on-going evaluation and self-evaluation of learners
- assessment of language skills in adult education and in-service training.

But before going into detail, we first need to consider the Polish education reform of 1999.

2. The context of the Common European Framework in Poland – the educational reform of 1999

All the needs listed above became even more prominent when the educational reform was implemented on September 1, 1999. The reform changed the structure of schooling which used to consist of an 8-year primary and a 4 -year secondary school, replacing it with a 6-year primary, a 3-year lower-secondary gimnazjum for all and a 3-year upper- secondary liceum or a 2-year vocational school with a 2 year follow-up leading to final examinations and giving the right to enter universities. A set of diagnostic tests will be offered at the end of the gimnazjum with the first testing to be organised in 2002. It is recommended that assessment should embrace three components: school grades, the learner's portfolio and the student's external test score. New Matura, i.e. a set of final school leaving examinations at the end of a three-year lycee, will from 2002 on combine external testing with a school-based oral exam.

The first foreign language enters the curriculum in grade 4 of the primary school for children aged 10, i.e. a year earlier than in the former system. Foreign language teaching from grade 4 is financed by the Ministry of Education. Primary schools are, however, encouraged to start language teaching earlier and to introduce two foreign languages whenever possible. A second foreign language is introduced from the beginning of the gimnazjum, i.e. for learners aged 13.

Structural changes were accompanied by curricular reform. What is required in the new system is that school teams of FL teachers prepare so-called authoring syllabuses. These authoring syllabuses must comply with a very generally outlined core curriculum (podstawa programowa) issued by the Ministry of Education. The core curricula for primary, lower secondary and upper secondary were prepared by the National Curriculum Committee for the whole educational system and included broad educational aims in the form of the so-called key competencies. Core curricula for particular subject areas were then worked out at the National In-Service Teacher Training Centre and presented to teachers for consultation. On this basis teachers in schools prepare their own authoring syllabuses which have to comply with the final version of the core curriculum and which are then submitted for approval to the educational authorities.

Instead of designing authoring curricula of their own, teachers can also select a curriculum they want to follow from the list of authoring curricula already approved by the Ministry.

The following trends are characteristic of the new educational reform:

- promoting integration of subject areas, which supports the idea of content-based instruction
- departure from obligatory segmentation of the teaching content into subjects, which encourages the formation of broader areas, cross-curricular links and team teaching
- encouraging descriptive evaluation, which calls for new forms of assessment, stresses the role of self-evaluation and shows the value of the language portfolio
- introducing external end-of-school testing which helps to combine summative external evaluation with formative school evaluation and subjective self-assessment.

All these tendencies have clear implications for syllabus design as well as for the choice of teaching methods. In their implementation the social and political context, however, plays an important role. After many years of a command economy, teachers, trainers and educators no longer tolerate top-down, rigid guidelines. What they look for are lists of options to choose from as well as examples or models they could optionally follow. This need can be satisfied in the work with the Council of Europe documentation since a set of options and models is what The Common European Framework actually offers.

3. The role of the Common European Framework in syllabus design for schools

The educational reform of 1999 presents new requirements vis-à-vis authoring syllabuses based on the core curriculum. According to ministerial recommendations issued at the outset of the 1999 reform, the authoring syllabus should include:

- an introduction including the presentation of the syllabus, its origin, information about the educational experience of the authors, characteristics of the students for whom the syllabus is suitable and minimum requirements related to school equipment indispensable for the implementation of the syllabus
- a set of objectives
- a list of teaching contents
- recommendations for educational procedures, methods and techniques
- exit competencies
- a list of testing techniques to be used in formative and summative evaluation
- samples of tests
- samples of lesson scenarios.

In the first stages of the reform schools found it difficult to work on syllabus design and looked for models. To help teachers in their work, The Polish School Publishers' WSiP worked out the first model syllabus series for English, German, French and Russian in primary education and for English, German, French, Russian, Italian and Spanish in secondary education. The model syllabuses were also the first ones based on the CEF in the sense that

- they presented language domains, activities, topics and text types required as well as outlining teaching contents and forms of assessment

- they followed the idea of different profiles for different foreign languages in the CEF syllabus design scenarios (e.g. oral and interactive skills were stressed in the English syllabus as this is commonly the first foreign language for most of the students and is in any case more widely used in the world, but the sociocultural component and reading comprehension skills were stressed in the French and Italian syllabuses, as these two languages are usually learnt as second foreign languages with educational and cultural intentions).

Teachers are free to use or adapt this model, they can also work out their own versions of the syllabus using lists of topics, texts, functions, domains and activities drawn from the Common European Framework or any other source. If their authoring version of the syllabus is favourably reviewed by two reviewers from the list of specialists recommended by the Ministry of Education who confirm that it is based on the national core curriculum and methodologically sound, the syllabus gets the official stamp of the Ministry, becomes officially approved and can be used in schools.

The need to work out individualised authoring curricula is constantly growing as contexts differ and as schools interested in an early start to language learning introduce modern language programmes of varied intensity at various stages of initial education. This makes the use of a uniform syllabus impossible and calls for designing a special syllabus for each particular school.

Various institutions are trying to help teachers in this work. So far, several foreign language syllabuses have been prepared by Polish and British publishing houses to facilitate work with their materials. These have then been approved by the Ministry. The syllabuses stress integration, cross-curricular content and activities which are prepared by teams of teachers, e.g. a foreign language teacher and a classroom teacher at the primary level or teachers of several subject areas at the secondary level. They also include clear practical suggestions related to evaluation and testing. Each of the documents, however, bears its own characteristic features. For example, the syllabus prepared by the School Section of the Polish Scientific Publishers' (Wydawnictwo Szkolne PWN) for initial education encourages integration of English with the primary curriculum through team teaching and full co-operation between the English teacher and the classroom teacher. The syllabuses prepared by the same publishing house for upper primary and lower secondary emphasise operationalising educational objectives in order to facilitate learners' self-evaluation and introduce them to autonomy. The syllabus prepared by Oxford University Press for the primary school pays special attention to ways of catering for needs springing from the individual learning styles of children while the syllabus prepared by the same publishing house for lower secondary concentrates on communicative methodology and can thus function as a great help for teachers in selecting teaching and testing techniques for the classroom. Syllabuses prepared by the Juka-91 publishing house for the primary and for lower secondary give special attention to cross-cultural understanding, language across the curriculum as well as to creativity and self-expression through music and art. The syllabus prepared by Cambridge University Press stresses an activity-based approach. Syllabuses designed by a team from Pearson Education offer an innovative system of evaluation and self-evaluation to help the learners plan, manage and self-evaluate their own work.

These examples demonstrate that each authoring syllabus complies with the requirements of the national core curriculum and follows a general communicative line, at the same time trying to promote characteristic features of its own. This guarantees a degree of uniformity in

the educational system, enabling teachers to make choices appropriate for their learners and suitable for their own teaching styles, which is perfectly in line with the ideas presented in the Common European Framework and the Portfolio.

4. The Common European Framework in curriculum design for teacher education

The laws of the new educational reform which relate to authoring versions of syllabuses springing from the common core curriculum also hold for institutions of teacher education. This means that a new core curriculum is needed for the teacher education system now in its 10th year of functioning, especially as teacher trainees in their future work will have to be able to develop skills listed in the core curriculum for primary, lower and upper secondary schools. These skills, often referred to in the ministerial core curriculum as key competencies include self-management, communication, information, work and study skills presented with lists of detailed sub-skills and are accompanied by sets of language skills to be developed. Last but not least, practical language syllabuses in teacher training colleges should also comply with the new approaches to language teaching and learning.

This means that out of the four components of the former, pre-1999 reform teacher education curriculum – three, i.e. the language component, the methodology component and the pedagogical component need upgrading, while the academic background studies component can remain without extensive modifications. Changes are introduced by a committee of university professors and college directors attached to the National Council of Teacher Education. The final draft will be approved by the Minister of Education.

Draft 1 now being completed stresses

in the language component:

- interactive and mediating skills in language activities
- savoir-être and savoir-apprendre in general competencies

in the methodological component:

- criteria for the selection of coursebooks and supplementary materials springing from the identification of domains, activities, topics and text types approaches to error
- learning and communicative strategies and their facilitation
- self-evaluation, self-learning and learner autonomy

and in the pedagogical component:

- the competencies-based approach to learning outcomes and
- the positive I CAN approach in learner-centred language education.

The influence of the Common European Framework can be seen in the competencies-based approach, in the introduction of the learning to learn component and the ideas of self-evaluation, self-learning and learner autonomy as well as in the emphasis given to the learning and not only to the teaching process.

5. The Common European Framework in on-going evaluation and self-evaluation of learners

Autocratic teaching styles as well as managerial and judgemental roles of the teacher are deeply rooted in the Polish educational tradition. Since the fall of communism the teaching of foreign languages has done a lot for the democratisation of classroom interaction. Learner autonomy, however, was not sufficiently stressed at the beginning of the 90s and, in consequence, learning outcomes are still very much the responsibility of the teacher.

The situation has been changing since Polish teacher trainers became acquainted with ideas presented in the Common European Framework and in the European Language Portfolio. Such ideas open the way to learner autonomy where the learner takes over the responsibility for his own learning, develops skills of self-evaluation and finds opportunities to use his new learning to learn skills. As this approach is more than needed in the Polish social and political context, it is now being promoted in pre-service and in-service teacher education. Introducing learner autonomy is gaining more and more popularity among teachers as it is seen as a tool which can save the teacher effort, at the same time ensuring the desired learning outcomes. Teachers of other subject areas welcome the newly acquired self-dependence, interpersonal skills and learning to learn know-how in learners and are beginning to introduce learner autonomy in non-language fields. It is not the first time in the post-war history of Poland that foreign language teaching has promoted innovative ideas throughout the whole of the educational system.

Under the influence of the Common European Framework new primary textbooks are now being prepared specially for the reformed school, e.g. the "Bingo" series published by the School Section of the Polish Scientific Publishers' (Wydawnictwo Szkolne PWN) and "Surprise!" published by Juka-91. Publishers include special I CAN and I LIKE\ I DON'T LIKE sections which help the learners to self-evaluate as well as enable the children to evaluate language tasks and communicative activities. The sections encourage learners to formulate their opinions on lesson and textbook activities and to plan further work on the basis of self-evaluation. At the same time these sections provide accountability for parents who pay for their children's language education. As the Ministry of Education covers the cost of 3 hours of language teaching per week from grade 4 onwards, expenses related to the early language start or to extended language programmes are usually covered by the local self-government and very often by parents.

Similar approaches have been implemented in teacher education through guided work, self-evaluated project work, reflective practice accompanied by self- prepared logs, peer evaluation and self-evaluation encouraged in many colleges.

The impact of the Common European Framework can, therefore, be seen in the role of learning to learn and in the significance attached to formative evaluation.

6. The role of the Common European Framework in the assessment of language skills

The Common European Framework has also influenced the process of designing language examinations. Under this influence examinations now no longer take a paper and pencil form only, but tend to include an oral part which tests interactive and mediating skills and a written part including skills based on real life listening and reading comprehension as well as writing tasks.

Among the new examinations a final exam at the end of secondary education in Poland is the most important. The decision to introduce the so-called New Matura resulted from general dissatisfaction with the system, its subjectivity, lack of comparability and the doubling of effort as final examinations are immediately followed by entrance examinations held at universities.

The present work on school-leaving examinations is regulated by the Act issued on April 16th, 1999 by the Minister of Education. The act states that:

- the examination is to be conducted in two parts: the oral school-based part and the written external part
- the examination is to be taken at the basic or at the advanced level to be chosen by the student one year before the examination,
- students can be exempted from the exam on the basis of recognised language certificates.

The school-based part of the basic level exam is to last 15-20 minutes and is planned to include two tasks. Task 1 will contain three guided role-plays based on natural communicative situations. Task 2 will contain a conversation on the basis of elicitation materials in the form of illustrations, excerpts from a larger text, cartoons, diagrams, tables or newspaper headlines with instructions for the examinee.

The school-based part of the advanced level exam will include one more task in the form of the examinee's presentation of a topic selected earlier from a catalogue approved by the regional examination committee as well as a conversation based on the presentation. This part of the examination adds an additional 15 minutes to the former part and is common to both levels.

The external examination at the basic level will last 2 hrs 10 min and include three parts: listening comprehension (20 min), reading comprehension (50 min) and writing (60 min). The advanced level will add three more parts of 2 hrs 50 min– another listening comprehension task (20 min), another reading comprehension task with a lexico-grammatical test (60 min) and another writing task (90 min).

In the near future the results of the New Matura are going to be presented in percentage form only as information which is parallel to the school-based assessment in the form of a grade.

New forms of assessment influenced by CEF have also entered adult education. Employers see the need for assessing the linguistic competence of their employees in particular language skills in order to make better use of their possibilities in the workplace. The Association of Polish Banks was the first important organisation which decided to introduce the idea of levels and skills to certify the linguistic competence of their employees.

The Common European Framework approach functioned as a model for the work on the problem. Since the Polish context of adult learning calls for slightly different solutions - two additional levels of very high professionally oriented competence have been introduced and the scale now embraces eight levels, i.e. A through H. The six existing levels have been grouped into three main levels of Basic, Intermediate and Advanced User while the new advanced levels formed a new Highly Advanced User's group. The scale is user-oriented as it is based on what the learner can do in each of the four basic skills, though some descriptors

related to the quality of performance were also added. Intuitive methods were used by a committee with a large number of consultants to arrive at the final scale through formulation of key concepts while identifying and defining levels. To refine the scale, however, qualitative methods based on performance analysis were also employed at a later stage.

The modular approach has also been introduced in developing the learners' ability to function in a particular post and a given professional situation. Ten different professional modules are offered which can be selected and \ or combined. Learners then take a series of modular skills-, situation-, or theme-oriented courses thus collecting credit points and certificates. The first examination held by the Association of Polish Banks took place in June 1999.

7. Promoting the Common European Framework. Financing promotion activities

The Common European Framework as a document is not easily available in Poland, though it can be found in the Documentation Centre of the Council of Europe in Warsaw, at the Ministry of Education and in several academic libraries. The approach adopted in CEF is promoted by multipliers and disseminators who participated in new-style seminars of the Council of Europe and in ECML workshops in Graz. Dissemination takes place at teacher training sessions and although teacher trainers refer to CEF as a source of their ideas, not many trainees read the original document. They are more interested in curricula, syllabuses, textbooks and supplementary materials, especially those which, according to their trainers, follow the recommendations of the Council of Europe.

The promotion of CEF takes place predominantly through **pre-service teacher education** of various forms, i.e.

- at methodology courses on syllabus design, individual differences in language teaching, methods and techniques of language teaching etc. at state universities offering 3-year philological programmes leading to a BA, 2-year post-BA or 5-year programmes leading to a philological MA with teaching qualifications,
- at university annual methodological conferences for academic and college staff which are usually financed by universities,
- at methodology courses offered by 60 teacher training colleges of English, and those colleges of French, German, Italian and Spanish where methodology teachers function as multipliers and disseminators for the region (these are usually financed by the local educational authorities).

The promotion of new approaches takes place through other channels also, e.g. through the **in-service teacher training system**. There are 14 regional Insett centres sponsored by the Ministry of Education all of which organise workshops ranging from single sessions to 6 week residential courses for practising teachers.

Regional centres often collaborate with local teacher training colleges using the same premises and the same equipment. New ideas are then promoted

- at in-service teacher training „methodology update" courses organised by colleges for local teachers
- at residential courses for unqualified teachers organised by regional in-service teacher training centres and

- at afternoon or weekend refresher courses for qualified teachers organised by regional in-service teacher training courses.

New ideas are also promoted by social and cultural institutes such as the British Council, Goethe Institute, Alliance Française and the Cervantes Institute at their FLT methodological conferences and through publications resulting from these conferences.

Powerful, though indirect, promotion comes from publishers who use CEF ideas in materials design and distribute their publications at methodological workshops for teachers.

Thus, though no special financing has been given to the promotion of the Common European Framework or the European Language Portfolio - universities' conference budget, local education authorities' budget, INSETT budget and the budget of social and cultural institutions are all involved in the promotion of CEF documents which, in consequence, are seen to have a strong impact on the Polish education.

8. Life with CEF - Conclusions

As has been demonstrated above, many ideas presented in CEF have been successfully promoted in Poland and many areas of language education have benefited from the Framework. The most important of these ideas are related to the modular approach, options for curriculum design, possible curriculum scenarios, clarification of objectives, general and specific competences, emphasis being given to the learning processes and to the role of affective variables.

The relatively successful dissemination of those ideas among teachers, however, does not mean that teachers are familiar with CEF as a publication. Though the approach adopted in CEF becomes part of the Polish teachers' professional knowledge and practical classroom skills, the document itself is not really widely known. So far, for reasons of availability, only top teacher trainers, academic teachers and syllabus designers had a chance to get well acquainted with CEF documentation, usually at Council of Europe conferences and workshops. Promotion and dissemination of CEF in Poland has, so far, meant in fact the promotion of those ideas presented in CEF which were less popular in Poland and deemed worth promoting by the experts. In this sense CEF documentation had been mentally processed by them before it reached the Polish teachers.

Lack of familiarity with the CEF document itself has both negative and positive aspects.

The negative aspect consists in the fact that limited access to documentation might open the way for manipulation, and although no signs of it can be seen at present, a society with a history of communist information management is always sensitive when it comes to strategies based on a document which is not commonly known.

The positive aspect relates to dangers avoided. CEF offers no clear-cut recommendations successfully trying to be non-evaluative and tolerant of any educational solution possible. This is not exactly a virtue in school systems where more than half of the teachers are still unqualified and do not always make efforts to develop professionally. With CEF in hand they would feel perfectly excused as all the teaching procedures are treated equally in the document.

Access to CEF documentation is going to be facilitated in the near future as the National Committee for the European Year of Languages plans to have the document translated into Polish. The effects are awaited with increasing interest.

The CEF documentation is lengthy, abstract, and not really reader-friendly with its multiple lists within each category. This has been noticed by the author of the present chapter who – for the last four semesters – introduced CEF documentation on MA courses for practising teachers at two universities: in Warsaw and in Bialystok. Tired teachers, working at several jobs, and looking for ready-made solutions find it difficult to go through the text and even tend to dismiss it as useless in their practical work. CEF is, therefore, likely to pass unnoticed without the experts' preliminary work on the selection of valuable ideas to be promoted in a given national context.

In this sense, it was beneficial that CEF reached broader circles of teachers through teacher training sessions and through syllabuses, textbooks and supplementary materials based on new approaches.

That is why countries which would like to use CEF – as has been shown by the Polish experience – might find it useful to divide work into three stages

- the expert stage at which CEF is analysed against the background of the national context and its needs and
- the materials design stage at which curricula, syllabuses, textbooks, supplementary materials and teaching aids are prepared and
- the teacher-training stage at which new methodological approaches are promoted and ways of implementing ideas in the language classroom are shown.

CHAPTER 2: THE COMMON EUROPEAN FRAMEWORK IN CATALONIA

Neus Figueras
Departament d'Ensenyament, Generalitat de Catalunya

Joan Melcion
Universitat Autònoma de Barcelona

Introduction

The use of the Common European Framework of Reference (CEF) in Catalonia is related to the use that its regional government has made of all Council of Europe documents since it received full responsibilities in Education from the Spanish government in 1981. The fact that Catalonia has two official languages, Catalan and Spanish, and also the fact that it has a General Directorate for Language Policy in charge of following up the process of normalization of Catalan since 1980, has contributed to a special awareness of issues related to language learning and a strong interest in any developments concerning the learning and teaching of languages.

In Catalonia, the Common European Framework has been received as the culmination of a process, as part of an ensemble of other very useful documents such as the Waystage, the Threshold and the Vantage levels, rather than as a new product likely to trigger off brand-new approaches. All Council of Europe documents have been used both as a point of reference and as a reflection tool, informing the teaching, the learning and the assessment of first, second, and foreign languages.

This chapter reviews briefly how the Common European Framework has been used in different institutional contexts, governmental (the Departament[1] d'Ensenyament, the Departament de Cultura) and academic (the Universitat Autònoma de Catalunya) and for which purposes, and makes proposals for a way forward in the achievement of real transparency in foreign language certification in Europe.

1. A point of reference

The Common European Framework of reference has been used as a point of reference both in curriculum design and in assessment.

It is perhaps the Escoles Oficials d'Idiomes, small in number (eleven in September 2000) but strategically distributed geographically, and with a considerable number of students (over 30,000 in the academic year 1999-2000) that have used the Framework most. The Escoles Oficials d'Idiomes are state-funded centres belonging to the Departament d'Ensenyament teaching foreign languages to adults. The most widely taught languages are English, French, German and Italian, but the Escoles also teach Spanish languages like Basque, Catalan and Castillian and other foreign languages like Arabic, Chinese, Dutch, Greek, Japanese,

[1] "Departament" is the term used in the Catalan Government to refer to the different ministries. The different departments – the ones quoted are "Ensenyament"/Education and "Cultura"/Culture - are granted full executive power from the President and the Parliament.

Portuguese, and Russian. The development of their new curriculum was half-way through when the Framework was published, so it was possible to use it for validation and to check for possible gaps in aspects of language use. Chapter 3, Common reference levels, contributed to a clearer definition of levels in terms of what learners can actually do with the language . Chapter 5, The competences of the user/learner, Chapter 6, Language learning and teaching, and Chapter 7, The role of tasks in language learning and teaching, were helpful in that they backed the frame initially envisaged for the curriculum, which was based on a consideration of language as an instrument of communication . It was particularly helpful to contrast the approach adopted in the curriculum.

"a communicative approach will need to be adopted, considering the learner as a protagonist of his/her own learning and following the learning process of each different learner, paying special attention to cognitive and metacognitive factors and to the strategies which facilitate and favour learning and communication "(Decret 312/1997, 437)

with the "action-oriented" approach of the CEF, which views

"users and learners of a language primarily as "social agents" i.e., members of a society who have tasks (not exclusively language-related) to accomplish in a given set of circumstances, in a specific environment and within a particular field of action" (p.19)

Chapter 4, Language use and the language learner, was particularly useful to contrast language teaching specifications, not only because it provided a very useful breakdown of language use but also because it related this very closely to the language learner. The fact that a document like the CEF dealt with the context of language use and with communicative language processes, including the competences that learners have acquired in previous experiences, validated the structure of a curriculum which attached a lot of importance to strategies and to the ability to learn a new language. This was a very important step, as there is a widespread tradition (even in the era of communicative approaches) to focus on form and, in the case of teaching foreign languages to adults, to pretend they start learning a foreign language as if they were a blank slate. The description of a learner's characteristics and the various uses of language contexts were extremely relevant in that it reinforced the idea that learner's interests are varied and, as such, have to be incorporated into language programmes. Chapter 8, Linguistic diversification and the curriculum, was particularly useful in the context of Catalonia, where two official languages are used interchangeably, as it evidenced the wealth of possibilities open to learners with experience of linguistic and cultural plurality and provided directions on how to build on it. Chapter 9, Assessment, was used to develop the manuals for item writers and examiners of the certificate exams which corresponded to the curriculum, and the marking schemes developed for writing and speaking tests drew on the very helpful guidelines and scale development methodologies in Appendix A, Developing Proficiency Descriptors.

The authors of the Catalan Threshold Level (1999) also used the Framework extensively as a point of reference, although the General Directorate for Language Policy had already participated in a pilot experiment for the Council of Europe Modern Languages Project back in 1982 and had already used the first (1975) and second (1990) versions of the Threshold. The methodological approach adopted in the Catalan Threshold Level was action-oriented and the description of communicative competence was made in terms of the factors which determine the interaction. The Framework also informed the strong socio-cultural component of the document in Catalan.

The Servei d'Idiomes Moderns-UAB Idiomes at the Universitat Autònoma de Barcelona, though still developing its own curriculum, has nevertheless has been using the Framework in the development of its testing specifications, and has focused mainly in the Illustrative Scales of Descriptors for Communicative Activities interspersed in Chapter 4, Language use and the language user/learner, and Chapter5, The competences of the user/learner. These testing specifications have been turned into the working document currently used by all the Universities in the district for the development of common exam certificates. The Serve d'Idiomes Moderns - UAB Idiomes is already making an overt use of the levels as defined on page 32 in the Framework, and the publicity brochures that it distributes show clear parallelisms between their levels and the Council of Europe levels. And this is strikingly different from the other foreign language centres, which forget the European dimension and link themselves to the levels and certificates of the different examination boards of different countries.

The Framework has also proved very useful for assessment purposes, as it has provided a long- awaited common basis for the interpretation of language certificates. The citizens of Catalonia, like those in many other European countries, can choose from different certificates when in need of accreditation in foreign languages. And variation is not limited to the already considerable number of examination boards represented (ICC, UCLES Examinations, Trinity College, Chambers of Commerce, Goethe Institute, Alliance Française....), as both the government and the universities also have their own system of accreditation in foreign language competence as well. Since being able to prove foreign language ability has become more important for job finding, the demand for foreign language certification has increased, and the urge to clarify what each certificate means has become a priority. The Departament de Cultura (responsible for certificates of Catalan as a Second or Foreign Language and a member of ALTE since its foundation) was the first to issue legal documentation which granted equivalences between the different certificates of Catalan provided certain conditions were met. The Departament d'Ensenyament followed in 1993 by creating a system which guaranteed that the certificates in any of the Escoles Oficials d'Idiomes in Catalonia for any of the 14 languages taught in them were equivalent, and has since then been providing the actual exams in the case of the most widely learnt languages (Catalan, English, French, German, Italian, Spanish and Russian). The biggest universities in the Catalan district are about to unite to offer common exams and therefore homogeneous certification for the languages that can be learnt in their centres for modern languages. The Departament de Cultura, the Departament d'Ensenyament and the Universities have thus all used the publications from the Council of Europe as an independent, European and stand alone point of reference, and also as an example of how to put into practice the latest research-based theories.

2. A reflection tool

Although very slowly, and mainly thanks to the fact that those responsible for writing up various documents related to language teaching and language testing have read, studied and used the Framework, the majority of teachers and test administrators are beginning to go to the document and use it as a reflection tool. Sections of Draft Two of the document, especially what in the 2001 version are Chapter 9, Assessment, and Appendix A: Developing proficiency descriptors, have been used in examiners' training sessions, and will be used even more in the future, especially as the new final edition of the Framework is - as promised – somewhat more attractive and user friendly. One of the advantages of the Framework, which will ensure that it is widely used in teacher training seminars, is that it actually contains the state of the art in language learning and teaching, as can be seen in the quotes relating

different projects in different parts of the world and in the bibliography, the inclusion of which is a very welcome improvement to Draft Two.

What has been achieved in the small period of time that the Framework has been around - and specially considering the fact that in countries with political decentralisation like Spain the thorough distribution of Council of Europe documents is often very poor - is that it is now possible to say, for instance, that the teaching and testing of languages as diverse as Russian, Italian, English, Arabic or Catalan, amongst others including, of course, English, in all the Escoles Oficials d'Idiomes de Catalonia and at the Servei d'Idiomes at the Universitat Autònoma de Barcelona has the same objectives and homogeneous points of reference, and that these relate to the Common European Framework.

3. Future developments

Having proved very useful to help homogeneous and transparent description for the different organisations specialised in the teaching of modern languages to adults, the Common European Framework is slowly beginning to be taken into consideration in primary and secondary education. As a partner of the European project Speakeasy (57882-CP-1-UK-Lingua-LD), which has as its main aim to provide practical materials for learners, teachers and assessors in five EU states which exemplify good practice in the field of Modern Languages oral teaching and assessment, the Departament d'Ensenyament is developing and piloting oral assessment materials in secondary schools using the level descriptors in the Framework as described in Chapter 3 and activities which draw on the Communicative Activities list in Appendix B . There are also plans to pilot the use of the European Language Portfolio (CC-LANG (96) 30) in all levels of education as from September 2000.

One of the features of the Common European Framework which has not been as yet subject to enough attention in Catalonia, though, is the wealth of characterisation related to learning to learn and self-assessment. It is hoped that future users of the self-access centres across Catalonia (now more than 50 for Catalan alone and 15 for a variety of foreign languages) will be able to assess their ability in relation to the reference levels presented in Chapter 3, which were presented at the opening speech of the 1999 annual meeting of the Departament de Cultura self-access centres and which were very well received. The focus of this presentation was on the way learners could profile their main language skills and self-assess their level of proficiency using the Common European Framework's self-assessment orientation tool (page 32, Table 2 in Chapter 3) and also how this was currently being implemented by the European project DIALANG, described elsewhere in this book. Special attention, as well as funding, should be given to the implementation of the various courses of action related to self-directed learning and self-assessment suggested in the Framework, as the cost of reorganising the present set-up of materials in self-access centres already in operation may hinder the development of the proposal for the reorganization of the materials according to "Can do" statements, otherwise considered very useful.

4. How far can we go? And who's in charge?

The various uses of the Common European Framework described in the chapters in this book should encourage the Council of Europe to continue working along the same lines, once a common ground seems well established and accepted. The Framework does fulfil its objectives, as agreed in the Rüschlikon Symposium, in that it is comprehensive, transparent and coherent, and in that it is also multi-purpose, flexible, open, dynamic, and non-dogmatic,

although one could not say it is user-friendly, not even in its last and much improved edition. The Common European Framework is not yet a perfect document, if that can ever be possible.

The challenge now is to check that all the energy and willpower mobilised in the interpretation of the document by diverse institutions in the different countries truly results in the mutual recognition of language qualifications across institutions and across countries. The European Language Portfolio currently being introduced as a document with international currency which documents the learners' progress towards plurilingual competence per se is not a solution. So far it is only a shell which different institutions and different countries may or may not adopt to document the learners' foreign language abilities, and the ones in circulation show how diverse interpretations lead to greatly differing results (compare, for example, the Swiss proposal for a portfolio with the French proposal or the EAQUALS proposal). If portfolios have to be developed following "Can do" statements, or at least be somewhat aligned to the Framework scaling system, curricula need to change and exams need to clarify their objectives in relation to the Framework levels in a transparent way, even changing them if necessary. And this will affect two very large groups, that of students in the compulsory school system and that of students who study foreign languages outside the school system. It will definitely be a change for the better, but it will surely have non educational implications which will affect small companies involved in foreign language learning, publishers, etc., which have not been evaluated so far.

However, this is just the beginning of the problem. Consider the possibility that in the near future all school leavers in Catalonia, or in Italy, or in France, have a portfolio which includes a description of their achievements in English (and possibly in a second foreign language) and a stamp by the Government that they are at level B1 in two languages, English and German, for example. Will that be enough for one of those boys or girls if they want to go and study in Germany? Similar cases can be discussed with reference to university students; a student of medicine in Barcelona goes to English classes and gets to level C1. Both the teaching and the exams at the university language centre he or she has been studying conform to the Framework guidelines and he or she gets a certificate from the university saying so. Will that be enough to attend university courses in Britain? There are thousands of different cases with the same final question that can be discussed, and the answer should be that any certificate granted by an institution, be it governmental or not, which has proved to be valid and reliable and conform to certain pre-established standards should automatically be recognised. What does "automatically recognised" mean, however? Does it mean that people in admissions offices at universities and personnel managers in companies will have a list of the different possible stamps and certificates? Surely not. Quick action is needed to develop an easily recognisable, truly European way of making what is accountable really accountable, no matter where it is from. This is as far as transparent accreditation in foreign languages should go.

But this poses still another question. Who is in charge? Who says this certificate can be recognised, and this one cannot? Who sets the rules and the standards? Who checks whether they are followed properly? These questions need to be answered soon but it is not clear by whom. Perhaps the next step to be taken by the Council of Europe should be to clarify whose competence it is to establish the usability and currency of the many certificates around Europe (and around the world!) so that teachers, course designers, examining bodies and educational administrators but also, and especially, learners across Europe can see that the publication of the Common European Framework of reference really has made a difference.

References:

Mas, M. and Melcion, J. (1999). Nivell Llindar per a la llengua Catalana. Direcció General de Política Lingüística. Departament de Cultura, Generalitat de Catalunya

Departament d'Ensenyament - Decret 312/1997, de 9 de desembre pel qual s'estableix l'ordenació curricular del primer nivell dels ensenyaments d'idiomes (DOGC núm 2588, de 27.2.1998).

Contact address details:

Neus Figueras
Departament d'Ensenyament
Via Augusta, 202
08021 Barcelona-Spain
Tel: 93-400 69 16
Fax: 93-400 69 80
e-mail: nfigueras@pie.xtec.es

CHAPTER 3: USING THE COMMON EUROPEAN FRAMEWORK OF REFERENCE IN COURSE AND MATERIALS DEVELOPMENT: THE UK OPEN UNIVERSITY EXPERIENCE

Cecilia Garrido and Tita Beaven
The Open University

Introduction

This case study is based on the development of the Spanish courses at the British Open University. This higher education institution was founded in 1969 and for over 30 years has offered a wide variety of undergraduate and postgraduate courses to adult students of all ages and backgrounds, completely at a distance. For a long time it was believed that it would be difficult to teach languages successfully away from the classroom and it was not until 1991 that the Centre for Modern Languages (CML) at the Open University was born. The Centre, now called Department of Languages, has Diploma courses in French, German and Spanish. The Diploma qualification also counts towards BAs in Humanities, European Studies and Language Studies.

The Spanish Diploma is the third in the Department's course portfolio and the development of its courses started in 1996. The first course, *En rumbo*, was launched in 1999. The second, *Viento en popa*, in 2000 and the third, *A buen puerto*, in February 2001. The work carried out by the Language Policy Division of the Council of Europe has been instrumental in the development of the courses in question, and this case study will illustrate in further detail various aspects and stages in the process.

Although the references made to the Common European Framework of reference in this case study relate specifically to the Open University Spanish courses, it is hoped that colleagues in other learning environments may be able to relate their experience developing language courses to our own, and may also find our experience useful.

1. The challenge

The Spanish team had a brief to develop a series of courses, geared to adult students with some knowledge of the language, and who hopefully would be looking to learn Spanish to enhance their intellectual profile or to acquire a qualification. However, the philosophy of the Open University is to be open to all and therefore students do not need to have formal qualifications to join its courses. Students' backgrounds differ widely not only in relation to their life experience, but also in their prior knowledge of the subjects they want to study.

The team therefore needed to design a suite of courses that on the one hand would appeal to a wide range of individuals, and on the other, would meet the University's requirements to award a qualification of a level comparable to that of "conventional" Higher Education institutions.

It was necessary to make a detailed definition of the student profile to be able to address their needs and interests, and to make a clear analysis of the features required in the courses so that the University's requirements could be met. Some assumptions regarding the profile of potential students were made based on the experience of the French and German courses, developed at the Department in previous years. This included assumptions regarding the students' age, social background, language learning experience and level of language competence, although the latter was more difficult to identify accurately.

2. Engaging with the Common European Framework of reference

It was clear to the Spanish development team that, to be able to meet students' needs, the courses needed to be interesting, relevant and provide the support needed by individuals who would be learning at a distance, often in isolation. To be able to meet requirements in terms of set levels of competences to be achieved at particular stages in the process, the courses needed to formulate very clear objectives, a set of outcomes through which the achievement of objectives could be measured, and the means for students to be able to achieve such goals. The curriculum should also provide the opportunity for students to become autonomous learners capable of evaluating their continuous progress.

Before the development of the Spanish courses, the Open University had loosely identified the point of entry and exit of the French and German courses in relation to established British competence levels, namely, the General Certificate of Secondary Education (GCSE) normally achieved by students after a period of study of one or two years, whether at secondary or further education, and the Advanced (A) level of competence, usually achieved by students over a period of two years after achieving GCSE, in the final two years of secondary education. The levels of competence to be achieved by prospective Open University language students were at the time described as follows:

Open University language courses: competence levels as described in 1996

	First course	**Second course**	**Third course**
Entry level	GCSE	End of the first year of an 'A' level course.	'A' level
Exit level	End of the first year of an 'A' level course.	'A' level	Level achieved by a student at the end of the first year of undergraduate language studies

Figure 1

The level descriptors for the third course became not only more difficult to specify but also less accurate, especially due to the fact that it is currently very difficult to compare competence levels across the Higher Education spectrum in the UK, as identified by Professor Jim Coleman in the survey he carried out in 1996: “The survey shows there is a need for universities to state clearly proficiency levels in all course documentation. There are hugely different levels of language performance and provision across the system. Students, parents and employers need to know their linguistic objectives and assessment criteria, and these should be defined in accordance with agreed national and international standards” (Coleman, 1996).

The description and declared levels of competence produced were not only the result of a deficient system to account for student proficiency, but also to the lack of a syllabus that outlined clear objectives at each stage of the learning process. The Spanish team was very aware of the need to design a syllabus where objectives were transparent and where students could be perfectly aware of the route to follow to achieve those goals, and of the means at their disposal to take control of the learning process to optimise results.

Once initial assumptions were made regarding the profile of potential students, the Spanish team carried out some research to find out how relevant the CEF would be in the design of the curriculum of the Spanish courses. Some members of the team were already aware of the work carried out around the definition of the Threshold level in Spanish, *"Nivel Umbral"*, but when consulting it in detail, they were disappointed to find that the specification was out of date and that the update to the English Threshold specification in 1990 never took place in the Spanish version. We then concentrated on using the CEF to inform our decisions regarding the content of the syllabi for our courses and, to a great extent, how these decisions would affect the design of the materials that would support their delivery. We needed an approach that was flexible and adaptable to our context. Before going very far we found that the objectives of the CEF fit very well with our own: "The construction of a comprehensive, transparent and coherent Framework of language learning and teaching does not imply an 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 should be:

- multi-purpose: usable for the full variety of purposes involved in the planning and provision of facilities for language learning
- flexible: adaptable for use in different circumstances
- open: capable of further extension and refinement
- dynamic: in continuous evolution in response to experience in its use
- user-friendly: presented in a form readily understandable and usable by those to whom it is addressed
- non-dogmatic: not irrevocably and exclusively attached to any one number of competing linguistic or educational theories or practices" (CEF, 2001, 1.6)

3. Putting pen to paper

The Spanish course team had extensive teaching experience and it already had ideas as to the type of syllabus that in their view should drive the objectives of the courses. The preference was for a notional-functional syllabus where successful communication is the main objective of language interaction and where linguistic content would be used to underpin such interaction. Wilkins' work (1976) on notional syllabuses and on competence level definitions, presented to the Council for Cultural Cooperation of the Council of Europe (CDCC), where the CEF originated, reflected the Spanish team's initial thoughts. The CEF was the tool that helped us put those thoughts in order, and to focus on various aspects throughout the development of the courses.

It was very useful to be able to refer to the CEF as a tool to develop the "picture" of the big puzzle, but it was also liberating to know that the rules were not fixed and that we had the flexibility of making the "picture" our own. This is one of the main advantages of the CEF. It is there to enable language practitioners but not to prescribe or restrict: "You may of course use the Framework document in any way you wish, as with any other publication. Indeed we

hope that some readers may be stimulated to use it in ways we have not foreseen. [..] One thing should be made clear right away. We have NOT set out to tell practitioners what to do, or how to do it. We are raising questions, not answering them. It is not the function of the Common European Framework to lay down the objectives that users should pursue or the methods they should employ" (CEF, ibid: xi)

Chapter 2, *Approach adopted*, and 4, *Language use and the language user/learner* proved to be the most useful for our purposes. They helped us to start putting the puzzle together. The difference made in Chapter 2 between General and Communicative competences helped us shape the content of our courses in terms of the means we should provide for our students to achieve communicative competence, and perhaps more importantly, in relation to the general competences we would need to foster to enable students to be independent learners, able to learn from each language interaction and develop strategies that would help them deal with the unplanned and the unexpected. Speakers' level of communicative competence (native and non-native) is very often reflected in their success or failure in achieving specific communicative goals. As native speakers we are always updating our communicative knowledge, our strategic repertoire. It is therefore essential that we promote the development of this strategic repertoire in our language students.

As Little says: "When we shift our attention to the language learner/non-native speaker it is necessary to make two points. First, if the learner's purpose is to achieve communicative competence in the target language, that purpose will be fulfilled to the extent that he or she succeeds in mastering (i) forms of routinized communication, (ii) the means to repair breakdown in communication, and (iii) techniques of conscious planning of communication" (Little, 1996: 16)

Chapter 5, *The competences of the user/learner*, highlights the importance of the development of such competences. Some of them are simply acquired, for instance the need to use particular linguistic structures to express wishes, hypotheses, etc. Many others are culture-related and therefore require a frame of mind and a positive attitude on the part of the learner, for instance when internalising differences in behaviour between his/her own culture and that of the culture being learned. This is a dimension that has been very important in the development of the Spanish courses. The CEF has reinforced the Spanish team's firm beliefs that the level of strategic competence a learner of Spanish has is highly dependent on his/her understanding of the various cultures that constitute the Spanish-speaking world today, and that unless students have a positive attitude to developing such understanding they will be unable to be truly competent language users. The figure below shows some examples of how we used the information from Chapters 4 and 5 to define language objectives:

Syllabus objectives in relation to the CEF

The Common European Framework	Objectives in the Spanish syllabus
General competences:	
"Existential competence: e.g. a **willingness to take initiatives or even risks** in face to face communication [...] also listening skills, **attention to what is said, heightened awareness of the risks of cultural misunderstanding in relation with others"** CEF (ibid., 21)	**"When participating in a Spanish debate:** - Identify uses of metaphor, irony and other "violations" of conversational maxims. - Identify implied attitude of the speaker: annoyance, sarcasm, reprimand, etc. - Identify strategies for interrupting." (Spanish Skills syllabus, block 6, *Usted qué opina*)
"…declarative knowledge is understood as **knowledge resulting from social experience** (empirical knowledge) and from more formal learning (academic knowledge)…" CEF (ibid., 20) "Sociocultural knowledge Strictly speaking, **knowledge of the society and culture of the community or communities where the language is spoken**…" CEF (ibid., 93)	**Roots** - The concept of multiculturalism in Spanish speaking societies - The Arab legacy in Andalusia - Music and dance as testimony of multicultural origin (Spanish thematic syllabus, Third course, block 1, *Raíces*)
Communicative language competence:	
"Sociolinguistic competence also includes the ability to recognise the linguistic markers of for example: social class, **regional provenance, national origin, ethnicity**, occupational group" CEF (ibid., 103),	**Identifying the Argentinean accent:** Special features Pronunciation and modulation of the voice - Grammatical structures - Lexicon - Linguistic influences: from the Amerindian cultures from the Canary Islands and Andalusia from Italy (Spanish Cultural syllabus, Third course, block 2, *Con mis propias manos*)
"Linguistic competence: Knowledge of and ability **to use formal resources from which well formed, meaningful messages may be assembled and formulated.**" CEF (ibid., 99)	- **Requesting and expressing opinion:** *¿Qué opina de…?* *Me gustaría saber su opinion sobre…* *A mi me parece que…* - **Expressing agreement or disagreement in various degrees:** *Estoy totalmente de acuerdo…* *Yo en cambio pienso que…* *¡Pero cómo puedes decir eso!* (Spanish Grammatical syllabus, Second course, block 6, *Usted qué opina*)

Figure 2

The sections on language and language user, tasks, activities and processes required to develop communicative competence (Chapters 4 to 7) helped us reflect on:

(a) Issues relating to the development of our course materials:

- Conditions and constraints under which communication occur (CEF, 4.1.3). In our case this applied, for instance, specifically to the development of specifications for the gathering of audio-visual and print resources.

- The user/learner's mental context (CEF 4.1.4). Here we had to take into consideration issues specially related to learners studying on their own, their needs, interests and factors that would help keep a high level of motivation throughout the process.

- Communicative tasks and purposes (CEF 4.3). This came into play when identifying the possible domains in which our students were likely to engage and the type of activities and roles that would help them to be competent in such situations.

(b) The content of the syllabus:

- Communicative themes (CEF 4.2) through which we would build the language interactions.
- Communicative language activities (CEF 4.4), to establish the balance to be achieved between the various language skills.
- Task description and performance (CEF 7.1 and 7.2), to focus our efforts on how to develop strategic competences in the context of the objectives to be pursued, and the outcomes expected from students at each level.

(c) The learning and teaching process our students and we would engage in:

- How do learners learn? (CEF 6.2.2), to help us rationalise the balance between the different components of the syllabus: situational vs. grammatical, vs. functional, vs. task oriented and problem solving. The final objective was to optimise the language experience and address the perceived needs of our students.
- Learning to learn (CEF 5.1.4), particularly relevant in our context, to help students learning at a distance to become autonomous, in control of their learning and able to transfer their skills to enhance their communicative interactions.

Chapter 8, *Linguistic diversification and the curriculum*, was helpful in that it reinforced the sense of flexibility in the setting of objectives, allowing for prioritisation but leaving open a path for us to formulate objectives conducive to pluricultural competence.

The information sifted from the CEF, together with our research into similar language provision at "conventional" universities in the UK, and the assumptions made of the profile of our potential Spanish students, allowed us to make informed decisions regarding the content of the syllabus and the levels of competence to be pursued at each stage in the learning process. We felt confident that we could formulate objectives that were transparent and achievable, and that our students' achievements would make them competent communicators, able to share their culture and understand those where the target language is spoken. Last but not least, our students' level of language competence could also be recognised nationally and internationally, a feature that will become more and more attractive in a Europe promoting increasing mobility and a sense of democratic citizenship.

The series of descriptors that appear in Chapter 3, *Common reference levels*, were extremely useful in helping us with the setting of objectives, description of the relevant domains and tasks, and in the development of activities geared to specific outcomes through which the achievement of goals can be measured.

The Spanish courses have been aligned to the levels of competence specified by the CEF as follows:

Open University Spanish courses: competence levels based on the CEF descriptors

	First course: ***En rumbo***	**Second course:** ***Viento en popa***	**Third course:** ***A buen puerto***
Entry level	**"Breakthrough" Basic User - A1** Can communicate at a simple level using common expressions in familiar situations.	**"Waystage" Basic User- A2**	**"Threshold" Independent User - B1**
Exit Level	**"Waystage" Basic User- A2** Can communicate in situations which require exchange of information about familiar subjects using expressions of immediate relevance.	**"Threshold" Independent User - B1** Can communicate and deal with most familiar situations and can cope with less familiar situations without undue degree of stress.	**"Vantage" Independent User B2** Can communicate with a degree of fluency and spontaneity, contributing own point of view on matters not always very familiar, and produce clear and detailed text in his/her field of specialisation.

Figure 3

4. The development of the audio visual materials

The Open University believes, for reasons both of quality assurance and of equal opportunities, in providing all its students with all the learning materials they will need in order to complete their course, and therefore apart from the set texts used, such as grammar reference books and dictionaries, all course materials at the Department of Languages are prepared by the course team; these include all course books, audio-visual resources, and assessment materials. Language courses at the Open University make heavy use of specially commissioned AV resources; the use of authentic audio and video materials and studio recorded audio activities serve the purpose of immersing the learner in the language in a way that, to some extent, replaces the interaction with the teacher and fellow learners that usually takes place in the conventional face-to-face language learning environment. Moreover, because of Equal Opportunities issues, course designers cannot assume that students have access to any other materials that will help them with their language learning (such as satellite TV, Internet access, etc.). The design of the AV resources is therefore of central importance to the courses. It is fair to say that, in the preparation of the third Spanish course, the CEF was not only useful in the development of the syllabus itself, but also played a major role in the planning and gathering of AV materials.

Before describing how the CEF informed the work on our AV resources, it might be useful to run through the production process of the OU materials, as it is a complex and lengthy process which involves a large team of contributors and usually takes three years to complete.

Once the courses have been produced, they have a life of 6 to 10 years. The production involves the following steps:

(a) production of a preliminary draft syllabus, with functions, topics and linguistic elements to be studied
(b) drawing up of the specifications of the audio-visual materials, prepared by the academic team in close liaison with the Production team at the BBC
(c) gathering of authentic AV resources in the target countries
(d) editing of video and selection of audio sequences.
(e) updating of the syllabus, based on the linguistic exponents present in the AV materials gathered.
(f) development of the course books
    (1) draft writing and review (involving external authors and in-house academic team)
    (2) editorial process (editorial queries to the course team)
    (3) production process (book design, production of art work and printing)

(g) production of the activities CDs (which include extracts of authentic audio and scripted activities recorded in the studio)
(h) in parallel to all the above, development of the assessment strategy and preparation of the assessment materials

As can be seen, the role of AV materials is extremely important. Although the academic team starts from a draft syllabus, its final version depends largely on the language arising from the authentic AV materials. Therefore, the AV specifications that the course team delivers to the BBC need to ensure that the AV materials, whilst being authentic and not scripted in any way, exemplify the broad functional and linguistic content the course requires.

For the production of *A buen puerto*, and indeed all the other language courses, the academic team drew up a draft syllabus which included the topics for each block of study, together with the broad functions to be covered and the academic skills we wanted the learners to develop. As an illustration, the draft syllabus for block 2, *Con mis propias manos*, which deals with the world of art, included the following:

Block 2: *Con mis propias manos*

Topic: Aspects of arts and crafts in Spain and Latin America

Sub-topics:

- Definitions (e.g.: What is art? What is the creative process? What is the difference between art and craft?)
- Art and society (the social function of art, public art, mural art in Latin America)
- Art as a means of protest (protest songs, Picasso's *Guernica* as protest)

Language focus:

- Speaking personally about likes, dislikes, preferences
- Expressing feelings
- Speaking about attitudes and responses to events/experiences
- Comparing and contrasting
- Talking about other people
- Defining

Academic skills:

- Quoting and paraphrasing
- Avoiding plagiarism

The team then analysed the targets for listening comprehension at Vantage level, which are divided into the following categories (CEF 4.4.2.1):

- Overall listening comprehension;
- Understanding conversation between native speakers;
- Listening as a member of a live audience;
- Listening to announcements and instructions;
- Listening to audio and media recordings.

These level descriptors for listening (CEF 4.4.2.1) indicated some of the points we should consider when drawing up the AV specifications. They were related to:

(a) The types of discourse exemplified by the AV resources. We would have to ensure that:

- some of the recording contained "propositionally and linguistically complex speech" and dealt with "both concrete and abstract topics",
- there should be some technical discussions,
- the learners would be exposed to "extended speech and complex lines of argument provided the topic is reasonably familiar, and the direction of the talk is sign-posted by explicit markers".
- the recordings would have to include speakers who demonstrated a variety of moods, tone, viewpoints and attitudes.

(b) The text types. Here the CEF (4.6.3) suggested the following: public announcements and instructions; public speeches, lectures, presentations, sermons; rituals; entertainment (drama, shows, readings, songs); sports commentaries; news broadcasts; public debates and discussions; inter-personal dialogues and conversations; telephone conversations and job interviews. To those text types, we wanted to add specific media genres, such as current affairs programmes, documentaries, live interviews, talk shows, phone-ins, etc.

(c) The recording environment. We would have to ensure that some of the recordings contained "a certain amount of background noise", to imitate as far as possible "real-life" conditions.

As a result of the considerations outlined above, we decided that each of the seven blocks of study would have a specially commissioned 30 minute video which would be related thematically to each block, but that we would not prescribe the linguistic content to the production team. In other words, we would not be concerned about the linguistic structures used by the participants or the functions exemplified. The purpose was to present the learner with a documentary that mirrored as far as possible a real-life documentary, but which would be highly relevant to the student in that it would relate to the cultural content of each block of the course. Similarly, we decided to make a 20 minute "radio programme" for each of the blocks which followed "real radio" conventions and which would enable students to have access to something very much like the real thing. Again, we were not prescriptive in terms of the language used, although we did specify we wanted a variety of genres within the programme format (interviews, current affairs, talk shows, phone-ins, etc.).

Although the descriptors for listening were useful in deciding some of the format of the AV materials, the main input for the AV specifications actually came from elsewhere. Studying the Council of Europe documentation, it became apparent to us that if our students were to reach the goals set in the Vantage level descriptors for speaking skills, they would only do it by first of all being presented with models of spoken language that matched those descriptors; they could then be shown how the discourse is organised, as well as being presented with structures and vocabulary items that would enable them to perform within that type of discourse; finally they would be encouraged to produce the language themselves. It is important to remember as well that, as the OU courses are distance learning courses, the AV material would have to demonstrate many of the features of interactive language which in the normal face-to-face classroom setting can be analysed and practised through actual interaction between students.

The Council of Europe Vantage level (B2) descriptors for speaking skills encompass the following (CEF 4.4.3.1):

Overall spoken interaction:

The learner can "highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments".

Understanding a native speaker interlocutor:

The learner can "understand in detail what is said ... even in a noisy environment".

Conversation

The learner can "convey degrees of emotion and highlight the personal significance of events and experiences".

Informal discussion (with friends)

The learner can "keep up with an animated discussion between native speakers;

express his/her ideas and opinions with precision, and present and respond to complex lines of argument convincingly"; "take an active part in informal discussion, ... commenting, putting point of view clearly, evaluating alternative proposals and making and responding to

hypotheses", and "account for and sustain his/her opinions in discussion by providing relevant explanations, arguments and comments".

Formal discussions and meetings

The learner can "keep up with an animated discussion, identifying accurately arguments supporting and opposing points of view", and can "contribute, account for and sustain his/her opinions, evaluate alternative proposals and make and respond to hypotheses".

Goal-oriented co-operation

The learner can "understand detailed instructions reliably"; "help along the progress of the work by inviting others to join in, say what they think, etc.", and "outline an issue or a problem clearly, speculating about causes or consequences, and weighing advantages and disadvantages of different approaches".

Transactions to obtain goods and services:

The learner can "explain a problem which has arisen and make it clear that the provider of the service/ customer must make a concession".

Information exchange:

The learner can "understand and exchange complex information and advice on the full range of matters related to his/her occupational role"; "pass on detailed information reliably"; "give clear, detailed description on how to carry out a procedure" and "synthesise and report information and arguments from a number of sources".

Interviewing and being interviewed:

The learner can "carry out an effective, fluent interview, departing spontaneously from prepared questions, following up and probing interesting replies" and can "take initiatives in an interview, expand and develop ideas with little help or prodding from an interviewer."

As will have become apparent by simply reading the above speaking descriptors, these were useful in three respects. First of all, they highlighted the types of communicative activities the participants in the audio recording would have to be involved in (such as conveying "degrees of emotion" and highlighting "the personal significance of events and experiences"; accounting for and sustaining "views clearly by providing relevant explanations and arguments"; helping along "the progress of the work by inviting others to join in, say what they think, etc." and so on). Secondly, it provided us with clear guidelines as to the types of oral interactions we would need to include in the materials (conversations, informal discussions with friends, formal discussions and meetings, goal-oriented co-operation, transactions to obtain goods and services, information exchange and interviews). Some of these posed greater challenges than others in terms of the recording of authentic materials, and the course team had to work closely with the audio producer to ensure all types of oral interactions were covered, and that the recordings did not only include those that are more easily set up, such as interviews and conversations. Some thought had to be given to the technical aspects of recording discussions and meetings, for instance, and, more importantly,

to the settings in which these would produce interesting, authentic language that would be useful for our purposes as language teachers.

Finally, the Vantage level descriptors also offered us pointers about the types of tasks that the learners should be engaged in when listening to the materials or doing further work related to the audio. These included tasks such as: "identifying accurately arguments supporting and opposing points of view"; outlining "an issue or a problem clearly, speculating about causes or consequences, and weighing advantages and disadvantages of different approaches"; or, synthesising and reporting "information and arguments from a number of sources", to name but a few.

By way of an example, we can look at some of the original AV specifications, and see how these then developed into the final materials and their teaching outcomes. Figure 4 illustrates the process by which the original ideas about AV materials were produced for some of the audio materials for block 2, *Con mis propias manos*, the syllabus of which has already been discussed.

Illustration of 5 independent examples of AV specifications and their realisation in the course materials

First draft of AV. specs to be refined after discussions with BBC and possible participants are identified			Final audio material and its use in teaching/learning	
Topic	**Participants**	**Type of interaction and communicative activity**	**Final audio**	**Teaching/learning outcome and activity/task**
Defining art What is art?	individual artists	**Conversation** that convey degrees of emotion and highlight the personal significance of events and experiences	two artists explain to the interviewer what art is for them personally and professionally	**Defining abstract terms** Learners define several abstract terms (music, painting, etc.)
Defining art How do you feel when you finish a work of art?	individual artists	**Conversation** that convey degrees of emotion and highlight the personal significance of events and experiences	two artists explain to the interviewer how they feel when they finish a work of art, both professionally and personally	**Expressing positive and negative feelings** Scripted speaking activity: learners have to several situations by saying how they feel
Difference between art and crafts	members of the public	**Conversation** in which the speaker accounts for and sustains views clearly by providing relevant explanations and arguments	Several vox pops in which members of the public are asked by interviewer to explain the difference between arts and crafts	**Expressing contrast**. Revision and practice of expressions and structures to denote contrast
Difference between arts and crafts	Group of art students	**Informal discussion:** animated discussion in which ideas/ opinions are expressed clearly, fluently and effectively and in which a point of view is put forward clearly by providing relevant explanations, arguments and comments	Discussion in which the students on a training course in traditional crafts have a heated discussion about the difference between arts and crafts - there are clearly opposing points of view about the similarity/differences between arts and crafts	**Work on learner training** Strategies for listening to an informal group discussion **Speaking** Scripted activity: student takes part in a discussion with two other participants and has to maintain his/her viewpoint regarding art and crafts following prompts
Traditions in public art, murals in Latin America	Group of artists discussing a project	**Goal-oriented cooperation** (discussing a document, organising an event...) Examples of how to help along the progress of the work by inviting others to join in, say what they think, etc.	Discussion between art teacher and her students on a mural art project they were about to embark on, which includes presentation of project, questions, brainstorming as to how to proceed, reaching an agreement and recapping.	**Listening and Speaking** Learners fill in a worksheet summarising discussion as they listen, then summarise orally the execution of the project **Language work around a project** discussing it, proposing a plan of action (e.g.: how to ask for clarification, ask for further information, explain the phases of the project, make suggestions, invite other people to express their point of view, resolve problems cooperatively, etc.)_ **Task**: students have to propose a plan of action for another project

Figure 4

As can be seen in Figure 4, block 2 consists of a number of sub-themes, and each sub-theme has one or several audio extracts. The aim was to present a variety of types of interaction and communicative activities. Some of these would then be used for providing examples of specific linguistic exponents, some would be exploited as part of a specific task, and others might also provide material for developing specific listening or speaking strategies.

As is usual in the production of any AV recording, the specifications indicated the topic to be covered and nature of participants, but our work with the CEF also enabled us to identify the type of interaction and communicative activity we wanted to illustrate, thus giving a much greater focus to the specifications, and resulting in audio material of high quality. We firmly believe that the quality of the materials stems from the fact that it offers a variety of types of interaction and communicative activities which are not usually present in language teaching materials.

5. Conclusion

To sum up, we hope to have demonstrated how, when producing a course, it is not only important to think carefully about the syllabus, but also to ensure the AV material offers the learner variety not only in the topics it deals with and in the linguistic functions it aims to exemplify, but also in the breadth of types of interaction it covers. The use of authentic audio in situations which are often under-represented in AV teaching materials, such as group discussions, can also provide the learner with great insights into how various conversation strategies occur in natural language, and provide the teacher with invaluable opportunities to work on the development of listening strategies that the learner will find necessary in the "real world".

The CEF is however very useful beyond the development of materials. Language practitioners will find it very valuable to:

- help them justify their own approach to language teaching;
- consider the main questions that will help them to define language course objectives and make decisions regarding transparent levels of language competence to be pursued;
- determine how to achieve those objectives taking into account the various types of competences required to develop students into autonomous learners capable of interacting effectively with the foreign culture, and in whichever role they are likely to perform;
- help them to identify the range of authentic materials (audio-visual or printed) they want to use in their own teaching, and decide the purpose for which they will exploit those materials;
- analyse the purpose of their assessment strategy and make decisions on how to implement it via formative and summative means.

The nature of the CEF is such that all language practitioners, while having a framework within which they can work, can feel free to make decisions that are relevant to their own context and to the needs of their students. Further more, language teachers can feel confident of being able to offer learners language learning opportunities that address their needs, and that are transparent in terms of what is expected of them to achieve set objectives, and the outcomes that will reflect such achievements.

References

Coleman, J. (1996). *Studying Languages: A Survey of British and European Students*. CILT

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge University Press.

Council of Europe (1998). *Modern Languages: Learning, Teaching, Assessment. A Common European Framework of Reference*.

Little, D. (1996). *Strategic competence considered in relation to strategic control of the language process in Strategies in language learning and use*. Council of Europe.

Wilkins, D.A., (1976). *National Syllabuses*, Oxford University Press.

CHAPTER 4: HOW TO PROMOTE LEARNING TO LEARN IN FIRST FOREIGN LANGUAGE CLASSES IN FINLAND

Hanna Jaakkola, Department of Teacher Education, University of Helsinki
Ursula Viita-Leskelä, Second Teacher Training School, University of Helsinki
Sirkka-Liisa Sävy, Simonkylä Lower Secondary School
Kaija Komsi, Simonkylä Upper Secondary School

Introduction

The purpose of the project reported in this chapter was to explore the practicability of the Common European Framework (CEF) in Finnish schools. The research and development work focused on those aspects of learning to learn described in the CEF. The aim of this work was to promote learning skills in first foreign language classes systematically from the primary level up to the upper secondary level. Eleven teachers, each from a different school, were involved in the project, which they carried out as classroom action research in their regular classes during the school year of 1998 – 1999.

The research questions dealt with learners' existential competence, their study skills, heuristic skills and self- and peer-assessment. The results showed that the learners profited from the explicit development work although the time available was too short to bring about permanent changes in their behaviour. The teachers felt that the CEF could support their work and found the experience of action research valuable for their professional growth.

1. Learning to learn

Learning to learn was chosen as the focus of the Finnish pilot project reported in this chapter because it has become an important goal of education in today's rapidly changing world. In addition to the practical needs of lifelong learning, the importance of this goal can be justified by present theories of learning, which consider learners to be active agents of their own learning (e.g. Kohonen 1992; Rauste-von Wright & von Wright 1994; Williams & Burden 1997). The purpose of the project was to find out in what ways the features of learning to learn included in the Common European Framework (CEF) can be applied to classroom work in first foreign language classes in Finnish schools starting at the primary level and continuing at the lower and upper secondary levels.

Some key concepts

Learning to learn can be seen as strategically efficient, intentional action directed to learning, and intentional learning is a skill that can be learned. However, learning skills do not automatically transfer to new situations, but they can transfer if attention is paid to the context of their use in the learner’s own life and activities (von Wright 1992; Niemivirta 1997). Learning to learn is naturally not restricted to language learning but is part of the learner's general competences (CEF 2.1.1; 5.1). Learning to learn languages is generally considered to include metacognitive knowledge and strategies, language learning strategies and cognitive styles.

Metacognitive knowledge is knowledge on two levels: we know something and we know that we know. Metacognitive knowledge helps learners to reflect on their inner action and learn to regulate it. It also includes the awareness of one's own personality, feelings, motivation, attitudes and cognitive style (cf. existential competence, CEF 5.1.3). Metacognitive strategies, on the other hand, are general skills that enable learners to utilize metacognitive knowledge: to plan, monitor and assess their learning (Rauste-von Wright & von Wright 1994; Williams & Burden 1997).

The deployment of metacognitive knowledge and strategies is referred to as **self-direction** in FL learning. Research evidence suggests that metacognitive knowledge is a prerequisite for self-directed language learning. Explicit metacognitive knowledge about task demands and appropriate strategies for task solution may even be the most important factor in determining the effectiveness of a person's language learning attempts (Chamot & O'Malley 1994; Wenden 1998). Although the CEF does not often use the terms metacognitive knowledge and strategies, it includes a great deal of subject matter in this area (e.g. 6.3.5; 6.4.6.5; 6.4.2.3; 9.3.13).

Learner reflection is the main tool in the formation of metacognition, i.e. metacognitive knowledge and strategies. The level of reflection indicates the quality of the learner's metacognitive knowledge. Huttunen (1996) discerns three distinct levels of learner reflection:

mechanical level: little or no reflection, enumeration of activities done
pragmatic level: reflection mirroring the practical demands of a task
emancipatory level: reflection extending beyond the prevailing situation, evaluation of teaching, one's own work and progress in learning. This level is typical of an autonomous, self-directed learner.

There are many, even conflicting, views on **language learning strategies** (for a more detailed account, see Cohen 1998). In the CEF learning strategies are actually not distinguished from communication strategies (2.1.5; 4.4; 7.2.3). Since many teachers in the project group were familiar with Oxford's (1990) classification of language learning strategies, her work was used as a starting point.

Learning strategies differ from **cognitive** or **learning styles** (see e.g. Skehan 1998) in the sense that strategies may be learned and modified but cognitive style appears to be part of an individual's identity. Cognitive style may be defined as "an individual's preferred and habitual approach to organising and representing information". Despite numerous different labels, cognitive styles seem to consist of two basic dimensions: wholist-analytic and verbal-imagery (Riding & Rayner 1998: 8-9). The CEF also presents cognitive styles as part of a learner's identity. It suggests that learners may improve their study skills by developing their own learning strategies corresponding to their cognitive styles (5.1.3; 6.4.6.5).

The CEF does not deal with cognitive styles in detail. In order to identify students' preferred learning styles, especially their favoured perceptual mode, the teachers of the lower and upper secondary levels used part of the Learning Styles Inventory by Dunn and Dunn (Dunn, R. et al. 1992), which some of them already knew. The basic assumption behind the Inventory is that learning achievements improve if learner preferences and the teaching are compatible.

2. Method

The pilot project was carried out as classroom action research with eleven teachers, each from a different school: four primary, four lower secondary and three upper secondary schools. The schools are situated in the Greater Helsinki area, apart from one primary, one lower secondary and one upper secondary school in the provincial town of Kouvola. The teachers were invited to join the project by the coordinator Hanna Jaakkola. They were known to be active in the teaching profession; many of them were also textbook and materials writers and / or pre- and in-service teacher trainers. The first foreign language taught was English, except in one primary and lower secondary school where it was German. The research and development work focused on first-year learners of each school level.

2.1 Action research

The case studies were conducted by the teachers in their regular classes. Action research has proved to be successful when undertaken by teachers in their own classrooms to improve their professional practice (Ellis 1997). Action research typically proceeds in cycles of activities: planning, action, observation, reflection, revision of plan, renewed action, observation, reflection and so on (see Carr & Kemmis 1986; Elliot 1991). Teachers as action researchers may use a variety of research techniques, both quantitative and qualitative. Since the generalisability of research findings is not of primary concern, there is normally little use of statistical techniques. A qualitative, interpretive approach is considered to be more appropriate (see Cohen & Manion 1994; Wallace 1998).

For data collection the teachers in the project group used questionnaires, logs, diaries, interviews and informal observation techniques. They analysed their data quantitatively, counting frequencies and percentages and / or interpreted it qualitatively.

2.2 Implementation of the project

There were five cycles of action research in the five periods of the school year of 1998 – 1999, each lasting six or seven weeks. The last cycle was used for data analysis and writing the report. The project work mainly consisted of contact sessions of the whole group with the coordinator, meetings of the subgroups of teachers teaching at the same school level and fieldwork periods. Various problematic issues were frequently discussed by phone, fax and e-mail.

Most contact sessions included a presentation by an expert on some features of learning to learn or on doing action research. Reading materials were also discussed. One two-day workshop was arranged. The coordinator gave feedback on the teachers' papers by e-mail, fax and phone. After having reflected on their previous cycle of research and development work and discussed it during the contact sessions, the teachers made a new or more focused action plan for each new school period. The refining of the research questions continued throughout the whole process. In the last contact session the teachers evaluated their experiences of the project work.

3. How to promote learning to learn in language classes at the primary

At the primary level (ages 7-12), Finnish children start learning a foreign language at the age of eight or nine (grade 3). Their ways of learning are strongly influenced by the individual stages of their mental development: young learners are still very holistic and concrete in their

ways of thinking and learning; they are not able to approach learning analytically. Therefore, in the beginning the teaching of the first foreign language has to be strongly action-based.

The four primary-level teachers in the project group were Anna-Liisa Ketovuori in Vahteron koulu, Pirkko Puonti in Ilolan koulu, Marjatta Sala in Päivänkehrän koulu and Ursula Viita-Leskelä in Helsingin II normaalikoulu. They first examined their goals within the CEF context: the aims of teaching a foreign language to children are 1) to teach them the rudiments of a foreign language, and 2) to provide them with the skills necessary to learn a language and to assess their learning. The teachers then concentrated on finding ways to promote learning to learn, the second aim. An important factor was to provide the learners with metacognitive skills, i.e. the means to talk about their learning.

For these purposes the most useful chapters of the CEF were about existential competence (5.1.3), study skills (5.1.4.3), heuristic skills (5.1.4.4) and awareness-raising (6.4.6.5). In order to promote these skills on a more concrete level, we decided to use Oxford´s (1990) lists of learning strategies as our tool, with special emphasis on memory strategies, memorizing being one of the young learners' strengths. The strategies were not taught as such, but always in an appropriate FL learning situation where they could be put to use immediately as Wenden (1991) suggests.

3.1 Research procedure

The teachers tried to find out their pupils' beliefs, knowledge and preferences about learning and then help them to develop a reflective approach to learning a new language. Procedures proposed by Wenden (1998: 531) for awareness-raising activities concurred with the research procedure:

1. the learners' opinions and preferences in general and in connection with a foreign language were elicited with two questionnaires and several self-assessment forms
2. the findings were discussed with the learners and partly with their parents
3. in the discussions the learners were confronted with alternative views and procedures, which the teachers hoped would lead to
4. future reflection on their learning and revision of their ideas and beliefs.

3.2 Tasks and situations to promote learning to learn

The teachers kept a log of the situations and tasks given to the learners in order to promote learning to learn. Besides normal FL classroom work, the pupils learned to deal with the new learning situation: they were given a new definition of the concept of homework and they became familiar with their FL books through a quiz.

Self-management began with remembering to bring one´s books and pencils along and proceeded to include the idea of having regular homework. Learners were given clear rules and advice for classroom work and homework. They learned cooperation by working in pairs and playing FL games in groups, and they practised maintaining attention by listening to tapes and repeating what was heard or answering questions.

Metacognition was developed through the teachers´ comments and by monitoring one´s own progress for example through self-assessment tasks. To help the development of metalanguage the self-assessments were discussed in class and the teachers used modelling

when going through homework. In connection with their language learning tasks, the pupils learned to use various memory strategies, analogy, grouping and comparison to Finnish.

3.3 Monitoring the learners: questionnaires and self-assessments

The questionnaire *I as a language learner* made for the upper classes was adapted for the young learners by making the questions concrete and keeping the questionnaires short and easy to answer. This did not completely succeed, partly because of a lack of time, so teacher supervision was needed in filling out the forms.

The answers mirrored the need to reconsider one´s routines in a new learning situation: for example parental help increased. Young learners tend to give an answer that they think pleases the teacher or corresponds to their beliefs and hopes: when asked about the skills needed in FL learning, pupils first claimed that they did nearly everything listed in the questionnaire, but later the answers became more realistic. The learners' preferred social forms broadened to include group work. Although many learners, both boys and girls, in grade three used lexical inferencing, boys also showed a marked tendency to ask somebody for help when confronted with a new FL word. In a follow-up a year later the same pupils had learned to consult the textbook first. The answers to the questionnaires gave the teachers many hints about what to add to language learning situations.

The aim of the structured self-assessment forms was to raise the learners' awareness of the importance of regular school work and the various aspects of learning a foreign language (see CEF 9.3.13). Although their answers did not always correspond to the teachers´ observations, they probably contributed to the reflection skills apparent in the open-ended self-assessments at the end of the school year. The majority of the third-graders´ open-ended assessments could be classified as examples of the mechanical level of learner reflection (Huttunen 1996): the pupils mentioned single items they had studied: *"The Snowman", "I know colours / animals / names of foodstuffs";* but the higher achievers´ self-assessments demonstrated a higher level: *"I can write well", "I can / cannot speak German / pronounce words wel", "I am satisfied with my learning"*. The assessments made by fifth-graders at the same school clearly show the development of metacognition: they were able to reflect more deeply on their learning: *"I have been quite active and attentive during lessons. I ought to learn vocabulary better."* or *"I have done my homework. I try to be as active as possible, but some things are difficult, for instance making different kinds of sentences."* On the other hand, another group of fifth-graders who had not practised self-assessment earlier were not able or willing to assess their learning. This proves that metacognitive skills and self-assessment can and should be taught early.

To sum up, one school year was probably not long enough to establish permanent changes in the learners´ behaviour, especially as there were only two 45-minute FL lessons per week. In order to become truly efficient, self-assessment and study skills need constant practice and discussion throughout the primary level and beyond, but the results obtained from these case studies are promising.

The teachers stressed the importance of cooperation with parents and among the colleagues at the four individual schools. Self-assessment and discussion with the pupils about their goals and study skills have become a working principle at the schools involved. On the whole, the CEF project provided the teachers with valuable information about their pupils, and served as an awareness-raising exercise for pupils and teachers alike.

4. Learning to learn at the lower secondary level

Three teachers at the lower secondary level: Sisko Heikkilä in Kouvolan yhteiskoulu, Krista Kindt-Sarojärvi in Helsingin II normaalikoulu and Sirkka-Liisa Sävy in Simonkylän yläaste did their action research with first-year learners, i.e. grade seven (aged 12-14); one teacher, Arja Pohjala in Olarin koulu, taught grade nine (aged 15-16). The first-year learners had just started in a new school after six years at the primary level. The time span reserved for the research and development work was rather short with two or three 45-minute lessons per week, which in two schools gave a total of about 70 lessons during the entire school year.

The teachers approached the learning to learn issues they had chosen (CEF 5.1.3; 5.1.4.3; 6.4.6.5; 6.4.7.1) with the help of three research questions which they found essential in their work:

a. How can learners be instructed to recognize their learning style and develop their learning strategies?
b. How can learners be assisted to take responsibility for their learning and how is the process surveyed?
c. How can awareness of learning and teaching processes be raised in the learning of vocabulary?

4.1 Learning styles and learning strategies

To help the students to recognize their learning styles and develop their strategies as suggested in the CEF (6.4.6.5), the teachers used part of the questionnaire designed by Dunn and Dunn (Dunn, R. et al. 1992). This part concerned the learners' perceptual strengths. As could be expected, most of the students were visually dominant while auditory learners came second. These are the two groups that teachers easily satisfy with texts accompanied by pictures, exercises, transparencies, tapes, oral instructions, videos etc. After having identified some kinesthetic and tactual learners, especially among the low achievers, the teachers provided tasks that took their preferences into account. The students made mind-maps, drew pictures, acted out weather conditions and everyday situations, used drama in learning the tenses, moved around the classroom to find the vocabulary tasks they were able to accomplish and used artists' mannequins to check verbs.

The teachers discussed the meaning of learning styles and strategies with the students both before the questionnaire was filled in and afterwards. Six months later, when the seventh-graders were asked about their dominant sense, few learners remembered what it was. Only one student reported to have changed her learning strategies in accordance with her perceptual strength. The teachers assumed that the learners were too young to benefit from this kind of awareness-raising directly. However, the information gained about the students' preferences was very important to the teachers because of its implications for teaching as described above.

It might be possible that learners are more interested in these kinds of features of their identities later in adolescence. The positive experiences of the teacher in grade nine seem to indicate this. She paid more attention to learning styles and introduced for example the concept of global and analytic learners to her students, who then tried to find their own category.

4.2 How to assist learners in taking responsibility for their learning?

All four teachers tried to develop this ability by means of self- and peer-assessment (CEF 6.4.2.3). They used questionnaires and learners' logs for self-assessment and awareness-raising games in peer assessment: the learners observed their partners in class either secretly or openly. Learner reflection also took place when the students talked about learning with their classmates or asked for more information when filling in the questionnaires. In their logs and self-assessment questionnaires the majority of the students reached the pragmatic level of reflection, evaluating the tasks in this way: "It was easy", "I learned a lot", "I liked it" (see Huttunen 1996).

On the whole, self-assessment had positive results. With the seventh-graders it appeared to increase goal-setting and motivation (see CEF 9.3.13). In some groups low achievers were able to raise their marks and in one group they benefited indirectly: they did not give up altogether. It seemed to be difficult for these students to predict their marks in a realistic way; usually they overestimated both their goals and their marks.

On the other hand, the ninth-graders showed increased planning and a greater ability to assess their own performance, evaluate their activity and predict their marks. Self-evaluation was a usual practice in their school. Their teacher assumed that metacognitive reflection progresses with age and intelligence. In general, self-assessment seemed to have a direct impact on the learners' metacognitive reflection, whereas peer assessment, in this form, appeared to affect it in an indirect way: the students learned about themselves by observing others.

4.3 Awareness-raising in the learning of vocabulary

In order to find out how the learners studied new words (CEF 6.4.7.1) the teachers prepared a vocabulary learning questionnaire. The results revealed that most learners studied words from the lists in their textbooks. Only a few practised the new words by using them in their own sentences. Outside school they learned new words mostly from television-programmes and songs.

In the seventh grade the teachers introduced several different ways of learning vocabulary. They tried to make the words more comprehensible by connecting them with perception and activities such as drawing or acting. These learning strategies were then practised in class for the next couple of weeks, and after that the students were occasionally reminded of these techniques. The activities showed an immediate effect, but after six months the students no longer remembered the strategies. This was an obvious result, one teacher observed, because the acquisition of new learning strategies takes time and regular practice. Some seventh-graders noticed that they had improved their listening skills, which could be attributed to an increase in vocabulary. The teacher of the ninth grade used a creative writing project with her mostly visual and global learners to foster vocabulary learning. She also tried out the scale for Creative Writing (CEF 4.4.1.2) to assess her students' written products.

To sum up, many of the methods or activities the teachers implemented were attempts to move in the direction they thought was right. All in all, awareness-raising took place. The teachers felt that they benefited from the whole process: studying the reading materials, cooperation, supervision, reflection and the actual work in class with a special focus. The teacher's log was a means to develop one's thinking and professional competence, or it served as an opportunity for instant therapy. It was instructive and therapeutic for the teacher to see later what kind of problems she had had and how she had managed to overcome them. In a

way the learners and teachers both developed through the same process: only their starting points and goals differed.

5. How to promote learning to learn English at the upper secondary level

The project was carried out by three teachers in three upper secondary schools: Tapio Erma in Olarin lukio, Kaija Komsi in Simonkylän lukio and Leena Säteri in Kouvolan Yhteiskoulun lukio. The research project took place in the first four periods of the school year. In Olari and Simonkylä the students did not stay in the same study groups or with the same teacher in every six or seven week period; thus the teachers faced new students at the beginning of every course, as is generally the case in Finnish non-graded upper secondary schools. In Kouvola, however, the research was carried out in permanent groups, who studied most of their subjects in English.

The objective of the research was to answer the question how the general competences in CEF 5.1.3 and 5.1.4 as well as self- and peer-assessment in CEF 6.4.2.3 can be applied to classroom situations. The research questions were:

1. How can the learner's existential competence be developed?
2. How can the study skills and heuristic skills mentioned in the Framework be improved?
3. How can self- and peer-assessment be applied to classroom work?

The content and results of the three main areas of the research and development work will be discussed in the following.

5.1 The learner's existential competence

In all three schools the students starting their upper secondary education answered a questionnaire *I as a language learner*, in which they reflected on their strengths and weaknesses, attitudes towards different ways of studying and their working habits, and also set themselves goals. The questionnaire was created by the participating teachers and based on some of the aspects of learning to learn described in CEF 5.1.3 and 5.1.4. The answers were discussed together in class. The teachers then concentrated on tasks that aimed at developing the students' self-awareness and self-confidence (CEF 5.1.3). Focus was also on the students' ability to learn from direct observation and participation in communication events, on their awareness of their strengths and weaknesses as learners and ability to identify their needs and goals and organize their strategies accordingly (CEF 5.1.4.3). They wrote learning logs, practised oral compensation strategies, assessed each other's oral presentations and were directed to observe their learning styles. During the course the students were asked to reflect on their work and evaluate their progress at the end of the course. In the Kouvola school the students spent more time on a group project where they set their own goals within a set framework and planned how to find, analyse, organize and present their information. In the Simonkylä school the students were asked to choose a well-done piece of work to hand in at the end of the course to be assessed for the course mark.

The teachers felt that the time spent on finding out as much as possible about the students' existential competence and working together with them to improve it was well spent. The teachers found that the students' attitudes towards learning English were generally positive, but some of the working methods they preferred were not very effective; for example, some students said that their preferred way of studying was listening to the teacher lecturing. Some

students lacked self-confidence, and again others felt threatened by certain communication situations, most notably by oral presentation. The research brought the teachers to the conclusion that the CEF can contribute to the learning process by defining the key elements needed in it and by making teachers aware of the importance of general competences in learning languages.

5.2 Self-direction and study skills

Particularly in the Kouvola upper secondary school a lot of emphasis was on students' self-direction. They were trained in the use of the heuristic skills listed in the CEF (5.1.4.4), in effective cooperation and how to set their goals and organize their strategies accordingly (5.1.4.3). In the first course the students were given the task of doing a group project on some feature of London. They were to set their goals, find material on the Internet or from other sources, analyse, organize and process it, and give an oral presentation on the topic. Among the students, those who were highly motivated high achievers were mature enough to plan and carry out their studies independently for two weeks, but the average and low achievers needed supervision by the teacher. In this research the second-year students seemed to benefit more from self-directed learning: they were able to set concrete goals for themselves, select the materials to be studied and monitor their studying well enough to achieve the goals. The general competences in CEF 5.1.3 and 5.1.4 should be considered important educational goals: learning them enables students to make use of self-directed lifelong learning.

In all three schools the students on one course concentrated on reading skills, which are essential in self-directed learning. The teachers compiled a questionnaire which they gave to the students at the beginning of the course to make them reflect on their attitudes to reading, their motivations and preferences (CEF 5.1.3) and on what they really did when they were assigned different texts to read. To develop the students' reading skills, they were offered a lot of reading material. Various ways to approach texts were introduced and practised: predicting, skimming, scanning, summarizing, mindmapping, asking and answering questions, cooperative reading, writing down key words and ideas and underlining. The aim was to provide students with a real choice in using different strategies.

In Kouvola and Simonkylä schools special attention was paid to one aspect of reading skills: lexical inferencing (see Vaurio 1998). Possible strategies to guess or deduce the meaning of new words were reviewed, the teacher modelled the inferencing process, and authentic texts from magazines and newspapers were used for practice. The students felt that practising lexical inferencing helped them in approaching challenging texts. In Simonkylä lexical inferencing was used as part of the course test, but many students felt that it was unfair as the skill was to a great extent explained by what might be called general knowledge or intelligence.

5.3 Self-assessment and peer assessment

In Chapter 6.4.2 of the CEF the roles of teachers and learners are discussed. In this research the teachers focused on the learners' role in evaluation: self-assessment and peer assessment. Peer assessment was used in all three schools in connection with the students' oral presentations. The students evaluated their peers in several ways: orally and in writing, answering questions, pointing out the best features in the presentations or suggesting improvements. They were told to focus on the positive: this was important as many students initially felt very uncomfortable making oral presentations. Peer assessment had a dual

purpose: it also made the assessing students realize why a presentation was good and thus learn through observation, another learning skill listed in the CEF.

In Simonkylä a simplified version of one of the illustrative scales in the CEF (Interviewing and Being Interviewed, 4.4.3.1) was used for self-assessment in interactive oral communication. The students were clearly interested in doing it; their evaluations were slightly higher than the teacher's. It appears that simplified versions of the CEF scales might be useful for self-assessment. For this kind of self-assessment to work, the teacher has to explain to the students exactly what skills they will be evaluated on (see CEF 9.3.13).

A learning log was used on several courses in all the schools with varying success. Many of the students, particularly the boys, were not very eager to do it. The boys, on average, seemed to be less motivated than the girls by reflective, fairly regular diary-type writing. Most of them never reached a level of real reflection, being satisfied with simply listing tasks they had carried out. When the log was part of portfolio work, the students' work was more thorough and reflective: they saw it as an integral part of the portfolio.

The teachers felt that a learning log can be a useful tool for the students in gaining knowledge about their cognitive processes and success as learners. The teachers realized that they had not paid enough attention to structuring and monitoring the writing of the logs: the students needed to understand the significance of the log for their learning and to get sufficient help in the form of questions and suggestions as to what they might reflect on. The better the log is integrated with the rest of their work and materials, the more positive and serious the students' attitude is.

6. Conclusions and evaluation

The results of classroom action research studies are not meant to be generalized. However, some conclusions can be drawn from the findings. Firstly, the results showed that the explicit development of features of existential competence (CEF 5.1.3), study skills (CEF 5.1.4.3), heuristic skills (CEF 5.1.4.4) and first attempts to foster metacognitive reflection (CEF 6.4.6.5) can be started at an early stage and should be continued throughout learners' school years.

The teachers at each level considered self-assessment (CEF 9.3.13) especially beneficial in developing learner reflection and consequently learners' metacognitive knowledge and strategies. When learners' metacognitive knowledge and strategies grow, their ability to take responsibility for their learning increases (CEF 6.4.6.5). They are able to plan, carry out and assess their own learning in a self-directed way. However, these studies seemed to support the research evidence that high achievers use metacognitive strategies more often and more effectively than low achievers (see Cohen 1998: 7).

Secondly, the teachers stated that systematic awareness-raising in the learning and teaching processes (CEF 6.4.6.5) greatly benefited not only the students, but also their teachers in their professional growth. Thirdly, a certain level of maturity seems to be required from the students to profit from the knowledge of their own learning preferences (CEF 6.4.6.5). However, this knowledge was important to their teachers as it helped them to plan suitable activities for different learners.

The teachers felt that it was a valuable experience to participate in the project. They noticed that reflection on professional action increased through this kind of activity. New information

was gathered, new teaching ideas were circulated, valuable information about learners' ways of thinking and learning was gained. The contact sessions and especially the cooperation with colleagues from the same school level were appreciated. However, all teachers complained about the constant lack of time and the increased work load. They also found doing classroom action research and reporting on it difficult because of too little preparation time and training before the project started. More support and advice were needed.

The positive outcomes and the problems above seem to be common in classroom action research projects (e.g. Nunan 1993; Huttunen 1999). The difficulties can be avoided to a great extent through good preparation, adequate training and thorough planning before the actual research and development work begins. In our project this phase was far too short for all involved, the coordinator included.

In general, the teachers considered their work with the Common European Framework rewarding. It was felt that the Framework could support teachers in their work. The CEF (1996) was found to be extremely rich in content but difficult to approach; it took time and effort before it opened up to the reader. One reason for this was that many concepts and terms differed from those we were used to. For example "existential competence" was one of those concepts, but gradually during the project work it was adopted and descriptions of learners' general competences, existential competence included, were found especially helpful. The revised final version of the CEF (2001) takes account of most of the points the teachers criticized. However, a subject index and a glossary of the main concepts and terms would make the use of the CEF as a handbook or reference book easier and the text more accessible.

As a consequence of the CEF pilot project, features of learning to learn have already been integrated into the school-specific FL curriculum in several participating schools. The textbook and materials writers among the teachers have started to apply the CEF to their materials in preparation. For example, in a textbook for primary school German and in one of the upper secondary school English series, revised in 2000, many of the learning to learn features in the CEF and this pilot project were included in the material. In-service training of foreign language teachers in learning to learn issues has also started with teachers from the pilot group as trainers. The report on the project in English (Jaakkola 2000a) and a volume for foreign language teachers in Finnish (Jaakkola 2000b) have been published. Although one school year was a very short time for a project like ours, it initiated a positive development, both among the teachers and the learners.

References

Carr, W. & Kemmis, S. (1986). *Becoming critical: education, knowledge and action research*. London: Falmer.

Chamot, A. U. & O'Malley, J. M. (1994). Language learner and learning strategies. In N. Ellis (ed.), *Implicit and explicit learning of languages* (pp. 371-392). London: Academic Press.

Cohen, A. (1998). *Strategies in learning and using a second language.* New York: Longman.

Cohen, L. & Manion, L. (1994). *Research methods in education*. Fourth edition. London: Routledge.

Council of Europe (1996) : *Modern languages: learning, teaching, assessment. A Common European Framework of reference*. Draft 2 of a framework proposal. CC-LANG (95) 5 rev.IV, Strasbourg.

Council of Europe (2000). *Common European Framework of Reference for Languages: learning, teaching, assessment.* Council of Europe. Cambridge: Cambridge University Press.

Dunn, R., Dunn, K. & Treffinger, D. (1992). *Bringing out the giftedness in your child*. New York: Wiley.

Elliot, J. (1991). *Action research for educational change*. Milton Keynes: Open University Press.

Ellis, R. (1997). SLA *research and language teaching*. Oxford: Oxford University Press.

Huttunen, I. (1996). Metacognition in the process of development towards learner autonomy. In U. Tornberg (ed.), *Focus on the language learner* (pp. 77-97). Uppsala Universitet. Centrum för didaktik.

Huttunen, I. (1999). The Finnish module: reflective teaching requires a lot of reflection. In E. Gierlinger (ed.), *Sharing the European language classroom: a project in international classroom action research* (pp. 51-66). Schriften der Pädagogischen Akademie des Bundes in Oberösterreich 3. Linz: Universitätsverlag Rudolf Trauner.

Jaakkola, H. (ed.) (2000a). *How to promote learning to learn the first foreign language. Piloting the Common European Framework of Reference in Finnish schools*. University of Helsinki. Department of Teacher Education. Research Report 215.

Jaakkola, H. (ed.) (2000b). *Kielenopiskelijan opiskelutaitoja kehittämässä. Kokemuksia opetuskokeilusta* [Developing language learners' learning skills. Experiences of a pilot project]. Helsinki: National Board of Education.

Kohonen, V. (1992). Foreign language learning as learner education: facilitating self-direction in language learning. In B. North (ed.), *Transparency and coherence in language learning in Europe* (pp. 71-87). Report on the Rüschlikon Symposium 1992. Strasbourg:
Council of Europe.

Niemivirta, M. (1997). Oppimaan oppiminen ja oppimisen säätely [Learning to learn and regulation of learning]. In J. Hautamäki, P. Arinen, A. Hautamäki, J. Lehto, S. Kupiainen, M. Niemivirta & P. Scheinin, *Oppimaan oppiminen Helsingissä. Peruskoulun kuudesluokkalaiset oppijoina keväällä 1996* (pp. 73-90) [Learning to learn in Helsinki. Pupils of grade 6 in primary school as learners in spring 1996]. Helsinki: Publications of the School Board of the City of Helsinki A 10.

Nunan, D. (1993). Action research in language education. In J. Edge & K. Richards (eds.), *Teachers develop teachers research. Papers on classroom research and teacher development* (pp. 39-50). Oxford: Heinemann.

Oxford, R. (1990). *Language learning strategies. What every teacher should know.* Boston: Heinle & Heinle.

Rauste-von Wright, M. & von Wright, J. (1994). *Oppiminen ja koulutus* [Learning and education]. Helsinki: WSOY.
Riding, R. & Rayner, S. (1998). *Cognitive styles and learning strategies. Understanding style differences in learning and behaviour.* London: Fulton.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Vaurio, L. (1998). *Lexical inferencing in reading in English on the secondary level*. University of Jyväskylä. Jyväskylä Studies in Education, Psychology and Social Research 145.

von Wright, J. (1992). *Oppimiskäsitysten historiaa ja pedagogisia seurauksia* [History and pedagogic consequences of views on learning]. Helsinki: National Board of Education.

Wallace, M. (1998). *Action research for language teachers*. Cambridge: Cambridge University Press.

Wenden, A. (1991). *Learner strategies for learner autonomy*. London: Prentice Hall.

Wenden, A. (1998). Metacognitive knowledge and language learning. *Applied Linguistics* 19 (4), 515-537.

Williams, M. & Burden, R. (1997). *Psychology for language teachers. A social constructivist approach.* Cambridge: Cambridge University Press.

Corresponding author:

Hanna Jaakkola
Siltavoudintie 8 C 30
FIN – 00640 HELSINKI
e-mail: hanna.jaakkola@helsinki.fi
Tel: 358-9-7287446

CHAPTER 5: MEETING THE ENGLISH LANGUAGE NEEDS OF REFUGEES IN IRELAND

David Little, Barbara Lazenby Simpson and Fiona O'Connor
Centre for Language and Communication Studies
Trinity College - Dublin

Introduction

This chapter reports on a project that has used the Common European Framework (CEF; Council of Europe 2001) and the European Language Portfolio (ELP) to develop English language provision for refugees admitted to Ireland. In it, we describe the use we have made of (i) the CEF to elaborate English language proficiency benchmarks that determine the needs and measure the learning achievements of individual adult refugees and (ii) the European Language Portfolio (ELP) as a pedagogical tool that encourages reflective learning and the growth of learner autonomy. We also briefly describe a related project designed to provide English language support for non-English-speaking pupils (mostly children of refugees and asylum seekers) in primary and post-primary schools in Ireland.

1. The background to our project

In 1996 the Irish Refugee Agency commissioned the Centre for Language and Communication Studies (CLCS), Trinity College Dublin, to investigate the provision of English language support for refugees in Ireland. The report that we wrote (Little and Lazenby Simpson 1996) outlined some general pedagogical principles, used these to evaluate current provision in Ireland, gathered information on practice in other countries, and made detailed recommendations for future action. The chief weakness we identified in current provision was a lack of direction and focus. For the most part refugees were assigned to general English courses in private language schools, which typically cater for the needs of young learners of English as a foreign language. This meant that refugee learners were exposed to the English needed by tourists in Ireland rather than by non-native-speaker members of Irish society. What is more, attendance at classes was patchy, records were sparse, and there was no clear path from reception, through a programme of English language learning, to vocational training and employment. Accordingly, we recommended that the provision of English language learning opportunities for refugees should be planned, monitored and evaluated with reference to a system of English language proficiency benchmarks and associated assessment procedures. When this recommendation was adopted, we undertook to develop the benchmarks and assessment procedures with reference to the CEF, and to use the ELP as a means of relating the teaching/learning process to the benchmarks. Progress towards this goal has come in two stages. First, in 1998 a consortium led by the Refugee Agency and including the Centre For Language And Communication Studies (CLCS) and FÁS (the national vocational training agency) secured funds from the European Union's Integra programme for a two-year project designed to improve refugees' access to vocational training and employment. This project, known as Interact Ireland, funded the work described below, which focuses on adult refugees preparing for vocational training and/or employment. Secondly, in 1999 the Department of Education and Science established a Refugee Language Support Unit on a two-year pilot basis under the aegis of CLCS. This has allowed us to begin to extend our work into the school sector as we report in our conclusion.

2. Pedagogical presuppositions

The general pedagogical principles outlined by Little and Lazenby Simpson (1996) assigned a central role to the development of learner autonomy, defined as the capacity to assume reflective responsibility for one's own learning (see, e.g., Holec 1981, Little 1991). Although arguably essential to all truly successful learning, the development of autonomy is particularly important in the case of refugee learners beginning a new life in a new country. If they are to become fully integrated members of their host society, they must profit from every language learning opportunity that presents itself, outside the classroom as well as inside; and the best way of equipping them to do this is by helping them to develop the capacity to set their own learning goals, monitor learning activities, and evaluate their own progress. This general orientation coincides with the proposal in the CEF (section 8.3.2) that all language learners should be helped to "establish metacognitive control over their own competencies and strategies".

Refugee learners often experience feelings of displacement and loss of independence and control (McPherson 1997), and this is another reason for working to develop their autonomy in the language classroom. As Little has pointed out (1991: 8), autonomy should produce more focused and more purposeful learning; because responsibility for the learning process lies with the learner, there should be no barriers between formal learning and other areas of life; and this should allow learners to transfer their growing capacity for autonomous behaviour to other spheres of activity. In other words, language teaching that is intent on developing learner autonomy can give refugee learners a great deal more than proficiency in the language of their host society.

But the chief purpose of all language provision for refugee learners must be to make them autonomous users of their target language (for an exploration of the two-way relation between autonomy in language learning and autonomy in language use, see Little 1996). This implies that communicative tasks should play a central role in the language classroom (see, e.g., Nunan 1989, Long and Crookes 1993, Pica et al.1993). Among the benefits claimed for task-based language learning are a focus on content rather than form and the need to negotiate meaning with other learners involved in the same task (Skehan 1998). In addition, when the task is based on the use of authentic texts, the learners are inevitably drawn into the communicative world of the target language (Little 1997: 225). These considerations help to explain why communicative tasks came to play a central role in the specification of our benchmarks; for a strong task-orientation allows the benchmarks to be used simultaneously as a basis for course design and the yardstick against which learner achievement is measured.

3. Developing English language proficiency benchmarks for use with adult refugees

3.1 Determining the benchmarks format

In order to determine the format of our benchmarks (RLSU 2000), we had first to establish the domains in which our learners needed to operate and the communicative activities in which they must be proficient. Acknowledging that the number of possible domains is indeterminate, the CEF suggests that for general purposes of language learning and teaching it may be useful to distinguish four: personal, public, occupational, and educational (4.1.1). Given the relatively narrow focus of our refugee learners, we found these categories too general. The four domains that we identified as an appropriate basis for our benchmarks were: language learning, interpersonal, public, and career planning. The first of these was dictated by our pedagogical orientation to learner autonomy, the second by our learners' need to pay

particular attention, in language learning as well as in language use, to politeness conventions and other culturally determined norms of interpersonal communication in Irish society, the third by their need to cope with public services, and the fourth by the fact that their language learning is intended to lead them to vocational training and/or employment.

Chapter 4 of the CEF considers communicative proficiency from the perspectives of communication themes, communicative tasks and purposes, communicative language activities and strategies, communicative language processes, and texts, while Chapter 5 addresses the user/learner's competencies and strategies. All of these perspectives are important in curriculum planning and the development of learning materials, but the CEF gives particular prominence to communicative language activities in its elaboration of scales and levels. Using a variant of the four skills approach, it divides communicative activities into four categories, some of which are divided into sub-categories: productive (oral production – speaking, written production – writing); receptive (aural reception – listening, visual reception – reading, audio-visual reception); interactive (oral interaction, written interaction); mediating. We began by attempting to apply this approach to the four domains we had identified, but it quickly became apparent that the result would be too complex and

LANGUAGE LEARNING	INTERPERSONAL
• Learning to learn	• Personal identification
• Setting course objectives	• Learning about others
• Cultural awareness	• Maintaining social relationships
• Present and get information	• Conversation
• Learning about the workplace	• Correspondence
• Learning about tests	
PUBLIC	**CAREER PLANNING**
• Using public services effectively	• Career planning
• Making complaints	• Job seeking
• Learning about the media	• Preparing a curriculum vitae
	• Using the telephone
	• Interview skills
	• Learning about computers

Table 1

domains and activities (RLSU 2000, p.21)
cumbersome for our purpose. Instead, we decided to specify activities in terms of task clusters, which could then be broken down into individual tasks. Clearly in line with the task-based pedagogical approach referred to above, this approach offered to yield benchmarks that would serve their intended dual function of providing (i) a transparent resource for course planning and (ii) criteria for assessing learners' achieved proficiency. The scaffold of domains and activities/task clusters on which we built our benchmarks is shown in Table 1.

3.2 Scaling and levels

The CEF proposes that "an outline framework of six broad levels gives an adequate coverage of the learning space relevant to European language learners" (3.2). These six levels are grouped in three broad bands: A – Basic User, B – Independent User, and C – Proficient User; and each of the three bands is divided into two: A1 – Breakthrough, A2 – Waystage, B1 – Threshold, B2 – Vantage, C1 – Effective Operational Proficiency, and C2 – Mastery. We knew from the outset that on entry to our courses refugee learners tend to cluster around the Waystage and Threshold levels, and we hypothesized that they would have no difficulty in coping with the communicative demands of vocational training or the workplace if they could be brought to Vantage level (though the variable demands of different training programmes and occupations warned us against being too dogmatic). Accordingly, these were the three levels on which we focused our benchmarks. However, because we wanted the benchmarks to be used by our learners as well as their teachers, we decided against using the labels Waystage, Threshold and Vantage. Instead we adopted the terms Access (to indicate that the learner has enough English to access our courses), Entry (to indicate that the learner is ready to enter a programme of vocational training), and Independent User (to indicate that the learner is ready to move into employment).

Once we had decided on these three levels, we had to break down each of our activities/task clusters into individual tasks, and then provide a descriptor for each task at each of the three levels. In doing this, we assumed that there is a progression of difficulty from familiar to less familiar tasks, and from tasks that require descriptive and narrative language to tasks that depend on the language of analysis and comment (for the background to this approach, see North and Schneider 1998). We aimed to make our descriptors as brief as possible, wording them positively as "Can do" statements. We elaborated them by shuttling back and forth between the sample descriptors given in Chapter 3 of the CEF, especially the profile of language skills summarized in Table 2, and the specific tasks implied by our scaffold of domains and activities/task clusters (Table 1 above). Accordingly, our benchmarks can easily be related back to the CEF.

3.3 The fully developed benchmarks – three examples

To illustrate our fully developed benchmarks we have chosen one example each from the LANGUAGE LEARNING, INTERPERSONAL, and CAREER PLANNING domains. *Setting course objectives* (RLSU 2000, p.23) is the activity with which we begin all our courses. By exploring different aspects of learning, including the learners' motivation and commitment, it helps the learners and their teacher to begin to develop work patterns that emphasize the importance of setting manageable learning targets, monitoring the learning process, and evaluating learning outcomes. In other words, it inaugurates the process of developing learner autonomy. Clearly, some of the tasks included in this activity – e.g., *devise*

a classroom contract, keep a course diary – accompany the learners throughout their course. *Conversation* (ibid., p.35) is an activity of central importance to the development of learners' communicative proficiency inside and outside the classroom. Refugees often come to Ireland from cultures that are very different from the one to which they must now adapt, and they often need to work hard to master appropriate conversational skills. For this reason we focus particularly on pragmatic features of conversational interaction, such as turn-taking and politeness conventions. *Career planning* (ibid., p.28) is integral to the participants' ultimate goal, which is to secure long-term employment. The tasks associated with this activity are heavy with cultural

ACTIVITY: **Setting course objectives** — *The learner can ...*

Tasks	**Access level (Threshold 1)**	**Entry level (Threshold 2)**	**Independent user (Vantage)**	
Discuss future hopes/intentions	... use general language, may need prompting with questions. Can also use repetition of phrase such as, I think and use present tense structures.	... express ideas with some precise language. Can use expressions of hope and future tenses.	... use a wide range of formulations. Can use hypothetical mode as well as future tenses.	*Spoken interaction*
Interpret and produce timetables	... understand and name some standard pictorial symbols and abbreviations.	... understand and reformulate information given and understand most abbreviations.	... explain timetable to others using language appropriate to the language level of the other.	*Spoken interaction*
Devise a classroom contract	... follow a model in order to write simple clauses appropriate to an informal contract.	... write short statements of intent and expectation with some dictionary use.	... write detailed statements of intent and expectation using complex sentences.	*Written production*
Recognize course structure	... learn classroom vocabulary and relate it to coursework.	... identify different aspects of coursework.	... discuss aspects of coursework, adding comment and opinion.	*Heuristic knowledge*
Keep a course diary	... use simple language with dictionary. Writing is inaccurate in both spelling and grammar.	... use descriptive language and add comments. Writing is simple and correct.	... use a broad range of vocabulary and idiomatic expressions. Writing is accurate and complex.	*Written production*
Review and examine objectives	... re-read initial objectives and compare them with current position.	... recognize when objectives have been reached and use old plans to formulate new objectives.	... examine old objectives and identify how they were achieved. Can use this information in setting new goals.	*Heuristic knowledge*
Discuss plans and arrangements	... talk about classroom activities in factual terms. Can ask basic questions about future plans.	... describe plans in some detail. Can ask others for more information.	... describe and comment on own plans. Can ask others for reasons and explanations.	*Spoken interaction*
Read about courses	... understand basic facts in course descriptions.	... obtain information about course content.	... obtain detailed information about course content and organization.	*Reading /spoken interaction*

Setting course objectives is fundamental at the beginning of a course. The learners and teacher(s) develop patterns of working which will continue to be used throughout the course. The process of review and re-evaluation is central to the development of autonomy and self-learning.

ACTIVITY: Conversation —*The learner can* ...

Tasks	Access level (Threshold 1)	Entry level (Threshold 2)	**Independent user (Vantage)**	
Turn-taking	... handle short social exchanges and simple routine exchanges.	... hold and yield the floor but with some hesitation.	... follow and take part in naturally flowing conversation.	*Spoken interaction*
Express opinions	... give opinions without the support of explanation.	... express personal opinions on topics of personal interest.	... support opinions on most general topics.	*Spoken interaction*
Express emotions	... express likes and dislikes and state beliefs or opinions using everyday language	... express and respond to surprise, sadness, pleasure, interest and indifference.	... convey degrees of emotion and highlight the personal significance of events and experiences.	*Spoken interaction*
Maintain conversation	... ask and answer questions and exchange ideas on familiar topics.	... maintain conversation but may be hard to follow when making a point.	... sustain relationships with native speakers without unintentionally amusing or irritating them.	*Spoken interaction*
React to others	... express agreement or disagreement politely. Uses mostly "yes", "no" to indicate agreement etc.	... give brief comments on the views of others. Uses some fillers to show interest agreement, etc.	... react appropriately to agree or disagree with others' opinions. Uses appropriate fillers and gestures.	*Listening/Spoken interaction*
Follow topic shift	... follow main points in an informal discussion in standard dialect. May get lost if topic shifts quickly.	... enter unprepared into conversations on everyday topics.	... take an active part in informal discussion in familiar contexts, commenting, giving point of view, weighing alternatives.	*Spoken interaction*
Follow others' conversation	... generally identify the topic and follow main points.	... follow much of what is said provided use of idioms and non-standard usage is avoided.	... catch most of what is said but may find it difficult to participate in groups of speakers who do not modify their language in any way.	*Spoken interaction*

ACTIVITY: Career planning —The learner can …

Tasks	**Access level (Threshold 1)**	**Entry level (Threshold 2)**	**Independent user (Vantage)**	
Assess skills and qualifications	... list his/her skills and qualifications using a dictionary.	... describe skills and qualifications, emphasizing relevant skills.	... describe skills and qualifications in some detail. Uses specialist vocabulary where appropriate.	*Heuristic knowledge*
Prepare a C.V.	... write down basic information required for a CV and prepare the document using a template.	... gather and organize information to prepare a CV, including interests and hobbies.	... organize a CV to reflect different aspects of work and experience.	*Written production*
Identify career opportunities	... match very specific job description with own skills and qualifications.	... read through course brochures and job ads to identify areas related to skills and qualifications. Makes implicit as well as explicit connections.	... read and access a wide variety of career related texts, including help manuals.	*Reading*
Examine work goals	... list different work goals, using a dictionary for precision.	... explore different work goals in relation to identified strengths and weaknesses.	... identify different work goals and explain how they relate to an overall career plan.	*Heuristic knowledge*
Discuss career options	... describe hopes and wants. May find it difficult to find correct vocabulary to give precise information about jobs.	... explain origins of career plans. Uses appropriate terms and clear statements of purpose.	... explain career plans, referring to past experience and future directions, exemplifying for clarity.	*Spoken interaction*
Discuss career plans	... answer simple questions in relation to job areas of personal interest. Needs assistance with vocabulary.	... describe different options and possibilities in relation to career path. Can also explain reasons for choices made and express feelings about future expectations.	... comment on and analyse different options in relation to career path.	*Spoken interaction*

Career planning is a central area in the pre-vocational sector as the objective of the course is to equip the learner to operate independently in the job market. This links with the activities of CV preparation and job seeking.

implication, and it is essential that our refugee learners master them if they are to compete on equal terms with Irish nationals when applying for jobs.

3.4 Using the benchmarks to plan and evaluate courses

Traditionally curriculum design is based on the needs, interests, and prior knowledge of a "typical" learner. This approach has often been criticized for paying insufficient attention to the needs, interests and developed capacities of the individual learner; and it is clearly inappropriate when the members of the target group come from different countries, with different educational backgrounds, different levels of proficiency in the target language, and different ultimate objectives. In such a situation there are clearly as many profiles of language learning needs as there are individual language learners. Accordingly, our benchmarks are designed to allow each course participant to set his or her own learning objectives by identifying the activities and tasks relevant to his or her personal goals. They also allow learners to compare their current level of proficiency with their target level and, with the assistance of their teacher, to set out a learning programme comprising short-term goals that can be seen as stepping stones to their desired proficiency level. Short-term goals may range from the relatively basic task of developing topic-specific vocabulary to complex tasks such as compiling a curriculum vitae. Individual learning goals for the course as a whole can be expressed in grid format as a profile of target proficiency, and this can be used to evaluate not only the progress of the individual learner but the effectiveness of the course overall. When there is a high degree of match between target profiles and learning outcomes, the course can be said to have achieved its aims. Any mismatch can feed back into future course development.

3.5 Developing assessment instruments

Besides facilitating the specification of learning targets, the benchmarks also identify proficiency thresholds; they state how well and with what kind of assistance learners at different levels can perform certain communicative tasks. Teachers can thus use the benchmarks for purposes of continuous assessment, and learners can use them for self-assessment. But the benchmarks can also be used as the basis for designing language tests to measure learners' proficiency at a particular stage in their learning. In the context of pre-vocational language provision for refugees, there are two points at which this kind of assessment is important. We need to know when learners are ready to benefit from a pre-vocational course; and once they are taking the course, we need to know when they are ready to meet the communicative demands of mainstream training or the workplace.

In designing fixed-point proficiency tests related to our benchmarks, we have again assumed a continuum of difficulty from familiar to less familiar tasks, and from tasks that require descriptive and narrative language to tasks that depend on the language of analysis and comment (cf. p.00 above). We have also assumed that it is possible to develop a set of tasks that are open to manipulation by learners, so that they can perform adequately at their own level on this continuum. Task performance can then be scored in relation to a performance index derived from the benchmarks; in other words, test scores can be used to map the learner's profile of proficiency onto the benchmarks (cf. Peirce and Stewart 1997). At the time of writing, assessment instruments have been designed and preliminary field trials have been carried out, but further progress has been delayed by the need to provide immediate support for teachers of non-English-speaking pupils in the schools.

4. The role of the European Language Portfolio in relation to the benchmarks

4.1 Pedagogical principles

As we explained in 2 above, the development of learner autonomy is central to our pedagogical approach, and its importance is strongly implied by the structure and content of our benchmarks. The European Language Portfolio (ELP), conceived by the Council of Europe as a companion piece to the CEF, is the tool that we use to stimulate the growth of learner autonomy by encouraging our learners to reflect on the goals, process and outcome of their learning. The ELP comprises three elements: a passport that summarizes the owner's language learning experience and achievements; a biography that reflects discursively on the owner's experience of learning and using languages, with particular reference to intercultural encounters; and a dossier that contains samples of what the owner can do in the language(s) he or she knows. The link between the ELP and the CEF is made explicit by the inclusion in the passport component of a self-assessment grid derived from the latter (specifically, Table 2).

The version of the ELP developed by CLCS was designed to be used in two quite different language learning contexts: two-year foreign language modules taken by university students who are not studying a foreign language as part of their degree, and English language courses for adult refugees at reception and pre-vocational levels. Although the language learning objectives of university-level and adult refugee learners are by no means identical, the fundamental pedagogical approach underpinning the use of the ELP is the same. The development of learner autonomy depends crucially on the development of conscious awareness (Schmidt 1994, 1995) of all aspects of the language learning process. Exploiting the capacity for autonomous behaviour that is present to one degree or another in all learners (see Little 1991, Usher et al. 1997), the ELP helps to activate a process in which the learner's gradually evolving understanding of what learning entails, runs parallel to an increasing ability to identify and categorize those learning activities that can meet a previously stated objective. Specifically, the ELP activates and supports the learning process by helping the learner to

1. understand the extent of his or her existing language knowledge (passport);
2. identify the language learning potential that exists in the classroom and in the world outside, in relation to his or her needs (biography);
3. engage in self-monitoring through the use of reflection, self-assessment and individual objective setting (biography and dossier);
4. compile a dossier that supports his or her individual objectives (dossier).

Learning starts with the identification and articulation of individual learning goals, proceeds as a series of recursive cycles in which the setting of long- and short-term objectives is followed by the identification of the means of achieving the objectives and the specification of a timescale for achievement.

In the case of refugee language learners the ELP fulfils a dual role. Because the ultimate objective of the majority of adult refugees is to progress to vocational training or employment appropriate to their educational level, it is inevitable that the individual will wish to provide evidence of his or her ability to perform through the target language. Thus in the final stages of formal learning the ELP becomes a product or showcase portfolio containing only samples of the holder's output that are relevant to an interview or some other selection procedure.

Prior to this, however, the ELP is a process portfolio, providing a means of recording objectives, a focus for self-monitoring, and a place in which learning activities can be organized and work produced.

4.2 The ELP's reporting function: the passport and self-assessment

As noted above, the ELP comprises three components, a passport, a biography, and a dossier. The passport section is a formally presented document in which the holder notes his or her existing level of proficiency in all the languages he or she has learnt. The self-assessment of global language proficiency is carried out using Table 2 from Chapter 3 of the CEF, which is reproduced in the passport section and is intended to support many of the uses to which the ELP will be put by the learner. Self-assessment may be based on the certified achievement of formal learning, but it may also include uncertified language learning to any level and under any conditions. It is not unusual for refugees to acquire a number of languages to varying levels of proficiency as they move from place to place, in addition to the language(s) they have acquired naturalistically in their country of origin.

The initial interaction with the CEF for the purpose of filling in the passport requires learners to reflect on their existing state of language proficiency, if not on their language learning experience. In many cases this proves illuminating, for it encourages learners to view their proficiency in language(s) other than their mother tongue as evidence of their effectiveness as learners and their potential employability. If there were no initial self-assessment of this kind, many refugee learners would be overwhelmed by the prospect of learning the language of the host society; as it is, they begin to understand themselves as language learners. What is more, through the experience of self-assessment they begin to construct their own understanding of the subject matter to be learnt and to perceive and internalise relevant and meaningful standards of achievement (see, e.g., Boud 1995: 11, Campbell et al. 1996: 278). In this way they begin to assume responsibility for their own learning.

4.3 The ELP's pedagogical function

In a pedagogy that emphasizes the growth of individual learner responsibility, there are a number of recurring activities that cumulatively give coherence to the learning process and thus constitute the central pillar of learning. Regular reflection is one of the most important of these activities, and it is particularly well supported by the biography component of the ELP, which provides a means of recording personal learning targets, identifying effective learning activities and contexts, and – particularly important for refugee learners – includes a section for recording experience of "heritage" languages. As a supplement to the biography, in our refugee language courses we require our learners to keep diaries that are used for process evaluation (Kohonen 1999: 279) or, more accurately, process self-evaluation.

The dossier requires learners to select formative and summative samples of their work and thus provides a clear justification for learning activities (see Cole et al. 2000: 12). This component of the ELP complements the biography in stimulating learners to reflection that takes the form of questions like "Why should this be included?", "Does this demonstrate my learning?", "Should I do this again in a different way?", or "How does this contribute to my learning objective?" Learners assemble materials in the dossier in chronological order, which helps them to see the progress they are making as their dossier expands.

Whereas the CEF is used in the passport as the basis for summative self-assessment, the biography and dossier are dynamic, providing the basis for all the formative activities that underpin the learning process. As they use these sections of the ELP, learners gradually bring their short- and long-term objectives into sharper focus. By recording and reflecting on their learning experiences, they begin to recognize and understand the nature of language learning as it occurs inside and outside the classroom, and the potential for advancing learning in many contexts becomes clearer to them. They may, for instance, recognize that language learning benefits can be gained from involvement in a sporting club or from regular visits to the cinema. Thus the barriers which may have existed between formal language learning and living in the target language society are lowered and eventually disappear (see Little 1991: 8).

The scope of reflective self-evaluation is wide-ranging, covering issues such as content of learning, material selected for learning activities, affective factors related to success and failure in learning, and evaluation of the overall effectiveness of learning. In the biography, the assessment that takes place during reflection is also related to the CEF. However, experience of using the CEF for regular self-assessment has raised one serious problem. Although learners generally appear to be capable of summative self-assessment (see Little 1999: 3), data gathered from users of the ELP suggest that they lack awareness of the learning process and especially of the timescales within which they are likely to achieve their objectives. As a consequence, there is a tendency for them to establish learning targets with unrealistically short timescales. They may express their targets in terms of the CEF, for example by identifying the achievement of a higher proficiency level as a short-term interim objective, when the movement from one CEF level to the next would normally involve a longer period of study. While it is clearly not possible to legislate for the pace of learning (indeed, rhythm and pace should properly be determined and controlled largely by the learner; see Holec 1981: 3), in the absence of any caveat with regard to the timescales inherent in the CEF, it seems inevitable that learners will misinterpret this aspect of the learning process. When learners are largely unaware of the learning process, any attempt to use the CEF as a means of staging learning is likely to result in their setting unrealistic and unachievable targets or skewing the CEF to bring it into line with achievable objectives.

Evidence of the difficulty that learners encounter in using the CEF to maintain on-going reflective self-assessment suggests a need for more detailed descriptions of proficiency relevant to particular domains of language learning. In the case of pre-vocational language courses for refugees, the elaboration of proficiency benchmarks has largely solved this problem, though the solution has yet to be reflected in a revision of our version of the ELP. The benchmarks provide sharply focused descriptions of learning tasks and the activities that support them; and the tasks are firmly embedded in the specific demands of pre-vocational language courses organized on a modular basis, so that they support short-term objective setting across a range of modules and self-assessment at regular intervals. The use of tasks and activities that are transparent to learners because they are carried out in the classroom, confers a high degree of face validity on the self-assessment process. Furthermore, the use of activity and task descriptors that focus closely on the learning demands of a particular course, but which in turn may be mapped at appropriate intervals on to the CEF, underpins the learning process in terms of supporting choice of materials and learning activities, individual reflection, and self-assessment without compromising the integrity of the CEF.

5. Conclusion: a further field of application

The first phase of our benchmarks project, focused on the English language needs of adult refugees seeking vocational training and/or employment, was completed towards the end of 1999. Although (as noted above) much work remains to be done on the associated proficiency tests, there is already ample evidence that the benchmarks facilitate the setting of individual learning targets and the development of appropriate learning activities and materials, especially when they are used in conjunction with the ELP. Since the beginning of 2000 they have provided an essential underpinning for the work of the Refugee Language Support Unit. Besides being used to assign adult refugees to appropriate English language programmes, they have helped to determine the design of a database that will track the progress of refugees from reception through various forms of learning to employment, and they are being drawn on to specify the English language provision for which private language schools continue to tender. Derived as they are from the CEF, they confirm its effectiveness as a planning tool.

The second phase of our benchmarks project, which has to do with non-English-speaking children in Irish schools, provides us with another field of application. By the autumn of 2000 two further sets of benchmarks had been developed, for use in primary and post-primary schools, as well as two further versions of the ELP. The elaboration of these new sets of benchmarks was constrained by two factors. First, it is by no means uncommon for children of refugees and asylum seekers to start attending school with zero proficiency in English; thus we focussed our benchmarks on the first three of the Council of Europe's six proficiency levels, BREAKTHROUGH, WAYSTAGE, and THRESHOLD. Secondly, it was necessary to reflect the structure and content of the primary and post-primary curricula. Both sets of benchmarks begin with a skills-based profile of communicative proficiency at three levels – essentially an adaptation of the relevant sections of Table 2 in Chapter 3 of the CEF (in the primary benchmarks our adaptation seeks to take account of maturational progression in language skills as summarized by Sainsbury [1999]). The two sets of benchmarks differ thereafter. In the case of the primary benchmarks we identified and elaborated fifteen units of work corresponding to recurrent themes in the curriculum: *myself; school; food and clothes; colours, shapes and opposites; people who help us; people and places in other parts of the world; the local and wider community; caring for the locality; time; weather; transport; seasons and holidays; festivals; animals and plants; water.* In their structure these units of work are closely similar to the benchmarks developed for the adult pre-vocational sector and illustrated above. In the case of the post-primary benchmarks we used our profile of communicative proficiency to elaborate twelve modules of learning in three categories: PERSONAL – *personal identification, learning how to learn, cultural awareness*; SUBJECT INDEPENDENT – *classroom interaction, academic writing, cross-curricular learning;* and CORE CURRICULUM SUBJECTS – *English language, English literature, science subjects, mathematics, history and human geography, physical education*. The modules that are pupil-oriented tend to be task-based, while those that are curriculum-focussed tend to be skills-based.

The primary and post-primary benchmarks, together with the two new versions of the ELP, form the basis of a national programme of teacher development and pupil support. In the light of our experience with the pre-vocational benchmarks and adult ELP, we are confident that they will quickly establish themselves as powerful tools for planning, implementing and monitoring teaching–learning. Both benchmark projects will remain incomplete, however, until we have designed and validated appropriate assessment procedures.

References

Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

Campbell, L., Campbell, B., & Dickinson, D. (1996). *Teaching and learning through multiple intelligences*. Needham Heights, Mass.: Allyn and Bacon.

Cole, D., Ryan, C., Kick, F., & Mathies, B. (2000). *Portfolios across the curriculum and beyond.* Second Edition. Thousand Oaks, Ca.: Corwin Press.

Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment.* Cambridge: Cambridge University Press.

Holec, H. (1981). *Autonomy and foreign language learning*. Oxford: Pergamon.

Holec, H. (1996). *Self-directed learning: an alternative form of training*. In D. Little, R. Richterich, & H. Holec, *Strategies in language learning and use* (pp. 75–127). Strasbourg: Council of Europe.

Kohonen, V. (1999). Authentic assessment in affective foreign language education. In J. Arnold (Ed.), *Affect in language learning* (pp. 279–94). Cambridge: Cambridge University Press.

Little, D. (1991). *Learner autonomy 1: definitions, issues and problems*. Dublin: Authentik.

Little, D. (1996). Strategic competence considered in relation to strategic control of the language learning process. In D. Little, R. Richterich, & H. Holec, Strategies in language learning and use (pp.11–37). Strasbourg: Council of Europe.

Little, D. (1997). Responding authentically to authentic texts. In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 225–48). London: Longman.

Little, D. (1999). *The European Language Portfolio and self-assessment*. Strasbourg: Council of Europe (DECS/EDU/LANG (99) 30).

Little, D., & Lazenby Simpson, B. (1996). *Meeting the language needs of refugees. Some general principles, a report on current practice in Ireland and elsewhere, and some recommendations.* Dublin: Trinity College, Centre for Language and Communication Studies.

Long, M. H., & Crookes, G. (1993). Units of analysis in syllabus design – the case for task. In G. Crookes & S. M. Gass (Eds.), *Tasks in a pedagogical context. Integrating theory and practice* (pp. 9–54). Clevedon: Multilingual Matters.

McPherson, P. (1997). *Investigating learner outcomes for clients with special needs in the Adult Migrant English Programme*. Sydney: NCELTR.

North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing* 15 (ii), 217–63.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.

Peirce, B. N., Stewart, G. (1997). The development of the Canadian language benchmarks assessment, *TESL Canada Journal* 14 (ii), 17–31.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communicative tasks. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning. Integrating theory and practice* (pp. 9–34). Clevedon: Multilingual Matters.

RLSU (2000). *English language proficiency benchmarks for the pre-vocational sector.* Dublin: Trinity College, Refugee Language Support Unit.

Sainsbury, M. (1999). *Tracking significant achievement in primary English*. Bath: Hodder and Stoughton.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics, *AILA Review* 11, 11–26.

Schmidt, R. (1995). Consciousness and foreign language learning: a tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Technical Report #9. Honolulu, Hawai'i: University of Hawai'i, Second Language Teaching and Curriculum Center.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Usher, R., Bryant, I., & Johnston, R. (1997). Adult education and the postmodern challenge. London & New York: Routledge.

CHAPTER 6: DEVELOPING THE SWISS MODEL OF THE EUROPEAN LANGUAGE PORTFOLIO

Peter Lenz and Günther Schneider
University of Fribourg, Lern- und Forschungszentrum Fremdsprachen

Introduction

The Swiss model of the European Language Portfolio (ELP) is a pioneer application of the Common European Framework of Reference (CEF). The two works mutually influenced each other during the development process. From a historical perspective the ELP is not just an application of the Common European Framework. On the one hand, fundamental considerations underlying the ELP concerning issues like the definition of communicative competence, the categories of its description and the approach to its assessment emerged from the work done within the Framework working party in which two authors of the Swiss Portfolio participated. On the other hand, some of the instruments developed primarily for the Swiss ELP model also became part of the Common European Framework.

In this chapter we describe a Swiss National Science Foundation project, designed to provide elements for an ELP, in which calibrated descriptors were produced which served as the basis for assessment tools now contained not only in the ELP but also in the CEF.

1. A broad view of assessment

There are strong historical links between the Common European Framework and the European Language Portfolio, in as much as the two works cover the same ground. The CEF as a whole provides a wide framework for a great variety of developments in the field of language learning, the ELP, however, focuses on assessment, and within assessment very much on self-assessment. The ELP is made for the use of the learners and aims to suit their needs, to empower them to assess and document their language proficiency and intercultural competence, thus enabling and motivating them to plan their further learning - and this in a life-long perspective of growing plurilingual competence. Besides being a pedagogic instrument, the ELP also has a presentation function. For that purpose it contains elements which may help learners to provide the kind of evidence of their language competence and intercultural experience which is at the same time comprehensive and understandable for others.

The ELP can be seen as a tool that shares with the CEF and puts into practice **a broad view of the assessment of language proficiency and intercultural competence** (cf. CEF: ch. 9). Such a *broad view* may have several facets (cf. Schneider 1999a: 5):

- the object of assessment: different aspects of communicative competence as well as partial competencies are taken into consideration: communicative competence as a complex phenomenon encompassing language and intercultural aspects including *pluri*lingual competence in itself;
- the time frame: assessment not only as summative events at certain points in time but also as a means of observing progress (formative), as a continuing activity stretching even beyond the end of formal education;

- those who assess: specialized institutions, testers, teachers, hosts, peers, and the learners themselves;
- the instruments of assessment: exams, tests, observation by peers or native speakers (based on descriptions of language proficiency), self-assessment tasks and checklists, dossiers containing learner products;
- the means of feedback: grades, diplomas, descriptions of competence and levels, progress reports - all stating explicitly the competencies and skills they refer to.

One constant in this variation is a strong commitment to transparent criterion-referenced assessment based on a coherent scale of communicative language proficiency.

There are good reasons for such an approach to assessment. They include considerations of completeness, transparency, coherence, and also fairness towards learners and the assurance of quality in language education. Self-assessment is strongly promoted not only for pedagogic but also for political reasons based on the assumption that citizens of democratic societies need to be autonomous personalities.

2. The instrument included in the Swiss model of the European Language Portfolio

The ELP can be seen as an attempt to use assessment - in the broad sense described above - as a driving force to make language learning more motivating, more efficient and more transparent.

The Swiss model is intended for learners from the age of about 15 years upward. Physically, it is an A4 ring binder consisting of three main parts and an additional section:

- the Language Passport section subdivided into a language passport booklet and a sub-section which can accommodate diplomas, certificates and attestations;
- the Language Learning Biography section;
- the Dossier section.

The fourth section contains master copies of all those pages that may be used up in the course of time and need to be replaced.

Each of the instruments provided in the three main parts of the Swiss ELP model serves at least one of the two main functions every ELP has:

- the pedagogic function: motivate to learn (more) languages; help to reflect on language learning and to become a self-directed language learner;
- the documentation and presentation function: document a learner's language and intercultural competencies transparently, reliably, explicitly and in a concrete manner for him-/herself or others.

With respect to their status and their relationship to the Common European Framework it makes sense to divide the instruments into *core elements* and *complementing elements*.

2.1 Core elements

The following instruments may be regarded as core elements of the European Language Portfolio:

- the *Self-assessment grid* in the Language Passport section;
- the *Global scale* in the Language Passport section;
- six detailed *Self-assessment checklists* (one for each level A1 through C2) in the Biography section.

They make up the core of the ELP in the sense that they

- establish explicit links to the CEF;
- enable transparency and comparability;
- link the different parts of the ELP;
- make it possible to compare self-assessment and assessment by others.

The scaled descriptors and their applications are core elements not only of the Swiss model of the European Language Portfolio: the *Self-assessment grid,* in particular, is a key element of many national or transnational ELP versions for adolescent and adult learners, thus providing a common point of reference and enabling comparability. Some other ELPs contain the *Global scale* and/or (often adapted) checklists. The *Global scale* and the *Self-assessment grid* are also part of the CEF (Table 1: *Common Reference Levels: Global scale*; Table 2: *Common Reference Levels: Self assessment grid*). The checklists consist of "I can" statements that are closely related to the *Illustrative scales of descriptors* included in the Common European Framework. Both were drawn from the same bank of calibrated descriptors that was compiled on the basis of empirical research.

As a matter of fact, all the core elements are based on such descriptors that were developed and scaled empirically in a **Swiss National Science Foundation (SNF) project** (cf. North 1996/2000; North/Schneider 1998; Schneider/North 1999a/b; Schneider/North 2000). Most of them describe, in easily understandable language, relevant communication tasks; other descriptors deal with qualitative aspects of the language produced or with communication strategies. Such descriptors make it possible to describe, assess and compare language proficiency in a transparent and coherent manner, even across educational systems and borders. Ever since the ELP has been used in Swiss schools and adult education it was these core elements that teachers were most interested in (cf. Schneider/North 1999: 21). In both evaluation rounds during the official piloting in Switzerland (1999-2000), it became obvious that the *Self-assessment grid* and also the checklists were those elements that had been used in almost every class. The use of other parts (particularly the *Dossier* section) and instruments (e.g. some attestation forms) had often been postponed or left aside. When the teachers were asked whether they found the Portfolio useful in assessing the communicative language competence of their learners, the positive answers dominated strongly: 90% of the teachers answered *yes* in the first questionnaire, around 80% gave a positive answer in the final questionnaire. Other questionnaire items made clear that the teachers based this feedback mainly on experiences made with the core elements. Many of the answers to the question "What is in your opinion the most positive aspect about the European Language Portfolio?" illustrate the prominent position of the core elements. The following point were frequently mentioned (Lenz 1999/2000b):

- the approach of the ELP which describes the communicative language competence of the learners only in positive terms;
- the detailed description of communicative language competence;
- the common language of description;
- relevant and understandable scales and checklists;

- the transparency provided by the descriptions and the six levels;
- the fact that communicative language competence becomes measurable and comparable;
- the checklists which can be used for self-assessment, assessment by others, and the planning of language courses;
- the development of pupils' ability to take responsibility through comparing and discussing self-assessment and teacher assessment.

2.2 Complementing elements

Despite their importance and their appeal, the Swiss ELP model cannot be reduced to the core elements. Other instruments contained in the three sections *Language Passport, Language Learning Biography* and *Dossier* supplement the core elements in important ways. In an attempt to implement assessment in the broad sense described above, and to combine documentation, presentation and pedagogic functions, various instruments were added.

The **Passport section** encourages learners to collect their language certificates and any other diplomas which include a language component. In order to make transparent the type and level of competence attested by these documents, two forms are provided which are destined for the institutions that issue(d) the documents, and which may help, when combined, to place an examination within the categories and levels of the Common Framework: one of them is a form based on the *Global scale*, the other one is an examination description form that helps to describe in detail the format and elements of an examination. Additionally, the Passport section contains several attestation forms:

- Attestation of a school visit abroad
- Attestation of a stay in a foreign language area
- Attestation of hosting a guest from a partner school or institution
- Attestation of a work practice period abroad
- Attestation of participation in a regular exchange of letters
- Attestation of participation in a project
- Attestation of participation in bilingual or immersion courses
- Attestation of participation in an exchange program

Attestations do not have the same status as official diplomas and certificates based on examinations, of course, but they may well serve as "evidence" for certain competencies and underpin e.g. claims made on the basis of self-assessment or in the Language Learning Biography.

The **Language Learning Biography** section intends to enable learners to keep in view and document – for themselves and for others – scholastic and extra-scholastic language learning experiences, intercultural experiences, needs and objectives. Besides self-assessment checklists for the six broad European reference levels, this section contains forms that encourage learners to give a chronological overview of their (plurilingual) language learning biography; to provide detailed information about important intercultural experiences, and activities which have contributed to widening their knowledge of other countries and the people, society and culture of foreign language areas. "Intercultural experiences" is interpreted in a very broad sense: they may include holidays abroad, country-specific cookery

courses as well as literature and art classes in school. Also, ideas are given of how stays abroad can be prepared for and reflected on from an intercultural perspective.

Another form is intended to hold information about foreign language learning in schools and language courses. Although it is part of the ELP which is the property of a learner, it is expected to be filled in by language schools and teachers, or to be substituted with appropriate standard forms issued by schools. Like the forms for the description of exams in the Passport section this course description form is part of an attempt to make transparent – and to influence in a positive way – existing structures with the help of learners. In the instructions to this form the schools are asked to provide information on course objectives and to indicate the means used to reach the objectives. The examples provided to illustrate this point are interesting because most of them belong to innovative forms of language teaching and learning: intensive phases, interdisciplinary projects, exchange and immersion programs.

Finally, the Biography section contains a form that encourages learners to reflect on and write down their language learning objectives more explicitly than might happen just from filling in the "My objectives" column of the checklists.

The third main section of the ELP is the **Dossier**. It is meant as a collection of personal work, of "learner products" which document what one has achieved or is able to do in different languages. These "products" may be written texts, audio or video recordings or more extensive results of project work. Being a collection of actual work, the Dossier is even more concrete and illustrative than proficiency statements made by referring to the descriptors provided.

3. Developing the European Language Portfolio in the Swiss context

Language specialists from various educational sectors in Switzerland – private and public – made important contributions to the development of an ELP ever since the idea of a Language Portfolio (at first *Language Passport)* was born in 1990 and concretised in the surroundings of the Rüschlikon Symposium in 1991 (cf. Council of Europe 1992). One of the conclusions of that Symposium was that "a comprehensive, coherent and transparent framework for the description of language proficiency" (Council of Europe 1992: 39) should be established and that a working party should be set up "to consider possible forms and functions of a "*European Language Portfolio*" to be issued under its aegis and held by individuals, in which they may record their cumulative experience and qualifications in modern languages." (Council of Europe 1992: 40). The general opinion at the Symposium was that the *Common European Framework* should be developed first, and the *European Language Portfolio* afterwards as one of its applications. At the Symposium, the Swiss delegation put forward the idea to conduct a Swiss National Science Foundation (SNF) research project in order to develop a system of descriptions of language proficiency (and possibly descriptions of other aspects like strategic or cultural competence). These descriptions should then be used to develop Portfolio prototypes. Apparently, the combination "research plus Portfolio" appealed to several public and private institutions in Switzerland: they supported the project, and it was launched at the end of 1993. The empirical part was conducted in two phases in 1994 and 1995, involving 2700 learners and 290 teachers from various language regions and from all sectors from lower secondary to adult education. Such a variety of groups were included because the Swiss project group had agreed to focus on intersections in the Swiss educational system, especially the end of compulsory schooling and the end of secondary education

(academic, professional, commercial), and on degrees and diplomas in tertiary and adult education.

Thanks to the project, the ELP was developed much sooner than had been planned in Rüschlikon. But more than that: the SNF project yielded a scientifically sound framework of calibrated descriptors of language proficiency, which gave reason to believe in the feasibility not only of the ELP but also of the description of common reference levels for the CEF.

It is quite interesting to look more closely at the relationship between the developments on the European level and those in Switzerland during that phase. There was more of a *mutual* influence than it may seem from the above. Originally, the Swiss project helped the Council of Europe to proceed with the CEF (and the ELP) project; at the same time the decision by the Swiss authorities to finance the SNF project was highly influenced by its expected usefulness for the language project of the Council of Europe. Later, the encouraging results produced in Switzerland were an important argument for the Council of Europe to pursue the ELP project further. Similar synergies developed among the experts involved: the discussions that took place during the development phase of *Draft 1* of the Common European Framework were an important source of inspiration for the Swiss project leaders. As a result the scales of descriptors that emerged from the project fitted so well into the concept of the CEF that they could be used as an illustrative example in *Draft 1* when it was finally distributed for consultation at the end of 1996.

A **first draft of the Swiss Portfolio model** integrating the results of the ongoing research project was developed at the end of 1995 by the members of the steering group of the project *Cadre de référence pour l'évaluation en langues étrangères en Suisse,* formed after the Rüschlikon Symposium. This group was an interesting combination of people who, together, had a strong background in teaching, research (e.g. SNF project) and educational policy-making at various levels. For this first draft some of the descriptors that were available as provisional results from the SNF project were rephrased so that they could be used for self-assessment. Besides other instruments the *Self-assessment grid* was created at that point. In the months that followed, the Swiss ELP model was elaborated further in English, French and German language versions and then translated into Italian before it was sent to all teachers involved in the research project, together with a request to try it out and to give feedback. The first Swiss ELP model contained basically all the instruments that were published in 1999 for the official piloting. A slightly improved version was distributed before the final conference for all participating teachers, which took place in Fribourg in September 1996. The participants exchanged their experiences and suggested improvements. In spring of 1997 the Swiss ELP and three other draft versions were presented in Strasbourg to an audience with delegates from 40 member states of the Council of Europe. It was then decided to proceed with the ELP project (Council of Europe 1997a/b: ch. II).

The ground for conducting such a project and for producing and field-testing an ELP in Switzerland had been prepared in various ways in the years before Rüschlikon:

- In an attempt to co-ordinate language learning in the 26 Swiss cantons – each of them autonomous in the field of education – and to facilitate mobility within Switzerland, the Standing Conference of the Cantonal Ministers of Education (EDK) recommended, in the mid-eighties, so-called "**points of encounter**" (cf. EDK, 1987) for the intersection at the end of compulsory schooling and the beginning of upper secondary education. These points of encounter described in quite general terms what

pupils should be able to do in speaking, listening, reading and writing at that point. Descriptions of objectives were supplemented by didactic recommendations. The cantons were also encouraged by the EDK to conduct corresponding examinations. Lower and higher levels for other educational intersections were not described as "points of encounter" but the need for such descriptions remained, or rather increased. Particularly the definition of objectives and the creation of exams for new school types like professional upper secondary schools and professional colleges, demonstrated the need for a reference framework.

- In accordance with recent pedagogic trends, projects had been launched in Swiss schools which promoted and examined alternative forms and instruments of assessment and feedback. In French-speaking Switzerland, for example, a Portfolio approach to assessment in Mathematics was piloted. Some of the projects introduced self-assessment.

In the years after the Rüschlikon Symposium in Switzerland, as in other European countries, mobility within and across the borders increased, out-of-school learning became more important, English language curricula were expected to change due to the increasing importance of English as an international language, etc. In the mid-nineties, in response to some of the more recent needs and in an attempt to co-ordinate cantonal initiatives, the Swiss Conference of Cantonal Ministers of Education commissioned a group of experts to elaborate a global concept of language learning during compulsory schooling. In their proposition (EDK, 1998b), the group of experts adopted the *Common reference levels* as well as the Swiss ELP model as central elements of their propositions. The approach they took was quite different from the one taken in the previous concept, the *Points of encounter* (EDK 1987). While the authors of the *Points of encounter* basically had communicative language teaching in mind as a means to reach the common objectives, the authors of the *Global concept* presented a variety of promising paths which might lead to the competencies described in terms of the newly elaborated CEF (and ELP) levels. The paths suggested include immersion programs, exchange programs, out-of-school learning, early language learning, plurilingual language awareness classes, modular and partial language learning etc., and, of course, regular language classes. By recommending such a wide range of learning opportunities, the *Global concept*, despite being written for schools, takes into account the fact mentioned in the Common European Framework that "the educational curriculum is not limited to school and does not end with it" (CEF: 8.4.1). It is obvious that common assessment practice, including commercially available language certificate examinations, would not mirror in a satisfactory way the competencies acquired when following such paths. The ELP, however, based on a comprehensive framework of reference, containing a widely varied toolbox of assessment instruments and ideas, and providing room for a structured collection of a multitude of "evidence" of its holder's language competences and intercultural experiences, can be expected to be much more suitable for such an open environment. It would also be a means to overcome the "compartmentalisation" still common in language teaching and testing, and document the *plurilingual* competence of its holder.

The relationship between the *Global concept* and the European Language Portfolio seems to be synergetic. While on the one hand the *Global concept* is complemented rather neatly by the ELP, the ELP on the other hand may depend for a successful implementation in the school systems on the *Global concept* and the changes it is expected to trigger and sustain. Evidence gathered during the ELP pilot phase hints in that direction:

- many teachers were reluctant because they doubted that the ELP would find enough political support for wide recognition. They realised how much the approach taken might change their teaching but, at the same time, they did not feel the need or did not dare to carry out the necessary changes without a previous political decision;
- decision-makers were heard to say that they would only support the introduction of the ELP if the *Global concept* were put into action at the same time.

No matter how much the European Language Portfolio will depend on supporting measures in the future, it is quite obvious that in the case of the *Global concept* the ELP has played the role of a catalyst: the ELP exists – it is tangible – its core elements provide a framework for planning and assessment, and it has been tested in practice in most cantons thereby spreading many of the ideas contained in the experts‘ proposition of a global concept for language learning and, of course, in the Common European Framework.

The definition of objectives for the end of the upper secondary level is beyond the focus of the *Global concept.* However, the European framework of levels and the ELP have been concretely used to define examination profiles and to develop exams for the newly established “Matura” types for professional upper secondary schools, especially in the Canton of Ticino. The Standing Conference of Ministers of Education also commissioned and published a *Dossier* (EDK 1998a) containing propositions and examples of such examinations. The generally more innovative character of language exams in the professionally oriented upper secondary schools means a permanent challenge for the long-established grammar schools and has provoked fruitful discussions and initiated some changes already.

4. Developing descriptors, grids, scales and checklists

The main goal of the Swiss National Science Foundation project was to develop a bank of meaningful stand-alone descriptors of relevant aspects of language proficiency with known difficulty values (for a detailed description see North/Schneider (1998) and Schneider/North (2000). The statistical methodology used to assign each descriptor a difficulty value and to develop a scale of descriptors was a Rasch analysis (item response theory analysis). See CEF (Appendix A) for a discussion of development methodologies.

Carrying through the SNF project involved the following steps:

- collection of proficiency descriptors from existing scales; breaking up complex descriptors; creation of new descriptors as necessary for some categories (e.g. communication strategies); organizing the descriptors in separate databases for interaction (oral/written), production (oral/written), listening comprehension and reading comprehension;
- qualitative validation of the descriptors in workshops for teachers from different language regions and educational sectors. The purpose was to find out:
  - which descriptors are easily understood and interpreted similarly?
  - which descriptors are suitable for learner self-assessment?
  - which descriptors are deemed relevant in order to describe language proficiency?
  - what types of descriptors are preferred or rejected?
  - which translations are not equivalent?
- data collection: teachers assess learners in their classes and on video tape using descriptors that had emerged from the workshops; in order to have external control,

some learners were asked to take tests; in view of the preparation of the ELP, some learners assessed themselves using "I can" statements;

- statistical analysis using the program FACETS (by Linacre) in order to calculate item difficulty, learner performance and teacher severity;
- identification of "cut-off points" on the scale of calibrated descriptors into levels: ten clearly distinct levels could be determined; this ten-level scale was reduced to six broader levels for its use in the CEF;
- creation of grids, scales and checklists:

- the *Global scale* and the *Self-assessment grid* were constructed using the most typical and stable descriptors; they both were integrated into the CEF (ch. 3.2) and the Swiss ELP drafts as descriptions of common reference levels;
- checklists for learner self-assessment were created drawing on the descriptor database which is structured according to categories of communicative competence and difficulty. The descriptors had to be converted into "I can" statements; some of them had to be simplified in wording. Since the research project focused on interaction and oral communication, there were not enough scaled descriptors available for written production to include in the checklists.

In order to develop more descriptors for Writing, the time-consuming and expensive procedure described above could not be repeated. More informal methods were chosen. The same procedure applied to descriptors for extended reading (e.g. of works of literature) which many teachers had asked for. The steps taken were:

- on the basis of existing, still unvalidated items from the initial pool, a group of experienced teachers formulated items they felt were relevant;
- against the background of the experience gained in the course of the project, the experts who had conducted the SNF project selected the most promising items for piloting, other items were rephrased, systematic gaps were filled with new items;
- groups of teachers were given different sets of descriptors, each descriptor written on a strip of paper, including already calibrated items; every item was given to at least two groups; the teachers then sorted the descriptors into levels; they also noted down comments explaining problems that occurred, e.g. that a descriptor needed reformulation because of ambiguity;
- ELP authors evaluated the assignments to levels and the comments made; some descriptors were rephrased on the basis of the comments, others were eliminated;
- in a final round, the experts decided on the suitability of the descriptors and took them over into the checklists.

Finding suitable ways to develop reliable descriptors may become crucial in the future because there is an obvious need for more descriptors, most conspicuously for the description of finer (or *narrower*) levels, so that progress can be made more visible, and for the specification of proficiency for professional or academic purposes.

5. Insights from the piloting of the European Language Portfolio in Switzerland

A first trial period for the Swiss ELP model took place in 1996, shortly after its creation. It came to a temporary conclusion with a conference in Fribourg in September 1996 when many of the teachers who had participated in the SNF project were present. The second – official and larger scale – piloting was part of the Council of Europe's piloting project and lasted from spring 1999 to summer 2000. The version that was experimented in the second round

corresponded roughly to the earlier draft versions but had been fine-tuned in many respects and laid out by a graphic designer.

During the trial periods information could be gathered on

- the suitability and quality of the product and its elements, mainly the items from the SNF project and from the *Illustrative scales* in the CEF, now adapted for self-assessment;
- successful ways of integrating the ELP into teaching and learning;
- the kind of supporting materials that should be developed for the various types of users etc.

Almost from the very first months of its existence, the Swiss ELP model had a strong influence on ongoing developments in the field of foreign language teaching and learning. It actually did play the role as a catalyst it was expected to play by its authors. Whenever new curricula were developed, or new diploma or entry levels were defined, the ELP played a key role. Apparently, it was used in some cases as a substitute for the second draft version of the Common European Framework, which appears not to be so easily accessible for users.

5.1 Early piloting

When in 1996 the first draft version was distributed to all teachers who had participated in the SNF project, it was accompanied by a feedback questionnaire. The feedback that came in was informal and sparse but mostly positive and helpful.

Some important positive experiences were gathered at university language centres and in teacher training courses:

Example 1:

It soon became quite clear that learners in higher education are able to assess themselves successfully using the instruments provided, after only a short practical introduction to the approach taken and to the system of levels. A successful introduction of the European Language Portfolio into language classes at that level typically involved the following steps:

1. presentation of the ELP, its context, its approach and its uses;
2. getting acquainted with the system of levels: small groups of students reconstruct the *Global scale* (CEF: Table 1) cut into six pieces; students justify their solutions pointing out the relevant elements of the descriptors;
3. provisional self-assessment using the *Self-assessment grid* (CEF: Table 2): students find the description for each skill that corresponds best to their own proficiency; if necessary: clarifications and explanations concerning the descriptions.
4. verification and concretisation of the provisional levels using the appropriate *checklist(s)* containing mainly descriptions of tasks that are typically mastered at that level.

Since the introduction of the ELP to the students usually took place at the beginning of a term, another step followed:

5. Setting personal objectives using the "My objectives" column of the appropriate checklist(s). Afterwards the checklists were normally turned in to the teacher in order to inform him or her about the proficiency level of the group and their objectives.

The fact that learners ideally own ELPs printed in their mother tongue sometimes posed problems – not only in international classes. Many teachers were confronted with negative reactions from learners who did not appreciate the extensive use of their mother tongue in a foreign language class. One way of coping with that problem was to discuss the usefulness of self-reflective activities. A simple way of by-passing this issue was to use copies of the ELP instruments in the target language. The experience showed that learners above level B1 mostly preferred the target language versions. That way, self-assessment was combined with language work which took more time but apparently made more sense to the learners.

The *Self-assessment grid* and checklists were often used again towards the end of the term in order to determine progress. Given the breadth of the levels, it is not surprising that many learners did not generally, for every skill, reach higher levels. It even happened that learners placed themselves on lower levels, one reason being a more critical attitude developed during the term. At this final stage when teachers knew their students well enough, learner self-assessment was usually complemented by teacher assessment, and then compared. The results generally corresponded well. At the Fribourg conference in 1996, this feeling was shared by a majority of the teachers present. In 1999, the teachers involved in the official piloting also reported a high degree of agreement. This positive result could not in any case be expected since self-assessment does not seem to be general practice in Swiss language classrooms. Information from learners in 1995 and from teachers in 1999 lead to the conclusion that at least two thirds of all language learners are not used to assessing their language skills. In 1995 only 3% of the learners involved said that they regularly got a chance to self-assess their skills.

Example 2:

In fall 1997 the German-as-a-foreign-language Unit of the University of Fribourg tried to find out whether the self-assessment instruments from the ELP could be used as an entrance "test" for new students. All learners except beginners were asked to assess their language proficiency on the basis of the core elements of the ELP, following the steps just described in Example 1. The situation was not quite authentic in so far as they had previously been assigned to their classes with the help of a well-tried test incorporating listening, structures and vocabulary, writing and speaking. So, the students knew and were told that the results of their self-assessment were not in any way eliminating. A comparison between the traditional entrance test and the learners' self-assessment was encouraging. An overall correlation of 0.8 between the two instruments was achieved. When the same students were asked to take the *Item banker* computer-adaptive test (North 2000b), roughly the same correlation was attained. The teachers involved had the impression that the results obtained using self-assessment might allow the substitution of the conventional entrance test with this kind of self-assessment procedure, provided the students would behave in the same manner once they knew that their own assessment would decide the classes they could take. Deviations in both directions would have to be expected.

Example 3:

Experiences with teachers showed that it might be useful to include additional instruments in the ELP which are primarily directed at the teachers. The existing Swiss ELP versions so far contained mainly instruments destined for use by learners. However, some forms, including the checklists, course and exam descriptions and basically all attestations imply that the learners should involve other people, mainly teachers and teaching institutions. The authors of the ELP did not include instruments originally developed for teachers. The *Oral assessment criteria grid* (Schneider/North 2000: Appendix X; CEF: Table 3: *Common Reference Levels: qualitative aspects of language use*) containing descriptors of qualitative aspects of spoken interaction and production, and its summary, the *Global speaking scale* (Schneider/North 2000: Appendix X), are such instruments. Both were developed using qualitative descriptors from the descriptor bank. Existing calibrated descriptors were rephrased for use by teachers, systematic gaps were filled and formulations were standardised. The five qualitative categories chosen are *range, accuracy, fluency, interaction* and *coherence*. (For a discussion of type and number of categories see CEF: ch. 9.4) These scales were first developed for the 1996 Fribourg conference because the need to introduce teachers to the system of levels and descriptors in a suitable way had emerged. On the one hand teachers felt insecure when they used the descriptors so that they wanted some initial rater training. On the other hand, evidence from the data collection phase of the SNF project showed that while many teachers handled the descriptors well, i.e. consistently and in quite the same way as their colleagues, others obviously handled the criterion-referenced scales like norm-referenced ones when they assessed their students: instead of comparing individual descriptions to their students' proficiency, they spread out their class over a whole range of levels, thus rating their weaker students too low and their better students too high (Schneider/North 2000: 95).

At the Fribourg conference the two newly developed tools were used rather successfully so that the *Oral assessment criteria grid* could be integrated into the second draft of the Common European Framework afterwards. One of the strengths of this grid and scale is that they are concerned with qualitative aspects (e.g. accuracy), which cannot always be self-assessed by learners and therefore represent a teacher-specific complement to self-assessment within the same comprehensive system of levels and competencies.

How was the work with the new tools organized at the Fribourg conference? Three target-language-specific groups of teachers assessed video-taped learners, whose levels of proficiency had been determined during the SNF project (per pair of students: less than 10 min. duration; first monologue then dialogue). The steps involved were:

1. each individual teacher determines roughly one student's level, using the global speaking scale;
2. he/she checks this provisional assessment considering the more detailed criteria on the same and adjacent levels on the grid; writes down (on a form) a level for each category plus a global level;
3. the individual assessments are compared and justified within the group; differences are discussed; finally the group consensus is compared against the level determined in the SNF project;
4. more learners are assessed until the participants are familiar with different levels on the scale.

This procedure developed for and confirmed at the Fribourg conference has been used successfully in teacher training workshops many times since. The participants usually learn remarkably fast. It seems to be an easy way of training teachers to keep the whole system of skills and levels in mind while focussing quickly on the appropriate levels.

Therefore the inclusion of the *Oral assessment criteria grid* and the *Global speaking scale,* as teacher-specific complements to the self-assessment instruments, into future ELP versions would make sense. Another possibility would be to include them in a guide for teachers. It would also be desirable to develop a corresponding writing scale and grid and to include it into the ELP or a guide for teachers as well.

5.2 Results from the official pilot phase

The official ELP piloting in Switzerland between May 1999 and June 2000 involved around 470 classes and was continually evaluated. Two questionnaire surveys (autumn 1999 and summer 2000), as well as discussions and interviews in between, helped to identify both positive and negative aspects of the Swiss ELP model and its uses in class.

Globally, the teachers considered the Swiss ELP model a good and innovative idea. The analysis of the quantitative data showed largely positive results. The following elements or aspects were particularly well received:

a) descriptors, grids and scales (use of positive descriptions; usability for self-assessment and assessment by others; comparability of communicative language competence; appreciation of partial competencies);
b) the high importance of self-assessment;
c) the ELP as a combination of the three parts *Passport*, *Biography* and *Dossier*, each of them having specific functions;
d) the European dimension of the ELP (common standards in Switzerland and Europe facilitating comparisons and mobility);
e) the usefulness of the ELP as a pedagogic instrument supporting reflective language learning and teaching.

The following elements or aspects were considered problematic:

f) the physical appearance (size and volume of the file);
g) a lack of guidance and visual aids that makes the file appear complex and hard to access;
h) a sustainable integration of the ELP into regular teaching and learning (compatibility with regular teaching and evaluation practice; motivation of the learners; time constraints);
i) the unclear status of the ELP during the pilot phase (doubts concerning a Swiss- and Europe-wide dissemination; doubts about its future value on the labour market; uncertainty whether curricula, final exams, school policy etc. would be adapted or not);
j) the use of the ELP at the lower secondary level (too demanding for 7th graders and non-academically oriented pupils; missing intermediate levels in the region A1 – B1; limited usefulness as a planning instrument when much is predefined by curriculum and textbook).
(Cf. Lenz 1999; Lenz 2000a/b)

Some of the problems encountered during the piloting were due to the experimental character of the ELP project. Using the European Language Portfolio as a comprehensive assessment tool requires an environment – educational and beyond – that is ready to allow and support the changes it entails. Decisions by the relevant authorities in combination with effective public relations activities are needed to help to create the necessary conditions for the ELP to be implemented successfully. Other problems encountered concerned the ELP itself. Its design and content have to be tailored to its target groups of users in a continued effort.

6. Further development of the European Language Portfolio

On the basis of the results from the piloting, the Swiss authorities decided to continue the ELP project and to prepare for implementation and wider dissemination. Among others the following decisions were taken:

- The European Language Portfolio to be launched officially in Switzerland during the European Year of Languages (2001).
- The existing ELP model to be refurbished in the light of the feedback obtained and the new *Principles and Guidelines* issued by the Council of Europe. On the basis of the results from the piloting the following modifications were envisaged:
- a graphic design that provides more guidance to users of the ELP and facilitates access to the file;
- a physical format that is lighter and less bulky while the A4 format is maintained;
  revision of the checklists: clearer instructions, equal and generally reduced number of items for each of the different skills, elimination of very similar items;
- development of a practical teacher's guide providing practical advice and containing supplementary material like the existing Oral assessment grid;
- the adoption of the common *ELP Passport* which was produced as a result of the Portfolio seminars that accompanied the Europe-wide ELP piloting.
- An ELP version for pupils in the second half of compulsory schooling (10 - 15 years old) to be developed. In this context checklists for one lower and several finer intermediate levels will be compiled.

The development of checklists for specific target groups is one type of activity that will have to be further pursued because there is an apparent need for them. By defining finer levels, shorter term objectives become available, and the learners can see progress more frequently (CEF:. 3.5; North 1999). The descriptor bank built up during the SNF project may be exploited for that purpose because it distinguishes between ten levels from (below) A1 to C2 instead of the six used in the CEF. The division into finer levels can be found in the *Illustrative scales* (CEF:. 3.5) where the passages from one sub-level to the next are marked with a horizontal line. The A and B levels are each split up in two sub-levels (cf. Schneider/North 1999a: 17; for the descriptor bank: Schneider/North 2000: appendices V and VI). However, not all descriptors needed can be drawn from the descriptor bank; others will have to be developed from scratch in order to suit the needs of the various potential users. This is particularly true for descriptions of *more specialized* objectives for learners in specific professional or educational environments and for very young learners.

Originally, the authors of the Swiss ELP model expected the teachers participating in the piloting to take over that task for their specific target groups and to suggest and try out descriptors. Mostly due to a lack of time but also because they were confronted with an admittedly difficult task, the teachers generally did not engage in this type of activity. Rather

exceptionally, a group of language teachers of several Swiss universities met with the intention of creating university-specific descriptors complementing the checklists on the six standard levels. The group chose the following procedure:

1. as a point of departure, UNICERT level descriptions (Eggensperger 1998) were exploited for items;
2. these items were complemented by descriptions of communicative tasks which the teachers present also deemed important;
3. the individual teachers matched the descriptors onto the existing six level scale using mainly the checklists as a reference;
4. items which had been assigned to more than two different levels or judged non-level-specific were eliminated or rephrased using the recommendations in Schneider/North (1999a: 13f.);
5. the new items were added to the checklists and used with students;
6. again, the items were edited and systematic gaps were filled; then the entire checklists were distributed to an even larger group of colleagues, and then discussed.

This experience showed that even for teachers who are used to self-assessment and reflection on domain-specific objectives, it was not easy to develop quasi-calibrated descriptors using such a procedure. Important preparatory work has been done but the effort will have to be continued for this target group as well as for many others.

It can be expected – and it is also desirable – that the development of new checklists for specific groups will not automatically lead to specific ELP versions for each of these groups. Undue proliferation would not make sense because, in principle, a learner owns only one ELP at the same time. Newly developed specific descriptors should be included in existing portfolios in a modular fashion.

Even before the end of the ELP pilot phase in Switzerland, some cantonal authorities took measures to create an environment that is more compatible with the ELP (and the CEF). For example, the cantons of central Switzerland developed a new curriculum for French on the primary and lower secondary levels by using (partly adapted) checklist items to define outcomes and thus making explicit reference to the ELP. It is interesting to see that they took the ELP checklists as a point of departure rather than the Common European Framework. As a consequence, there is a certain tendency in that curriculum to overemphasize proficiency in the different language skills in comparison with intercultural or more general pedagogic objectives.

Also before the ELP piloting was over, the cantons of the German-speaking part of Switzerland started to prepare a fairly large assessment project with the main intention to create a coherent assessment system for French and English based on objectives related to the Common European Framework, and fully integrating the future ELP for pupils (see above). In the course of the project, assessment instruments for various purposes will be developed: self-assessment tools, tests and standardized exams. At the same time measures for their implementation will be planned and prepared for: networks of resource persons, teacher training workshops, media-based information and training packages etc. It is clear that by means of such a project conditions may be created for a successful and sustained use of the European Language Portfolio.

7. Conclusion

Even before the European Language Portfolio was officially piloted in Switzerland – and certainly *during* the pilot phase – it started changing things in directions consistent with the Common European Framework. Beyond Switzerland, the Swiss ELP model functioned as an early model and source of inspiration because it was one of the first and most comprehensive ELPs available. It can be expected that through the official launch and wider dissemination the influence of the various ELP models will further grow within and beyond Switzerland, and thereby help the Common European Framework to have an impact in even more contexts.

References

Council of Europe (1992). *Transparency and coherence in language learning in Europe: objectives, assessment and certification. Symposium held in Rüschlikon, 10 - 16 November 1991.* (Edited by North, Brian). Strasbourg: Council for Cultural Co-operation.

Council of Europe (1996a). *Modern languages: learning, teaching, assessment. A common European framework of reference. Draft 2 of a framework proposal.* Strasbourg: Council for Cultural Co-operation (CC-LANG (95)5 rev. IV).

Council of Europe (1996b). *European Language Portfolio. Feasibility study prepared by Christ, Ingeborg / Schärer, Rolf / Debyser, Francis / Dobson, Alan with Kohonen, Viljo / Trim, John L.M. / Sheils, Joseph*. Strasbourg: Council for Cultural Co-operation (= Language Learning for European Citizenship CC-LANG (96)30).

Council of Europe (1997a). *European Language Portfolio. Propositions for elaboration with contributions by Christ, Ingeborg / Debyser, Francis / Dobson, Alan / Schärer, Rolf / Schneider, Günther / North, Brian et al. / Trim, John L.M.* Strasbourg: Council for Cultural Co-operation (= Language Learning for European citizenship CC-LANG (97)1).

Council of Europe (1997b). *Language learning for a new Europe. Report of the final conference of the project "Language Learning for European Citizenship", Strasbourg, 15-18 April 1997*. Strasbourg: Council for Cultural Co-operation (= Language Learning for European citizenship CC-LANG (97) 7).

Council of Europe (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment. Cambridge: CUP.*

EDK - Schweizerische Konferenz der kantonalen Erziehungsdirektoren (ed.) (1987). *Herausforderung Schweiz. Materialien zur Förderung des Unterrichts in den Landessprachen*. Bern: EDK (= Studien und Berichte 2).

EDK - Schweizerische Konferenz der kantonalen Erziehungsdirektoren / Celio, Claudia / Flügel, Christoph & Simon, Wolfgang (1998a). *Die Fremdsprachen in den gymnasialen Maturitätsprüfungen und den Berufsmaturitätsprüfungen.* Bern: EDK (= Dossier 53A). Online: http://edkwww.unibe.ch/dossiers/d53a.pdf

EDK - Schweizerische Konferenz der kantonalen Erziehungsdirektoren (ed.) (1998b). *Welche Sprachen sollen die Schülerinnen und Schüler der Schweiz während der obligatorischen*

Schulzeit lernen? Bericht einer von der Kommission für allgemeine Bildung eingesetzten Expertengruppe "Gesamtsprachenkonzept". Bern. Online: http://www.romsem.unibas.ch./sprachenkonzept/

EDK - Schweizerische Konferenz der kantonalen Erziehungsdirektoren (ed.) (1998c). *Informationsbroschüre Europäisches Sprachenportfolio, Schweizer Version.* Bern: EDK (französische Fassung: *Brochure d'information*; italienische Fassung: *Opusculo informativo*; englische Fassung: *Information brochure*). Online: http://www.unifr.ch/ids/Portfolio

EDK - Schweizerische Konferenz der kantonalen Erziehungsdirektoren (ed.) (1998d). *Forschung und Praxis im Fremdsprachenunterricht - Begegnung, Austausch, Umsetzung. Tagungsbericht Yverdon 1997.* Bern: EDK (= Dossier 52), Online: http://edkwww.unibe.ch/dossiers/d52.pdf

Eggensperger, Karl-Heinz / Fischer, Johann (Hg.) (1998). *Handbuch UNICERT*. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 22).

Lenz, Peter (1999). *Piloting the Swiss version of the European Language Portfolio, May - December 1999. Evaluation report.* (Internal report on behalf of the CDIP and the Council of Europe). [26 p.] Online: http://www.unifr.ch/ids/Portfolio/

Lenz, Peter (2000a). Erfahrungen mit dem Europäischen Sprachenportfolio in der Schweiz. *Babylonia* 1, pp. 23-28.

Lenz, Peter (2000b). *Piloting the Swiss Model of the European Language Portfolio May 1999 – June 2000. Evaluator's Final Report, May 1999 – July 2000.* (Internal report on behalf of the CDIP and the Council of Europe). [54 p.] Online: http://www.unifr.ch/ids/Portfolio/final-report.htm

North, Brian (1996/2000). The development of a common framework scale of descriptors of language proficiency. PhD Thesis. Thames Valley University. Reprinted 2000, New York: Peter Lang.

North, Brian (1996). Description and assessment of foreign language learning proficiency in the Swiss educational system. *VALS/ASLA (Bulletin suisse de linguistique appliquée)*, 64, pp. 129-143.

North, Brian (1997). Perspectives on language proficiency and aspects of competence: a reference paper discussing issues in defining categories and levels. *Language Teaching*, 30, pp. 93-100.

North, Brian (1998). Language proficiency descriptors. In EDK - Schweizerische Konferenz der kantonalen Erziehungsdirektoren (Ed.), *Welche Sprachen sollen die Schülerinnen und Schüler der Schweiz während der obligatorischen Schulzeit lernen?* Bericht einer von der Kommission für allgemeine Bildung eingesetzten Expertengruppe "Gesamtsprachenkonzept", Anhang. Bern. Online: http://www.romsem.unibas.ch./sprachenkonzept/North.htm

North, Brian (1999). The European Common Reference Levels and the Portfolio. *Babylonia* 1, pp. 25-28.

North, Brian (2000). Adaptive testing. *Babylonia* 1, pp. 50-54.

North, Brian & Schneider, Günther (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15, pp. 217-262.
Schneider, Günther (1999a). *Wozu ein Sprachenportfolio? Funktionen und Merkmale des Europäischen Sprachenportfolios (Schweizer Version)*. Online: http://www.unifr.ch/ids/Portfolio.

Schneider, Günther (1999b). Funktionen und wissenschaftliche Grundlagen der Kompetenzbeschreibungen im Sprachenportfolio. *Babylonia* 1, pp. 29-33.

Schneider, Günther (1999c). Evaluation und Selbstevaluation der fremdsprachlichen Kompetenzen: Kompetenzskalen und Sprachenportfolio. In Mercier, Olivier, & Dutoit, Yves (Eds*.), Wie wirksam ist die Berufsbildung in der Schweiz?* Akten der 3. Schweizerischen Berufsbildungstage, Lausanne, 19. - 21. März 1998. Lausanne ISPF (CD-ROM).

Schneider, Günther, & North, Brian (1997b). European Language Portfolio for young people and adults. In Council of Europe (Ed.), *Language Portfolio. Proposals for Development* (pp. 75-88). Strasbourg: Council for Cultural Co-operation.

Schneider, Günther, & North, Brian (1999a). *"Dans d'autres langues, je suis capable de..." Echelles pour la description, l'évaluation et l'auto-évaluation des compétences en langues étrangères*. Berne/Aarau: PNR 33/CSRE (rapport de valorisation).

Schneider, Günther, & North, Brian (1999b). *"In anderen Sprachen kann ich..." Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommmunikationsfähigkeit*. Bern/Aarau: NFP33/SKBF (Umsetzungsbericht).

Schneider, Günther, & North, Brian (2000). Fremdsprachen können – was heisst das? *Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Chur/Zürich: Verlag Rüegger.

Schneider, Günther / North, Brian / Flügel, Christoph & Koch, Leo (1999). *Europäisches Sprachenportfolio - Portfolio européen des langues - Portfolio europeo delle lingue - European Language Portfolio, Schweizer Version*. Ed. EDK - Schweizerische Konferenz der kantonalen Erziehungsdirektoren, Bern: EDK (The Swiss ELP is available in German, French, Italian and English language versions.) Online: http://www.unifr.ch/ids/Portfolio

Contact details:

Prof. Dr. Günther Schneider
Lern- und Forschungszentrum Fremdsprachen
Universität Freiburg
CH-1700 Freiburg
Fax: +41 26 300 97 17
e-mail: guenther.schneider@unifr.ch

Peter Lenz
Lern- und Forschungszentrum Fremdsprachen
Universität Freiburg
CH-1700 Freiburg
Fax: +41 26 300 97 17
e-mail: peter.lenz@unifr.ch

CHAPTER 7: DEVELOPING DESCRIPTOR SCALES OF LANGUAGE PROFICIENCY FOR THE CEF COMMON REFERENCE LEVELS

Brian North, Eurocentres Foundation[2]

Introduction

In this chapter, I describe the origin and development of the scales that appear in the CEF. After initial expert discussions, consultation of the literature and inspection of many proficiency scales, descriptors were developed that were refined in various ways in teacher workshops and by empirical analysis. Reference is made to other projects, also reported in this volume, that have in some sense replicated the work described here.

1. Background

At the intergovernmental Symposium hosted by Switzerland at Rüschlikon, near Zürich in November 1991, the idea of a common European set of language levels was presented in connection with the proposal for a European Language Portfolio (Schärer 1992). North (1992) presented options for what such a common framework scale defining different levels with "Can do" statements might look like. In a panel discussion on the Portfolio, Hargreaves (1992) put forward the idea of "natural levels" represented by the Council of Europe specifications of objectives: *Waystage* (Van Ek and Trim 1991) and *The Threshold Level* (Van Ek and Trim 1990) and other levels used by publishers and examining bodies. The Symposium recommended that the "Common Reference Levels" should be an integral part of the Common European Framework itself, rather than relegated to the Portfolio. The question was how to define them.

At this period, there were a large number of existing sets of defined language levels, sometimes called language proficiency scales, or yardsticks, or guidelines. Such tools were considered to be potentially very useful in contributing to transparency and coherence in language learning, in that they could be used to:

- provide "stereotypes" against which learners can compare their self image and roughly evaluate their position (Trim 1978; Oscarson 1978/9).
- increase the reliability of subjectively judged ratings, providing a common standard and meaning for such judgements (Alderson 1991).
- provide guidelines for test construction (Alderson 1991).
- report results from teacher assessments, scored tests, rated tests and self-assessment all in terms of the same instrument (Alderson 1991).
- co-ordinate pre-course testing, syllabus planning, materials organisation, progress assessment and certification (North 1993a).
- establish an educational framework of reference to describe achievement in terms meaningful to different partners involved (Trim 1978; Richterich and Schneider 1992).
- enable comparison between systems or populations using a common metric (Liskin-Gasparro 1984; Carroll B.J. and West 1989).

[2] The Eurocentres Foundation (www.eurocentres.com) is a Swiss-based not-for-profit educational foundation subsidised by the Migros Federation of Cooperatives. Eurocentres offer language learning stays abroad for 7 languages in regions where the languages concerned are spoken. Eurocentres has been an official consultant to the Council of Europe since 1968, and coordinated the 1991 intergovernmental Symposium which recommended the development of the Common European Framework and the European Language Portfolio.

However, what the language proficiency scales existing in 1992 had in common was that they had been written intuitively either by an expert or by a committee, or had developed in a particular learning context over a period of time. Whilst they could often be claimed to be effective in the contexts for which they were designed, they could equally be claimed to suffer from a fundamental lack of demonstrated validity. Indeed it was rare for any validity information to be provided for scales at all. Where such validity information existed, it was generally limited to inter-rater reliability statistics from assessors using the scale *after* it had been produced. There was usually little or no attempt to demonstrate the validity of the verbal formulation of the descriptors which made up the scale *before* the scale was published. There was little or no attempt to investigate through analysis how statements were actually interpreted. Even more fundamentally, scales tended to present a rather simplified, not to say impoverished and outmoded model of language use, which could not be related to theory (McNamara 1995).

In addition, a number of specific criticisms had been levelled against the genre, among them the following:

- The wording of many scales tended to be relative. One only understood the descriptor for Level 4 when one read it in conjunction with those for Levels 3 and 5.
- Logic was often circular: tasks placed at Level 4 were considered Level 4 because Level 4 people could do them. Learners were considered to be Level 4 because they could perform tasks deemed to be Level 4.
- The descriptors were seldom stand-alone criteria in the sense one could answer "Yes" or "No" to each one (Skehan 1984).
- The descriptors for the lower half of the majority of scales were worded negatively, with the result that while they might be useful in a screening operation, they could not serve as learning objectives (Trim 1978).
- Wording was often word-processed systematically in order to create the semantic appearance of a scale, without actually describing anything significant at all (North 1992).

To put this another way, most scales of language proficiency appeared to have been produced by appeal to intuition, the local pedagogic culture and those scales to which the author had access. In the process of development it was rare that much consideration was given to the following points:

a) using a model of communicative competence and/or language use;
b) checking that the categories and the descriptors made sense to users;
c) using a model of measurement;
d) avoiding the dangers of lifting rank-ordered scale content from one context and then using it inappropriately in another.

As De Jong (1990: 72) summarised, as a result: "the acceptability of these levels, grades and frameworks seems to rely primarily on the authority of the scholars involved in their definition, or on the political status of the bodies that control and promote them."

The question was: would it be possible to produce something better for the Common Framework? Would it be possible to identify categories which could be related on the one hand to theory through the descriptive scheme of the Framework, and on the other hand to categories teachers and learners could relate to? Would it be possible to scale the difficulty of such descriptors for those categories? Would there be a sufficient communality of interpretation in different contexts for such scale values to be considered "common?".

After the Rüschlikon conference, a small group of the presenters considered the feasibility of setting up a framework of descriptors, which would offer an instrument to profile attainment. The group recommended the description of both communicative language activities and aspects of learner competence at succeeding levels (North et al 1992). There was a consensus that the activities might be grouped under the headings Reception, Interaction and Production, which had been used by North (1992) in presenting a mocked-up scale for the Language Portfolio. These categories Reception, Interaction, Production had been developed from Brumfit (1984). There was also a consensus that categories used for aspects of competence needed to be intelligible to teachers. It was thought that practical rating categories might be used, similar to those employed successfully in UK graded objectives schemes (Page 1983) Eurocentres (North 1991), and other communicative assessment schemes related to a notional-functional approach (Morrow 1977; Cambridge/Royal Society of Arts 1990).

However, there was considerable concern about (a) the need to situate categories used in relation to theoretical models, (b) the need to find a methodology which would reduce the subjectivity and cultural specificity of the authorship, (c) the need for a development context which would be multi-lingual and which could cover relevant educational sectors, and above all, (d) the need to build on the – Anglo-Saxon – experience in proficiency scaling without allowing this testing tradition to dictate developments. Metaphors involving maps were common in the discussions: covering different types of territory, experiencing different "slices of life" as opposed to climbing a ladder.

These issues were discussed through 1992 and 1993 without any hurried decisions. It became apparent that Switzerland, because of a series of happy coincidences, might be in a position to provide the development context. Firstly the Swiss educational authorities were prepared to invest in a follow-up to the Rüschlikon Symposium which would assist the development of the European Language Portfolio. Secondly, 1993-5 was the turn of applied linguists and second language researchers to put in bids for research funding from the Swiss Science Research Foundation. Thirdly, Eurocentres, with a very developed framework of proficiency descriptors, was based in Zürich, and the present author was awarded a 3-month Mellon Visiting Fellowship by the US National Foreign Language Center (NFLC) in Washington DC. This opportunity offered the chance to study relevant second language acquisition and measurement literature, find out more about competency description in other disciplines and identify possible development methodologies.

During this interim period, and following the 3-month research stay in the US, North put together three background documents. The first, published as an Occasional Paper by the NFLC (North 1993b) criticised existing scales of language proficiency and put forward a possible methodology to pre-test categories and descriptors, and then to scale descriptors by analysing the way teachers used them when assessing their students. The second document was a detailed analysis and critique of existing scales of language proficiency (North 1994a). The third was a discussion of the relationship between categories in theoretical models and categories which might be scaleable, plus a proposal for how the common reference levels in the Common Framework might be organised (North 1994b / forthcoming).

Meanwhile, the network of partners who had organised the Rüschlikon Symposium put together a teacher network, and Günther Schneider, together with René Richterich, put together a proposal for a Swiss Science Research Foundation (NFP)[3] project based on the methodology North (1993b) had proposed. The main aim of the research project would be to develop and scale transparent and coherent proficiency descriptors for the Framework and Portfolio.

[3] Swiss Science Research Foundation = Nationalforschungs-Program: NFP. The applied linguistics projects in 1993-5 were action 33, hence: *NFP 33: Projekt Schneider & Richterich*)

In September 1993 the first meeting took place of a Working Party charged with developing the Common European Framework (CEF). Following Council of Europe practice, the Working Party consisted of (a) national delegates from countries who had a particular interest in the CEF, (b) invited experts, including the "Swiss" group (Daniel Coste – actually French; Günther Schneider – actually German; Brian North – actually English – and René Richterich) as well as Michael Milanovic from the University of Cambridge Local Examination Syndicate (UCLES), plus (c) the Project Director John Trim and Joe Sheils from the Language Policy Division Secretariat.

The Working Party delegated the actual drafting of the CEF to an authoring group consisting of John Trim, Daniel Coste (by then back at CREDIF in Paris) the present author, and Joe Sheils. Inside the authoring group, the task of defining the Common Reference Levels for the CEF categories was delegated to the author. It was accepted that the results from the Swiss NFP research project would be the basis of this description.

2. The research project

The NFP project ran from November 1993 to November 1996. The project personnel were Günther Schneider, the present author, and a francophone person: first Barbara Tscharner (1993-4), then Leonor Comet (1995), then René Richterich (1995-6). The project produced Common Reference Levels and the illustrative scales in the Framework, and developed a prototype European Language Portfolio. The full project report, mainly in German, is available as Schneider and North (2000). Project results were published more widely in North (1995) and North and Schneider (1998). The theoretical background and the methodology is described in detail in North (1996; 2000).

A pilot project for English conducted in 1994 was focused primarily on spoken interaction and spoken production (extended monologue). Some descriptors were also included for written interaction (letters, questionnaire and form filling) and for written production (report writing, essays etc.). In 1995, data collection was repeated for English, and extended to French and German. Descriptors were also added for reading and for non-interactive listening.

The project took place in four steps in each of the two years 1994 and 1995.

2.1 Comprehensive documentation: Creation of a descriptor pool

The survey of existing scales of language proficiency (North 1994a) provided a starting point. The scales fell mainly into four groups:

- North American and Australian scales derived from or related to the US Foreign Service Institute scale
- Scales from the British sphere of influence derived from or related to the English Language Testing System (ELTS) and its international successor IELTS.
- Scales produced in relation to the proposed unit-credit scheme (Trim 1978) or created by members of the Modern Languages Project network.
- European educational frameworks of objectives and assessment criteria such as the Eurocentres Scale of Language Proficiency, the British National Language Standards, The Royal Society of Arts schemes for French and German, The English National Curriculum, and a draft Dutch framework of levels.

Altogether forty-one proficiency scales were pulled apart with the definition for each level from each scale assigned to a provisional level. Each descriptor was then split up into sentences which were each allocated to a provisional category. This process was undertaken at the same time that the CEF authoring group were creating the descriptive scheme of the CEF. In an interactive process involving the workshops with teachers described below, an effort was made to develop categories of descriptors which related to the descriptive scheme, but which were still intelligible, relevant and user-friendly for teachers. The descriptive scheme draws on the consensus between existing models of communicative competence and language use (e.g. Canale and Swain 1980; Van Ek 1986; Bachman 1990; Skehan 1995 McNamara 1995). In addition, an organisation of language activities under the headings Reception, Interaction and Production was adopted, as mentioned above. Space does not permit detailed consideration of the way in which descriptor categories were related through the descriptive scheme to theoretical models. Readers are referred to North (1994b; 1997; 2000).

Descriptors were produced for various communicative language activities. The following example, taken from a Dutch scale, is a descriptor for *conversation*:

Can enter unprepared into conversations on familiar topics.

Descriptors were also produced for qualitative aspects of language proficiency. The following example, edited from entries in a number of scales, is a descriptor for *fluency*:

Can produce stretches of language with a fairly even tempo; although he/she can be hesitant as he or she searches for patterns and expressions, there are few noticeably long pauses.

Since few descriptors for strategies were available, some 80 were written for interaction and production strategies (following Tarone 1980) with others added for reception. The following example, written after Kellerman et al's (1987: 106) concept "analytic strategy", is a descriptor for *compensating*:

Can define the features of something concrete for which he/she can't remember the word.

2.2 Qualitative Validation: Consultation with teachers through workshops

Qualitative validation of the pool of some 2,000 descriptors was undertaken through wide consultation with foreign language teachers representative of the different sectors in the Swiss educational system. These included teachers in the lower secondary, upper secondary, adult education, vocational education and university sectors. They taught levels from absolute beginner to near-native speaker. In each of the two years 16 workshops were run, a total of 32 meetings. The consultation aimed to ensure that the way teachers thought about proficiency was catered for in the descriptor pool and that all the categories used in the pool were relevant and usable for the teachers. In addition, the procedure identified the most effective descriptors. Two main techniques were used:

The first technique was adapted from that reported by Pollitt and Murray (1993/6). Teachers were asked to discuss which of a pair of learners talking to each other on a video was better - and justify their choice. The aim was to elicit the metalanguage

teachers used to talk about qualitative aspects of proficiency and check that these were included in the categories in the descriptor pool. These discussions were recorded, transcribed in note form, analysed and, if something new was being said, formulated into descriptors.

The second technique was based on that used by Smith and Kendall (1963). Pairs of teachers were given a pile of 60-90 descriptors cut up into strips of paper and asked to sort them into 3-4 labelled piles which represented related potential categories of description. At least two, generally four and up to ten pairs of teachers sorted each set of descriptors. A discard pile was provided for descriptors for which the teachers couldn't decide the category, or found unclear or unhelpful. In addition teachers were asked to indicate which descriptors they found particularly clear and useful and which were relevant to their particular sector. At some workshops, descriptors were also sorted by level, rather than by category. All data from the sorting tasks was coded in a detailed item history for each descriptor.

2.3 Quantitative Validation: Main data collection & Rasch scaling

2.3.1 Data Collection Instruments

A selection of the best descriptors was then used in a questionnaire survey. Those descriptors which had consistently been sorted correctly by category and by level, and which had been consistently stated to be clear and useful, were used for this purpose. These questionnaires were used in two kinds of assessment:

a) Teachers' assessment of the proficiency of 10 learners in their classes using questionnaires made up of 50 descriptors. This was a genuine assessment at the end of the school year, at the same time that the teachers were preparing their end of year grades. Adjacent questionnaires shared a certain number of descriptors. It was the analysis of this data which was used to establish the difficulty values of each of the descriptors and to create a common scale. Technical details are available in North (1996; 2000).

b) Teachers' assessment of video performances of selected learners in the survey using “mini questionnaires”. All teachers rated all videos for the language in question within two weeks of completing their class questionnaire(s). The video performances were structured to a standard format including monologue description and spontaneous interaction. The video “mini-questionnaires” used a small selection of descriptors drawn from the main questionnaires for the level concerned. The analysis of this data was used to establish the severity value of each of the teachers in order to get an objectified overview of the achievement of Swiss learners in relation to the common scale (Linacre 1989). Technical details are available in North (1996; 2000).

2.3.2 Subjects

Exactly 100 teachers took part in the English pilot in 1994, most rating 5 learners from two different classes (total 945 learners). In the second year 192 teachers (81 French teachers, 65 German teachers, 46 English teachers) each rated 10 learners, most rating 10 learners from the same class. Thus data on some 2,500 learners was collected and

analysed. In each year about a quarter of the teachers were teaching their mother tongue, and the main educational sectors were represented as follows:

	Lower Secondary	Upper Secondary	Vocational	Adult
1994	35%	19%	15%	31%
1995	24%	31%	17%	28%

Table 1: Subjects in the data collection for Rasch scaling

2.3.3 Analysis

The analysis method was an adaptation of an item-banking technique in which a series of tests (here questionnaires) are linked by common items called "anchor items" in order to create a common item scale. In this project learners were rated between 0 and 4 on each of the 50 descriptors on the questionnaire. This data was analysed with the Rasch "Rating Scale Model"[4] (Wright and Masters 1982), in a variant taking rater severity into account (Linacre 1989), as operationalised by the program *FACETS*. Details of problems which occurred and solutions which were adopted in the analysis are available in North (1995; 1996/2000).

During this phase of scale construction, a classic technique was used to check the stability of the difficulty estimates for the common items being used to anchor the separate questionnaires into a common scale (Wright and Stone's 1979: 92-96). The same technique was also used to check the stability of the difficulty values of the descriptors as interpreted in relation to different sub-groups of the population.

In the 1994 analysis for English, there was very little variation in the interpretation of difficulty, and that which occurred was easy to explain.

- Adults were judged to find it significantly easier than lower secondary learners to *use basic greetings, to greet and introduce people.* This is not particularly surprising considering their experience of life.
- Adults were judged to find it significantly easier to *use gesture to clarify what they want to say*, *use a dictionary,* and *ask how to say something in the foreign language*. It is not surprising that teenagers were considered by their teachers to use such strategies less naturally than adults. These strategies represent ways of coping despite linguistic limitations with real life tasks, with which secondary learners may not have had experience.
- Gymnasium students were judged to find it significantly easier than adults to: *relate the plot of a book and give their reactions*; *give clear detailed descriptions of complex subjects; write reliable reports on extended spoken or written information.* This has a certain logic. All these tasks involve what Canale (1984) has described as autonomous as opposed to communicative competence, a rather academic ability to produce sustained, coherent, logically structured discourse.

[4] The Rasch Model (Wright and Stone 1979) operationalises "Item Response Theory" (IRT), a branch of "Latent Trait Theory". Latent trait theory is so named because the methodology scales the trait latent within a set of test scores. Item Response Theory (IRT) is so named because, unlike classical test theory (CTT) it is the responses to individual items which are analysed rather than the test as a whole. With Rasch "items are calibrated and persons are measured on a common interval scale" (Wright 1988:286). In other words the Rasch model allows one to calibrate items onto the same measurement scale - onto the same meter. Woods and Baker (1985) and Baker (1997) give good introductions to the Rasch model.

In 1995, some of the variation appeared to reflect curriculum effects; some appeared to reflect the life experience of the learners concerned. Among the trends identified were:

- Vocational education students were judged to find it more difficult to *understand the main information content of most talks and short lectures on familiar topics* and to *read literary texts, including contemporary prose, poetry, and theatrical writing.* Gymnasium students were judged to find found both these academic activities particularly easy.
- Gymnasium students were also judged to find it easier than learners in vocational education, who in turn were judged to find it easier than adults, to: *summarise the plot of a book; and recognise the line of argument in the treatment of the issue presented in written text.* Again this is not surprising. This is the autonomous competence at which it was assumed in 1994 that Gymnasium learners tended to be better.
- Adults were judged to find it easier than learners in vocational education, who found it easier than Gymnasium students, to: *make short contributions in conversation (at elementary);* and *to produce stretches of language at a fairly even tempo*. Adults were also judged to find it easier to *interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain to either party.* The main effect is a suggestion that fluency comes easier to adults.
- Adults were also judged to find it easier to *understand radio news, to extract specific information from announcements and to understand (spoken) complex technical information.* This sounds like a curriculum effect concerning practical listening comprehension.

Whilst for technical reasons one should be cautious in attributing statistical significance to these differences, a possible *trend* for learners in certain sectors to be relatively stronger in certain aspects of proficiency is not surprising. The existence of a certain amount of explicable variation in the difficulty of the tasks described for different sub-groups is not in itself a reason to reject the items concerned. Statements describing such tasks offer an opportunity to profile such strengths and weaknesses. However, when constructing summary descriptions of levels, it is better to focus on descriptors which are consistently interpreted the same way in different contexts. To facilitate this, a "quality hierarchy" was established in the bank of descriptors. Items with a more stable interpretation were used when constructing the more holistic summaries used to present the Common Reference Levels in Tables1 and 2 found in Chapter 3 of the Framework.

Certain categories of descriptors proved to be troublesome in the analysis, or in technical terms to "misfit." Misfit is caused by responses which appear inconsistent in relation to the pattern of the main body of responses. Removal of such "outliers" clarifies the picture and increases the reliability and the precision of the measurement. Four groups of items showed noticeable "misfit":

a) Socio-cultural competence.
It is not clear how much this problem was caused (a) by the concept being a quite separate construct from language proficiency and hence not "fitting"; (b) by rather vague descriptors identified as problematic in the workshops, or (c) by inconsistent responses by the teachers.

b) *Work-related*
Those descriptors asking teachers to guess about activities (generally work-related) beyond their direct experience: Telephoning; attending formal meetings; giving formal presentations; writing reports & essays; formal correspondence.

c) *Negative Concept*
Those descriptors relating to dependence on / independence from interlocutor accommodation (need for simplification; need to get repetition/clarification). These aspects worked better as provisos in positively worded statements, for example:

Can generally understand clear, standard speech on familiar matters directed at him/her, provided he/she can ask for repetition or reformulation from time to time.

Whilst it was regrettable to lose three interesting categories of items from the descriptor bank, the decision had to be made to abandon them in order to safeguard the accuracy and hence usefulness of the main results.

2.3.4 Interpretation

The product of the statistical analysis was a vertical scale of descriptors, rather like a meter which has sentences attached to each millimetre of the ruler. The next question was: how does this vertical list of statements relate to levels? Each descriptor has a "score" on the scale, but how do these "scores" relate to levels? At what "score" does one level (e.g. A1) stop and the next level (e.g. A2) start? That point is called a "cut-off point" or a "cut-score." Setting cut-offs is always a subjective decision. As Wright & Grosse (1993:316) put it: "No measuring system can decide for us at what point *short* becomes *tall*." However, one can take a systematic approach to the problem.

The first question was how many levels could be distinguished in the data. As Pollitt (1991:90) shows there is a relationship between the reliability of a set of data and the number of levels it will bear. In this case a scale reliability of 0.97 justified 9 or10 levels. The first step was therefore to set provisional cut-offs at approximately equal intervals. The second step was to fine-tune these cut-offs in relation to descriptor wording in case there were "threshold effects" between levels. In some cases there was a gap on the vertical scale in approximately the right place. In other cases there was no gap on the scale, but there appeared to be a radical, qualitative shift between the levels implied by two adjacent descriptors. These two (gap; shift) were both examples of a "threshold": a place where the nature of the proficiency changed. In other words, progression up the scale appeared to be in the form of a series of stepped up plateaux rather than a simple continuum.

The procedures described above produced a set of levels of approximately equal intervals, with the switches between levels making apparent sense. The next phase of interpretation involved (a) looking at how the positioning of descriptors on the scale reflected the intentions of the authors of the proficiency scales from which those descriptors had been edited, and (b) identifying which levels corresponded to concepts generally associated with *Waystage* and *Threshold*. The author was helped in this intuitive process by several years' experience of developing the Eurocentres' scale, and tracking student progress data in order to situate examinations onto it.

The end product was the following set of bands on the mathematical scale. The bottom half of the scale has negative values because the kind of scale produced by this analysis, a "logit scale," is a standardised scale of logarithmic units.

Levels	Cut-off points
A1	-4.29
A2	-3.23
A2+	-2.21
B1	-1.23
B1+	-0.26
B2	0.72
B2+	1.73
C1	2.80
C2	3.90

Table 2: Cut-off points between Levels

The finer bands are marked out on the difficulty scale at almost equal intervals of approximately 1 logit (0.97 in the middle; 1.10 at the ends). Notice the symmetry, or normal distribution. The central part of the scale discriminates more finely than the two ends. Intermediate bands exist between the criterion levels A2, B1, B2 and C1. At the time, some commentators thought that factors in the data set or in the analysis may have been responsible for the fact that the difference between A1 and A2, or between C1 and C2, appeared to be only half as great as the differences between the other levels. However, recent research by ALTE (Association of Language Testers in Europe) in relation to their "Can do" statements appears to confirm that the space between C1 and C2 is in fact half that between A2 and B1; B1 and B2, or B2 and C1.

The existence of these "plus" bands between the main levels, however, led to considerable discussion in the Framework Working Party over whether one should stay with the widely accepted 6 levels, or adopt a 9 band scale. Part of the argument was that fewer, broader bands might lead to clearer distinctions in the criterion descriptions presented in Tables 1, 2 and 3 in Chapter 3 of the CEF. This could in turn be expected to produce higher reliability in the calibration of qualifications, and in the accuracy of self-assessments. Narrower bands, on the other hand, are attractive from an educational point of view because they enable a learner to see progress. Yet, the problem is that whereas the broader "natural" levels are common to most systems, narrower bands might prove a complication since educational institutions divide the broader levels in different ways. In the event the decision was taken to focus on the widely accepted levels as criterion levels, but to offer the added detail of the "plus bands" for users to exploit if they wished to do so. Thus in the illustrative descriptors distributed in Chapters 4 and 5 of the Framework document, descriptors representing a "plus band" (a strong performance above the criterion for that level, which does not reach the next criterion level) are placed above a horizontal line drawn across the entry for the level concerned – see Table 3 below.

	INFORMAL DISCUSSION (WITH FRIENDS)
C2	*As C1*
C1	Can easily follow and contribute to complex interactions between third parties in group discussion even on abstract, complex unfamiliar topics.
	Can keep up with an animated discussion between native speakers Can express his/her ideas and opinions with precision, present and respond to complex lines of argument convincingly.
B2	Can take an active part in informal discussion in familiar contexts, commenting, putting point of view clearly, evaluating alternative proposals and making and responding to hypotheses. Can with some effort catch much of what is said around him/her in discussion, but may find it difficult to participate effectively in discussion with several native speakers who do not modify their language in any way. Can account for and sustain his/her opinions in discussion by providing relevant explanations, arguments and comments.
	Can follow much of what is said around him/her on general topics provided interlocutors avoid very idiomatic usage and articulate clearly. Can express his/her thoughts about abstract or cultural topics such as music, films. Can explain why something is a problem. Can give brief comments on the views of others. Can compare and contrast alternatives, discussing what to do, where to go, who or which to choose etc.
B1	Can generally follow the main points in an informal discussion with friends provided speech is clearly articulated in standard dialect. Can give or seek personal views and opinions in discussing topics of interest. Can make his/her opinions and reactions understood as regards solutions to problems or practical questions of where to go, what to do, how to organise an event (e.g. an outing). Can express belief, opinion, agreement and disagreement politely.
	Can generally identify the topic of discussion around her which is conducted slowly and clearly. Can discuss what to do in the evening, at the weekend. Can make and respond to suggestions. Can agree and disagree with others.
A2	Can discuss everyday practical issues in a simple way when addressed clearly, slowly and directly. Can discuss what to do, where to go and make arrangements to meet.
A1	*No descriptor available*

Table 3: CEF scale for Informal Discussion

Sometimes the difference between content scaled at the criterion level and at the "plus level" is not very marked. At other times it appears quite significant. For example, in the scale for Informal Discussion reproduced in Table 3, the descriptors concerning following the discussion of the other speakers show a regular progression through each band:

- identifying the topic (A2+)
- generally following the main points - given clear standard articulation (B1)
- following much of what is said – given avoidance of idiomatic usage (B1+)
- catching much of what is said (same as B1+), - but still find it difficult to get into the discussion (B2)
- keeping up – even when animated (B2+)
- easily following and contributing (C1)

Once the question of the number of bands and the cut-off points between them had been settled, the next step in the interpretation was to compare the placement of descriptors on the CEF scale with their level on the scale from which they originally came. Despite editing, the

majority of the descriptors could still be traced back to their original source. Space does not permit a discussion of the analysis of the relationship to all the source scales. Descriptors from the Eurocentres scale had survived well in the qualitative and quantitative analysis, probably because of their user-friendly positively worded "can do" formulation, so the descriptors from that scale are used as an example. 73 of the 212 descriptors for spoken interaction and production calibrated in 1994 originated from the Eurocentres scale. Therefore it is interesting to plot the position of descriptors shared by the two scales, as in the figure below. This relationship has a correlation of **0.884**

CEF Levels											
C2											
C1								2	1	
B2+							2	1	1	
B2						2	1			
B1+				1	1	1	1		1	
B1		1	2	5	7	1				
A2+	1	8	6	1	1					
A2	8	3	6	1	1					
A1	6									
	1	2	3	4	5	6	7	8	9	10

Eurocentres Levels

Figure 1: Descriptors from the Eurocentres Scale

The shaded, boxed area represents the relationship thought to exist between the two scales. As can be seen there is a very definite pattern. That eight Eurocentres Level 1 descriptors are calibrated at A2 (*Waystage*) rather than A1 (*Breakthrough*) is almost certainly caused by the fact that, as mentioned above, the kinds of simple real-life tasks referred to will tend to be considered easier in the acquisition-rich environment of a stay abroad at a Eurocentres school. In addition, the CEF values at A1(*Breakthrough*) are mainly based on ratings of lower secondary school learners, and, as pointed out whilst discussing the analysis, there was a tendency for these items to be rated as harder for teenagers than for adults. The four items at Eurocentres Levels 4–5 (*Threshold*) placed at A2 (*Waystage*) and A2+ (*Waystage Plus*) are a group of description tasks judged in the study to be simpler than had been intended by scale writers. These descriptors talk about describing one's background, work environment etc. Although most proficiency scales place such tasks at *Threshold*, they are actually among the first themes about which users with a modest language level swap information when they meet on holiday. The plausibility of the calibrations for *Describing* was investigated in some detail in the study, and confirmed. The placement which is really odd is the Level 9 item calibrated at B1+ (*Threshold Plus*). This was the following descriptor edited from Eurocentres Level 9 and from entries at that level on scales from the American FSI (Foreign Service Institute: Wilds 1975) family of scales:

Pronunciation is clear and intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.

This description is likely to be true of a very wide range of level, including a very advanced level. What the Swiss teachers were saying is that it started to be true of their learners just below Level B2. This highlights the fact that it is notoriously difficult to define fine distinctions of level for pronunciation.

The final step in the interpretation phase was to investigate in detail the content coherence of the entire set of descriptors. Was there a logical progression in all the categories scaled? Were there any contradictions between content scaled in one descriptor or category and content scaled in another? Were elements which occurred in descriptors (e.g. topics mentioned) scaled equally coherently as the descriptors themselves? To assist in this final analysis, all the descriptors were broken up into elements, and the elements were presented in charts. These charts enabled a visual check on (a) the logic of progression in each category and (b) the consistency with which elements like for example "everyday situations of a concrete type" appeared in the same band on tables for different descriptive categories. To give an example, the items in the sub-scale for Listening in Interaction were split into the elements "Setting," "Speech" and "Help" as shown in Table 4.

A very clear progression is visible in all three columns. Speech must at first be *very clear, slow, carefully articulated repeated speech directed at* the recipient. Then comes *clear, slow, standard speech, directed at him/her* followed by *clearly articulated standard speech* (which no longer has to be especially slowly or carefully adjusted for him/her) and finally the *standard spoken language*.

Listening in Interaction: Calibrated Elements

Level	Setting	Speech	Help
B2+	-animated conversation between native speakers		
B2	-even noisy environments	-standard spoken language	
B1+	(topics which are familiar)	*As Threshold*	-none; extrapolate unknown words; deduce meaning
B1	-extended everyday conversation	-clearly articulated standard speech	
A2+	-simple, routine exchanges -familiar matters	*As Waystage*	-ask for repetition & reformulation
A2	-simple everyday conversation	-clear, slow, standard, directed at him	-if partner will take the trouble
A1	-everyday expressions aimed at the satisfaction of needs of a concrete type -short, simple questions & instructions	-very clear, slow, carefully articulated repeated speech directed at him	-sympathetic partner -long pauses to assimilate meaning

Table 4: Listening in Interaction: Calibrated Elements

Where categories appeared to have related elements they were juxtaposed on the same page to check consistency. The coherence shown by the different charts was extremely high, with one single contradiction. This was identified when Grammatical Accuracy and Monitoring & Repair Strategies were juxtaposed. At Level B1+ one was said *not to make mistakes which*

lead to misunderstanding, while at the next level, Level B2, one is said to be able to *correct mistakes which have lead to misunderstanding*—which should presumably not have happened. This may reflect the fact the inevitable mistakes at a low level are less likely to lead to misunderstandings than the occasional bad choice of phrase or wrong tense from a more advanced learner whose language is more likely to be accepted at face value. It is ironic that the descriptor concerned was one of the items formulated from a statement made during the teachers' discussion of aspects of learner performance in the workshops. It was one of the very few teacher formulations which had survived the full qualitative and quantitative validation process.

3. Replication

The multi-lingual, multi-sector context of the project, and the coherence of the scaling gave grounds for cautious optimism that the scale values of the descriptors would be found to be stable in future projects. The aim of the study had been to produce an open-ended set of descriptors with *a priori* validity for some of the categories of the Framework. It was hoped that future projects might further validate and extend the bank of descriptors.

Three projects, each separately reported in this volume, have in effect provided further validation for the scaling reported in this chapter. What is interesting is that all three of these projects have concerned self-assessment rather than (as in the original project) teacher-assessment. They have also used radically different data collection instruments, with different rating scales attached to their descriptors for the learner to do their self-rating (Yes / No; 0-4 etc.). Yet impressive inter-correlations of scale values have been obtained.

DIALANG: Kaftandjieva and Takala report elsewhere in this volume on the validation of descriptors for reading, writing and listening taken from the Framework, and used in self-assessment format by predominantly Swedish-speaking learners to rate their proficiency in Finnish. With the removal of one descriptor, a correlation of **0.899** was achieved between the reported scale values of descriptors in the Swiss and Finnish projects. This is all the more impressive when one considers that the Swiss project reported here focused predominantly on spoken interaction and production. Writing was not scaled empirically. The CEF descriptors for writing were produced by recombining elements scaled in relation to speaking. Reading was scaled as a separate construct, and then the reading scale was equated to the spoken interaction/production scale through the employment of listening items as anchors. The DIALANG validation suggests that the combination of empirical scaling and qualitative analysis which produced the CEF descriptors worked. It is also encouraging that descriptors originally scaled for English and then used successfully for French and German in the second year of the Swiss project, proved to have stable values in relation to a non Indo-European language, Finnish.

Basle University: North reports elsewhere in this volume on the validation of descriptors for listening, reading, speaking, writing, interactive skills and language knowledge taken from the Framework. The descriptors were used in a self-assessment instrument to rate foreign and mother-tongue language proficiency on arrival at university. Coincidentally, the correlation between the scale values produced by the Basle data and the original project for 27 common items was again **0.899**.

ALTE: Jones reports in another chapter in this volume on the equation of the ALTE scale of "Can do" statements to the Common Reference Levels. He used the self-assessment grid

presented as Table 2 in Chapter 3 of the CEF for qualitative comparison, and a set of 16 very stable descriptors of fluency and related categories as scaling anchors. These 16 descriptors met the criteria set by North (2000) for "Excellent Items." This was the top category in the quality hierarchy referred to earlier. Such items are particularly suitable for use as linking items to expand the original bank of CEF descriptors. They encompassed both a narrower, psycholinguistic definition of fluency as automaticity and flow, plus a broader, communicative definition. As well as descriptors from the CEF Fluency scale, CEF descriptors for comprehension in interaction, flexibility, and co-operating strategies were also included. Fluency was suggested as the most suitable category to anchor the two scales for equating purposes for three reasons. Firstly, fluency in Brumfit's (1984) Accuracy/Fluency is an easily understood, generic category very suitable for self-assessment. Secondly, work by Lennon (1990) and Fulcher (1993) demonstrated that communicative fluency is interpreted in a very stable manner, builds a linear scale and correlates very well with overall language proficiency. Thirdly, the scale values for fluency in the Swiss project showed the least variation by educational sector, target language, mother tongue etc. The correlation between the original scale values for the 16 anchor items and their values in the ALTE data for well over 1,000 learners was **0.97**.

4. Conclusion

The aim of the project described here was to develop and scale criterion statements related to the categories of the descriptive scheme of the CEF for the Common Reference Levels. The high degree of inter-correlation between the interpretation of difficulty in teacher judgements (Swiss project) and in self-assessments (DIALANG, Basle, ALTE) and the high degree of content coherence that was being demonstrated suggests that there is at least some consensus, some common framework underlying people's impressions. This consensus could become a basis for common European standards for standards-oriented assessment.

The next logical step is to try and link those impressions to actual test results. This is currently being undertaken in DIALANG and ALTE research. In the ALTE project, on-going research is linking self-assessments on "Can do" statements to actual performance on test tasks from the UCLES suite of exams. In DIALANG, data will be available to compare self-assessment to test results for each candidate. Over a period of time it is possible that using Rasch and newer techniques, researchers will be able to build up a fuller and more accurate consensus description, and feed this into future editions of the CEF.

However, this description will never be complete, or fully accurate, or represent the cutting edge of second language acquisition research. It will always remain just an operational model. Models provide a simplified representation of reality - that is why they are called models. If they didn't do this, they would be too complex to be usable. As Mislevy puts it:

> "A learner's state of competence at a given point in time is a complex constellation of facts and concepts, and the networks that interconnect them; of automatized procedures and conscious heuristics, and their relationships to knowledge patterns that signal their relevance; of perspectives and strategies, and the management capabilities by which the learner focuses his efforts. There is no hope of providing a description of such a state. Neither is there any need to." (Mislevy 1993: 28)

There are those who argue that if this is the case, one should give up the idea of describing stages of attainment. Yet teachers and raters need defined standards as a reference point for the judgements they make in assessment. As Scarino (1997: 249) points out there is a tension

arising from the fact that learning and progress are specific to the individual (and little understood) and yet educational systems need to provide mechanisms for monitoring and continuously improving standards across the system as a whole. Mislevy's response to the problem is to suggest the concept of a student model - "a simplified description of selected aspects of the infinite varieties of skills and knowledge that characterise real students" (Mislevy 1995:343). This is what the CEF descriptors seek to provide.

References

Alderson, J.C. 1991: Bands and scores. In Alderson and North, 71-86.

Alderson, J.C. and North, B. 1991: (eds.): *Language testing in the 1990s:* Modern English Publications/British Council, London, Macmillan.

Bachman, L.F. 1990: *Fundamental considerations in language testing*, Oxford, OUP.

Baker, R. 1997: *Classical Test Theory and Item Response Theory in Test Analysis. Extracts from: An Investigation of the Rasch Model in Its Application to Foreign Language Proficiency Testing*. Language Testing Update. Special Report No 2.

Brumfit, C.J. 1984: *Communicative methodology in language teaching: the roles of fluency and accuracy*. Cambridge: Cambridge University Press.

Canale, M. 1983: On some dimensions of language proficiency. In: Oller, J.W. (ed.) *Issues in language testing research*. Rowley (Mass.), Newbury House: 333–342.

Canale, M. 1984: On some theoretical frameworks for language proficiency. In: Rivera, C. (ed.): *Language proficiency and academic achievement*. Clevedon: Multilingual Matters.

Canale, M. and Swain, M. 1980: Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1-47.

Carroll B.J. and West, R. 1989: *ESU (English-speaking union) framework. Performance scales for English language examinations*. London: Longman.

Council of Europe 1992: *Transparency and coherence in language learning in Europe: objectives, assessment and certification*. Strasbourg, Council of Europe; the proceedings of the intergovernmental Symposium held at Rüschlikon November 1991 (ed. North, B.).

De Jong, H.A.L. 1990: Response to Masters: Linguistic theory and psychometric models, in De Jong,

H.A.L. and Stevenson D.K. *Individualising the assessment of language abilities,* Cleveland, Multilingual Matters, 71-82.

Fulcher, G. 1993: The construction and validation of rating scales for oral tests in English as a foreign language, PhD thesis, University of Lancaster.

Hargreaves, P. 1992: Round table discussion on the European Language Portfolio. In: Council of Europe 1992, 150–158.

Kellerman, E., Bongearts, T. and Poulisse N. 1987: Strategy and system in L2 referential communication. In Ellis R., *Second language acquisition in context.* Englewood Cliffs, Prentice Hall: 100-112.

Lennon, P. 1990: Investigating fluency in EFL: a quantitative approach. *Language Learning* 40, 387-417.

Linacre, J.M. 1989: *Multi-faceted measurement*. Chicago, MESA Press.

Liskin-Gasparro, J.E. 1984: The ACTFL proficiency guidelines: a historical perspective. In Higgs, T.C. (ed.) *Teaching for proficiency, the organising principle*. Lincolnwood (Ill.): National Textbook Company: 11-42.

McNamara, T. 1995: Modelling performance: opening Pandora's box. *Applied Linguists* 16, 159-179.

Mislevy, R. J. 1993: Foundations of a new test theory. In Frederiksen, N., Mislevy, R. J. and Bejar, I. I (eds.) *Test theory for a new generation of tests*. Hillsdale N.J. Lawrence Erlbaum Associates: 19-40

Mislevy, R. J. 1995: Test theory and language learning assessment. *Language Testing* 12, 3: 341-369.

Morrow, K. 1977: *Techniques of evaluation for a notional syllabus*. Royal Society of Arts, London

North, B. 1991: Standardisation of continuous assessment grades. Alderson, J.C. and North, B. 1991: (eds.): *Language testing in the 1990s:* Modern English Publications/British Council, London, Macmillan: 167-177.

North, B. 1992: *European Language Portfolio: Some options for a working approach to design scales of proficiency*. In: Council of Europe 1992, 158–174. Reprinted in Schärer, R. and North, B. 1992: *Towards a common European framework for reporting language competency*. NFLC Occasional Paper, National Foreign Language Center, Washington D.C.

North, B. 1993a: Transparency, coherence and washback in language assessment. In Sajavaara, K., Takala, S., Lambert, D. and Morfit, C. (eds.) 1994: *National Foreign Language policies: practices and prospects*. Institute for Education Research, University of Jyväskyla: 157-193.

North, B. 1993b: *The development of descriptors on scales of proficiency: perspectives, problems, and a possible methodology.* NFLC Occasional Paper, National Foreign Language Center, Washington D.C., April 1993.

North, B. 1994a: *Scales of language proficiency: a survey of some existing systems*, Strasbourg, Council of Europe.

North, B. 1994b / forthcoming: *Perspectives on language proficiency and aspects of competence: a reference paper defining categories and levels*. Strasbourg, Council of Europe CC-LANG (94) 20.

North, B. 1995: The development of a common framework scale of descriptors of language proficiency based on a theory of measurement. *System* 23, 445-465.

North, B. 1996: *The development of a common framework scale of descriptors of language proficiency based on a theory of measurement*. 1996 PhD thesis, Thames Valley University.

North, B. 1997: Perspectives on language proficiency and aspects of competence. *Language Teaching* 30, 93-100.

North, B. 2000: *The development of a common framework scale of language proficiency* New York, Peter Lang.

North B., Page B., Porcher L., Schneider G. and Van Ek J. 1992: A preliminary investigation of the possible outline structure of a common European language framework: Synthesis of the outcome of

the short intensive meeting of invited experts to consider principles, hypotheses and options for further investigation. Council of Europe CC-LANG (92) 12

North, B. and Schneider, G. 1998: Scaling descriptors for language proficiency scales. *Language Testing* 15, 2, 217-262.

Oscarson, M. 1978/9: *Approaches to self-assessment in foreign language learning*. Strasbourg, Council of Europe 1978; Oxford, Pergamon 1979.

Page, B. 1983: Graded Objectives in Modern Language Learning. *Language Teaching* October 1983: 292-305.Pollitt, A. 1991: Response to Alderson: Bands and scores. In Alderson and North: 87-94.
Pollitt, A. and Murray, N.L. 1993/1996: What raters really pay attention to. Paper presented at the 15th Language Testing Research Colloquium, Cambridge and Arnhem, 2-4 August 1993. In Milanovic, M. and Saville, N. (eds.) 1996: *Performance testing, cognition and assessment*. Studies in Language Testing 3. Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem, 2-4 August 1993. Cambridge: University of Cambridge Local Examinations Syndicate, 74-91.

Richterich, R. and Schneider, G. 1992: Transparency and coherence: why and for whom? In Council of Europe: 43-50.

Scarino, A. 1997: Analysing the language of frameworks of outcomes for foreign language learning. In Proceedings of the AFMLTA 11th National Languages Conference: 241-258.

Schärer, R. 1992: A European language portfolio - a possible format. In Council of Europe: 140-146. Reprinted in Schärer, R. and North, B. 1992: *Towards a common European framework for reporting language competency*. NFLC Occasional Paper, National Foreign Language Center, Washington D.C.

Schneider, G. and North, B. 2000: *Fremdsprachen können - was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit.* Chur, Zürich: Verlag Rüegger.

Skehan, P. 1984: Issues in the testing of English for specific purposes. *Language Testing* 1, 202-220.

Skehan, P. 1995: Analysability, accessibility and ability for use. In Cook, G. and Seidlehofer, S. (eds.), *Principle and practice in applied linguistics*. Oxford: Oxford University Press.

Smith, P.C. and Kendall, J.M. 1963: Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology* 47, 149-154.

Tarone, E. 1980: Communication strategies, foreigner talk and repair in interlanguage. *Language Learning* 30, 417-431.

Trim, J.L.M. 1978: *Some possible lines of development of an overall structure for a European unit/credit scheme for foreign language learning by adults*. Strasbourg, Council of Europe.

University of Cambridge/Royal Society of Arts 1990: *Certificates in communicative skills in English Teachers' Guide*.

Van Ek, J.A. 1986: *Objectives for foreign language teaching, volume I: scope*. Strasbourg, Council of Europe.

Van Ek, J. A. and Trim, J.L.M. 1990: *The Threshold Level 1990*. Strasbourg: Council of Europe.

Van Ek, J. A. and Trim, J.L.M. 1991: *Waystage 1990*. Strasbourg: Council of Europe.

Wilds, C.P. 1975: The oral interview test. In Spolsky, B. and Jones, R.: *Testing language proficiency*.

Washington D.C., Center for Applied Linguistics: 29-44.

Woods, A. and Baker, R. 1985: Item response theory. *Language Testing* 2, 117-140.

Wright, B.D. 1988: Rasch measurement models. In: Keeves, J. P. (ed.) 1988: *Educational research, methodology and measurement: an international handbook*. Oxford / New York: Pergamon Press, 286–292.

Wright, B.D. and Grosse, M. 1993: How to set standards, *Rasch Measurement, Transactions of the Rasch Measurement Special Interest Group of the American Educational Research Association* 7/3, 315-6.

Wright, B,D. and Masters, G. 1982: *Rating scale analysis.* Chicago, Mesa Press.

Wright, B.D. and Stone, M.H. 1979: *Best test design*. Chicago, Mesa Press.

CHAPTER 8: COUNCIL OF EUROPE SCALES OF LANGUAGE PROFICIENCY: A VALIDATION STUDY

Felianka Kaftandjieva & Sauli Takala
Centre for Applied Language Studies, University of Jyvaskyla, Finland

Introduction

The purpose of this chapter is to present validity evidence for the Council of Europe scales of language proficiency in Listening, Reading and Writing, as gathered in the DIALANG project. We report on the collection, which took the form of a sorting task, where raters – specialists in teaching Finnish as a second/foreign language – were asked to sort the scale descriptor units for Listening, Reading and Writinginto six successive piles representing the six levels of Council of Europe scales: Breakthrough, Waystage, Threshold, Vantage, Effective-Proficiency and Mastery. We report the statistical results in some detail, in the hope that this will encourage other developers of scales relating to the CEF to report in similar detail the results of their validation studies, and we present a detailed content analysis of some of the descriptors, to illustrate how some of the statistcal data can be interpreted in terms of the wording of the descriptors in the scales.

1. Background

DIALANG (Project 39441-CP-1-96-1-FI-LINGUA-LD), funded by European Commission DG XXII (Education and Culture) under the SOCRATES Programme, LINGUA Action D, aims to develop a self-access Internet-based language assessment system. The system will allow language learners/users to be informed about the strengths and weaknesses in their language proficiency in 14 languages (the eleven Community languages plus Norwegian, Icelandic, and Irish), and this is meant to guide learners in their efforts to improve their language ability.

One of the key features of DIALANG is that the feedback given to the clients about their language proficiency will be expressed in terms of the Council of Europe (CoE) six-level scales of language proficiency. Combining self-assessment tools with objective testing, DIALANG provides an opportunity for a comparison between self-assessed and objectively assessed language proficiency as well as an opportunity for a comparison of language proficiency in different languages.

The use of a common yardstick (CEF scales of language proficiency) is an attractive DIALANG feature, but at the same time, this makes the whole assessment system strongly dependent on the quality of this yardstick. The validation of CEF scales of language proficiency is crucial and it constitutes a starting point in the development and validation of the DIALANG language assessment system, since ‘validation logically begins with an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the interpretation to the proposed use’ (Standards for Educational and Psychological Testing, 1999, p. 9).

The pre-eminence of validity in developing and evaluating assessment tools is generally acknowledged, but it seems that “the very breadth and complexity of the concept makes it

difficult to work with in practice and, therefore, threatens continuing neglect of validity in the monitoring of the quality of the assessment" (Crooks & Kane, 1996, p. 266).

In the field of language testing for example, the increasing use of different scales of language proficiency entails extensive validity studies, the need for which is well recognized (Matthews, 1990; Shohamy, 1995; McNamara, 1996; Brindley, 1997; Butler & Stevens, 1998; McKay, 2000), since without enough validity evidence "there is no guarantee that the description of proficiency offered in a scale is accurate, valid or balanced" (North & Schneider, 1998, p. 219).

The purpose of this chapter is to present the results of a validation study, which was designed and conducted as a part of a pilot study of a standard setting procedure specifically designed for the purposes of DIALANG. For more detail and background on DIALANG, see the chapter in this volume by Huhta et al. "DIALANG – A Diagnostic Language Assessment System for Learners".

2. Method

2.1. Subjects

Two groups of language specialists in Finnish as a second language took part in this study.

The first group (Gr1) consisted of five teachers in Finnish in the upper secondary school, who had no previous experience with the CoE Framework.

The second group (Gr2) included seven specialists in language testing (with a teaching experience in Finnish as a second language), and most of them had some preliminary knowledge and experience with the CoE Framework.

2.2. Materials

The level descriptors of the CEF Scales for Listening, Reading and Writing (Council of Europe, 2001, Table 2 in Chapter 3) in their Finnish translation (Huttunen & Huttunen, 1998) were split into independent meaning units.

For example, the descriptor - *I can understand long and complex factual and literary texts, appreciating distinction of style* (Reading, Level C1) was divided into three units as follows:

- *I can understand long and complex factual texts.*
- *I can understand long and complex literary texts.*
- *I can appreciate distinction of style.*

As can be seen from the example above, the suggested partition is not unique, but only one of the many possible ways of dividing the general description into independent meaningful statements which can be more easily sorted in order of growing language proficiency.

The process of partitioning general level descriptors into independent units led to a list of a total of 64 units (19 for Listening, 20 for Reading and 25 for Writing) which are presented in Table 1 together with some summary statistics.

TABLE 1a. Listening: Descriptor units and summary statistics

CoE level	ID	Descriptor units	CoE level	Rating		
				Median	Range	Scale value
A1	*11*	I can recognise familiar words and very basic phrases concerning myself when people speak slowly and clearly.	1	1	1	–0.67
A1	12	I can recognise familiar words and very basic phrases concerning my family when people speak slowly and clearly.	1	1	1	–0.95
A1	13	I can recognise familiar words and very basic phrases concerning immediate concrete surroundings when people speak slowly and clearly.	1	1	1	–0.44
A2	21	I can understand phrases and normal vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local geography, employment).	**2**	**2**	1	+0.95
A2	22	I can catch the main point in short, clear, simple messages.	**2**	**2**	2	+0.44
A2	23	I can catch the main point in short, clear announcements.	**2**	**2**	1	+0.44
B1	31	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc..	**3**	**3**	1	+2.29
B1	32	I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when I am spoken to relatively slowly and clearly	**3**	**3**	1	+3.26
B2	41	I can understand extended speech provided the topic is reasonably familiar.	**4**	**3**	2	+3.95
B2	42	I can understand lectures provided the topic is reasonably familiar.	**4**	**4**	2	+5.46
B2	43	I can understand and follow even complex lines of argument provided the topic is reasonably familiar.	**4**	**5**	**3**	+6.06
B2	44	I can understand most TV news programmes.	**4**	**4**	**3**	+5.19
B2	45	I can understand most current affairs programmes.	**4**	**5**	2	+5.20
B2	46	I can understand the majority of films in standard dialect.	**4**	**4**	**3**	+5.19
C1	51	I can understand extended speech even when it is not clearly structured.	**5**	**5**	**3**	+5.89
C1	52	I can understand extended speech even when relationships are only implied and not signalled explicitly.	**5**	**6**	**3**	+6.19
C1	53	I can understand television programmes without too much effort	**5**	**4**	2	+5.24
C1	54	I can understand films without too much effort.	**5**	**5**	1	+5.29
C2	61	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.	**6**	**6**	0	+6.51

TABLE 1b. Reading: Descriptor units and summary statistics

CoE level	ID	Descriptor units	CoE level	Rating		
				Median	Range	Scale value
A1	11	I can understand familiar names, for example on notices and posters or in catalogues.	**1**	**1**	0	0.00
A1	12	I can understand familiar words, for example on notices and posters or in catalogues.	**1**	**1**	0	0.00
A1	13	I can understand familiar and very simple sentences, for example on notices and posters or in catalogues.	**1**	**2**	1	+0.44
A2	21	I can read very short, simple texts.	**2**	**2**	1	+0.20
A2	22	I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables.	**2**	**2**	**3**	+0.59
A2	23	I can understand short, simple personal letters.	**2**	**2**	2	+0.72
B1	31	I can understand texts that consist mainly of normal everyday language.	**3**	**2**	2	+1.09
B1	32	I can understand texts that consist mainly of job-related language.	**3**	**3**	2	+2.53
B1	33	I can understand the description of events in personal letters.	**3**	**3**	2	+1.92
B1	34	I can understand the description of feelings in personal letters.	**3**	**3**	1	+2.43
B1	35	I can understand the description of wishes in personal letters.	**3**	**4**	1	+2.63
B2	41	I can read articles concerned with contemporary problems in which the writers adopt particular stances or viewpoints.	**4**	**5**	1	+4.04
B2	42	I can read reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.	**4**	**5**	1	+4.04
B2	43	I can understand contemporary literary prose	**4**	**5**	2	+4.42
C1	51	I can understand long and complex factual texts.	**5**	**5**	2	+5.16
C1	52	I can understand long and complex literary texts.	**5**	**5**	2	+4.68
C1	53	I can appreciate distinctions of style.	**5**	**5**	2	+4.80
C1	54	I can understand specialised articles even when they do not relate to my field.	**5**	**6**	1	+5.51
C1	55	I can understand longer technical instructions, even when they do not relate to my field.	**5**	**5**	2	+4.54
C2	61	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.	**6**	**6**	0	+5.31

TABLE 1c. Writing: Descriptor units and summary statistics

CoE level	ID	Descriptor units	CoE level	Rating		
				Median	Range	Scale value
A1	11	I can write a short simple postcard, for example, sending holiday greetings.	**1**	**1**	1	-1.41
A1	12	I can fill in forms with personal details, for example, entering my name, nationality and address on a hotel registration form.	**1**	**1**	1	-0.95
A2	21	I can write short simple notes relating to matters in areas of immediate need.	**2**	**1**	2	–0.20
A2	22	I can write short simple messages relating to matters in areas of immediate need.	**2**	**2**	1	+0.00
A2	23	I can write a very simple personal letter, for example, thanking someone for something.	**2**	**2**	1	+0.00
B1	31	I can write simple connected text on topics which are familiar or of personal interest.	**3**	**3**	1	+1.65
B1	32	I can write personal letters.	**3**	**3**	1	+1.65
B1	33	I can describe experiences.	**3**	**3**	2	+1.72
B1	34	I can describe impressions.	**3**	**3**	1	+1.56
B2	41	I can write clear detailed text on a wide range of subjects related to my interests	**4**	**4**	2	+2.92
B2	42	I can write an essay passing on information or giving reasons in support of or against a particular point of view.	**4**	**5**	2	+2.96
B2	43	I can write a report passing on information or giving reasons in support of or against a particular point of view.	**4**	**4**	4	+2.88
B2	44	I can write letters highlighting the personal significance of events or experiences.	**4**	**3**	1	+1.80
C1	51	I can express myself in clear well-structured text expressing points of view at some length.	**5**	**4**	3	+2.81
C1	52	I can write detailed expositions of complex subjects in an essay underlining what I consider to be the salient issues.	**5**	**5**	3	+3.43
C1	53	I can write detailed expositions of complex subjects in a letter underlining what I consider to be the salient issues.	**5**	**6**	1	+4.92
C1	54	I can write detailed expositions of complex subjects in a report underlining what I consider to be the salient issues.	**5**	**6**	1	+4.45
C1	55	I can write different kinds of texts in an assured personal style appropriate to the reader in mind.	**5**	**6**	2	+4.85
C2	61	I can write clear smoothly-flowing text in an appropriate style.	**6**	**5**	3	+3.26
C2	62	I can write complex letters.	**6**	**5**	2	+3.53
C2	63	I can write complex reports.	**6**	**6**	1	+4.69
C2	64	I can write complex articles.	**6**	**6**	1	+4.92
C2	65	I can present a case with an effective logical structure, which helps the recipient to notice and remember significant points	**6**	**5**	2	+3.43
C2	66	I can write summaries of professional or literary works.	**6**	**5**	3	+3.58
C2	67	I can write reviews of professional or literary works.	**6**	**5**	2	+3.55

2.3. Procedure

The Common European Framework was first briefly presented to the subjects and its use for teaching and assessing was discussed. Its six levels were introduced only in general terms without specific skill descriptions (Council of Europe, 2001). The language specialists were then asked to sort the CEF scale descriptors for each of the three skills into six ordered categories (the **Method of Successive Intervals** – Torgerson, 1958; Dunn-Rankin, 1983).
In this way, each of the 64 descriptor units was rated by each of the 12 raters on a six-point ordinal scale.

A statistical analysis of the resulting matrices (units by raters) for Listening (19 units x 12 raters), Reading (20 units x 12 raters), and Writing (25 units x 12 raters) was conducted in order to investigate:

- The degree of association and agreement across raters, and between the aggregated rating of descriptor units and their original CEF levels.
- The construct validity of the CEF scales of language proficiency for Listening, Reading and Writing.
- The scalability of the CEF scales of language proficiency.

3. Results and Discussion

3.1. Inter-rater reliability

Inter-rater reliability is an indicator of the degree of consistency and/or agreement across the raters. The analysis of inter-rater reliability is an important part of the current validation study, because "to validate an interpretive inference is to ascertain the degree to which multiple lines of evidence are consonant with the inference" (Messick, 1989, p. 13).

On the other hand, since each of the raters has his/her own perception of the constructs of language proficiency in Listening, Reading and Writing, the analysis of inter-rater reliability can be seen as testing for convergence across different measures of the same constructs which is an integral part of the construct validation process (Cook & Campbell, 1979, p. 61).

There are two main approaches in the analysis of inter-rater reliability, which are based on either association or the degree of agreement between raters. Bearing in mind, however, that there is no single perfect statistical index for the estimation of inter-rater reliability, the best approach will be to compute different indexes and to compare them taking into account that the study conditions (e.g. sample size, number of raters, shape of the distribution, level of measurement, etc.) might affect the different indexes in different degree (Abedi et al., 1995).
The results of different measures of inter-rater reliability and agreement are summarized in Table 2.

TABLE 2. Inter-rater reliability – summary statistics

Indexes (k = 12 raters)	Listening (*n* = 19)			Reading (*n* = 20)			Writing (*n* = 25)		
	Mean	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
r [a]	**0.87***	0.76*	0.96*	**0.90***	0.79*	0.98*	**0.86***	0.70*	0.95*
α [b]	**0.99**			**0.99**			**0.99**		
PRL [c]	**0.99***			**0.99***			**0.98***		
W [d]	**0.87***			**0.91***			**0.86***		
H [e]	**0.78***	0.57*	1*	**0.78***	0.59*	1*	**0.76***	0.54*	0.94*
RAI [f]	**0.89**			**0.90**			**0.89**		

a) *Pearson product moment correlation coefficient –* **r** *(Cronbach, 1990, pp. 161 - 167)*

b) *Cronbach'* **α** *index of internal consistency (Cronbach, 1990, pp. 202 - 204)*

c) *Proportional reduction in loss index (Rust & Cool, 1994) – an extension and generalization of α for qualitative data and a limited number of raters, based on the pairwise exact agreement.*

d) *Kendal's coefficient of concordance (***W** *) measures the degree of association among all raters and assumes an ordinal level of measurement*

e) *Homogeneity coefficient -* **H** *(Aiken, 1985) – a measure of agreement among all raters for a single descriptor unit assuming ordinal level of measurement*

f) *Rater agreement index –* **RAI** *(Burry et al., 1996) – an average rater agreement over all descriptor units*

****)*** *Statistically significant at level p < .05*

The most commonly used index of the degree of association - **Pearson product moment correlation coefficient** (*r*) indicates that there is a strong positive linear relationship between raters for each of the three sets of descriptor units. All correlation coefficients were statistically significant ($p < .01$).

This strong inter-rater correlation indicates that inter-rater reliability in terms of internal consistency (**Cronbach'** α) should also be high. As can be seen in Table 2, the inter-rater reliability was above 0.98 for all the three scales of language proficiency and this suggests that the raters were applying similar constructs when they sorted the descriptor units into six ordered piles.

Although these results are very encouraging they have to be interpreted cautiously, since the sample sizes in our case are very small ($n \leq 25$) and the Pearson product moment correlation assumes interval level measurement and normal distribution of ratings.

On the other hand, Pearson product moment correlation and Cronbach' α are indexes of association, and thus they do not provide information about the degree to which two or more raters agree on the rating of a particular descriptor unit.

To overcome these limitations, we computed some other indexes of association and agreement that are more appropriate for lower levels of measurement and smaller samples and they are presented in the last four rows in Table 2.

As Table 2 shows, there is complete agreement between Cronbach's α and its extension and generalization for qualitative data - **Proportional reduction in loss (PRL) index**, proposed by Rust and Cool (Rust & Cool, 1994). This agreement between the two indexes warrants

generalization about a high level of inter-rater consistency regardless of small sample sizes and the level of measurement.

The values of Kendal's coefficient of concordance (W) - which is one of the most widely used non-parametric measures of correlation - for Listening, Reading and Writing are very close to those of the average Pearson correlation and they all are also statistically significant and high enough to conclude that the inter-rater consistency for all three language scales is quite satisfactory.

As far as the agreement between raters is concerned, the last two indexes presented in Table 2 confirm that the raters taking part in the study are very homogeneous in terms of exact agreement as well as in terms of association. In other words, there is enough evidence to conclude that the inter-rater reliability for all the three analyzed scales of language proficiency is very high.

The results of the inter-rater reliability analysis suggest that the twelve raters in the validation study share a common view about the scales of language proficiency in Listening, Reading and Writing, and the level of their agreement on the continuum of the progression of language proficiency is also high. Whether these views coincide with the original constructs, as presented in the CEF scales of language proficiency, is a question which will be discussed in the next section.

Although a high degree of inter-rater consistency was observed for all three skills it should be mentioned that the indexes for Writing were slightly lower in comparison with Listening and Reading. The differences are negligible but they are in evidence for all computed indexes and, therefore, suggest that the construct of the language proficiency in Writing is not quite so well defined as the other two constructs.

3.2. Scale Validity

Exploratory factor analysis is a statistical method for investigating the structure of inter-correlations between a set of variables and it makes it possible to reduce the larger number of variables to a smaller number of unobservable (latent) factors (constructs). Since exploratory factor analysis allows the exploration of the structure of the construct and the relationships between its elements, it has been considered one of the main tools in construct validity studies (Messick, 1989, p. 52; Thompson & Daniel, 1996, p. 198).

A **Principal Component Analysis** was conducted twice for each of the scales of language proficiency: once for the group of language teachers (Gr1) and once for the group of language testing specialists (Gr2). The high inter-rater consistency suggests that a one - factor solution might be expected and the results of the principal component analysis confirm this – in all six cases (2 groups x 3 scales) there was only one factor with an eigen value greater than one, and it explained a very substantial amount (between 88 – 93 %) of the total variance (Table 3).

TABLE 3. Principal component analysis – summary results

		Listening			Reading			Writing		
		Gr1	**Gr2**	**Total**	**Gr1**	**Gr2**	**Total**	**Gr1**	**Gr2**	**Total**
KMO[a]		.83	.85	.74	.90	.87	.86	.91	.88	.79
% of variance[b]		91%	89%	88%	90%	93%	91%	88%	89%	87%
Correlation[c]		.968**		.959**	.973**		.971**	.974**		.927**
Mean factor score	**Level A1**	–1.42 (*n* = 3)			–1.41 (*n* = 3)			–1.75 (*n* = 2)		
	Level A2	–1.11 (*n* = 3)			–1.03 (*n* = 3)			–1.50 (*n* = 3)		
	Level B1	–0.34 (*n* = 2)			–0.26 (*n* = 5)			–0.59 (*n* = 4)		
	Level B2	+0.57 (*n* = 6)			+0.31 (*n* = 3)			+0.21 (*n* = 4)		
	Level C1	+0.81 (*n* = 4)			+1.00 (*n* = 5)			+0.31 (*n* = 5)		
	Level C2	+1.61 (*n* = 1)			+1.53 (*n* = 1)			+0.41 (*n* = 7)		

a) *Kaiser-Meyer-Olkin Measure of Sampling Adequacy.*
b) *Percent of the total variance explained by the first factor*
c) *Correlation between: (i) factors scores, based on the two groups of raters or (ii) the factor scores of the total group of raters and the original CoE levels (** means significance at the .01 level).*

The correlation between the factor scores for the scale descriptors based on the different groups of raters was above 0.96 for each of the three scales of language proficiency, and this implies that the two groups of raters applied the same construct when they did the sorting task. This conclusion led to the next step when the principal component analysis was repeated with the whole group of 12 raters. Again a one-factor solution, explaining between 87 and 91 % of the total variance, best describes the structure of each of the scales of language proficiency. The correlation between the factor scores of descriptor units and their original CEF levels was above 0.92 for all three scales, which is evidence for the scale validity of the analyzed scales of language proficiency. Moreover, a closer look at the average factor scores per levels (Table 3 – last six rows) shows that their mean factor score increases with the levels. In other words, descriptor units belonging to lower levels of language proficiency were scored by raters on average lower than those belonging to higher levels of language proficiency. This means that, in general, raters perceived the stages in language acquisition in a similar way to that described in the CEF scales of language proficiency for Listening, Reading and Writing.

The small number of descriptor units per level does not allow the application of any statistical test and, therefore, we cannot say whether the differences between level means are statistically significant or not but, for all three language scales, there is a clear tendency of growth in the mean factor scores with the progression of CEF levels.

The constructs of language proficiency in Listening, Reading, and Writing consist of a number of verbal descriptors which are closely interrelated. Since the constructs were defined on the basis of certain observed similarities/dissimilarities in patterns of second language acquisition, the raters as experienced teachers should be able to recreate the original structure

of the analyzed constructs. The comparison between the original structure and the empirical structures of organization is, therefore, an integral part of construct validation (Messick, 1989; Fleishman & Mumford, 1991) and **Pattern Matching**, proposed by Trochim (1999) to be used as a tool in validity studies, was also applied in this study.

Pattern matching aims to compare the structure of the theoretical construct with some empirical structure (in our case the structure of 12 ratings). Since the constructs of language proficiency in Listening, Reading and Writing are described in the form of six-point scales it can be assumed that the two descriptor units belonging to one and the same level will be much more similar than two units belonging to different levels of proficiency.

For example, let us consider the first four descriptor units for Listening (Table 1a). According to the CEF scale for language proficiency for Listening the first three belong to the lowest level – A1 (1), while the fourth one belongs to the next level A2 (2). If we consider the absolute difference between the levels of two descriptor units as a measure of their dissimilarity, then the dissimilarity between each of the three first units will be equal to 0, while the dissimilarity between each of the first three units and the fourth one will be equal to 1 and they can be presented in the form of a triangular matrix of dissimilarity, as is shown in Fig. 1.

Fig. 1 Pattern Matching – Dissimilarity Matrices - Illustrative Example

A. Theoretical Pattern (Coe Scale)

Unit's ID (Level)	1 (1)	2 (1)	3 (1)	4 (2)
1 (1)	**0**	**0**	**0**	**1**
2 (1)		**0**	**0**	**1**
3 (1)			**0**	**1**
4 (2)				**0**

B. Empirical Pattern (Rater 1)

Unit's ID (Level)	1 (1)	2 (1)	3 (2)	4 (2)
1 (1)	**0**	**0**	**1**	**1**
2 (1)		**0**	**1**	**1**
3 (2)			**0**	**0**
4 (2)				**0**

However, according to the first rater, the first two units belong to level A1 (1), but the third and fourth unit belong to level A2 (2). This leads to another triangular matrix of dissimilarity, which is slightly different from the first one.

The correlation between the elements of these two matrices will indicate the degree of the concordance between the theoretical and observed patterns of the analyzed construct or, in other words, the degree of construct validity of the scales of language proficiency.

Applying this approach to all descriptor units for each of the three analysed scales of language proficiency and all twelve ratings, the following correlations between the theoretical and observed patterns were obtained:

- Listening +0.89***
- Reading +0.91***
- Writing +0.84***

All the three correlation coefficients are statistically significant at level p = .001 and – what is more important – sufficiently high to conclude that the CEF scales of language proficiency in

Listening, Reading and Writing are valid enough to be used as a framework for learning, teaching and assessment.

This conclusion is supported by the results (Table 4) of a **Discriminant Analysis** (using as the grouping variable the original levels of scale descriptors and the 12 ratings of the language specialists as independent variables), which showed that all of the original grouped cases (scale descriptor units) were correctly classified. As can be seen in Table 4, for all 64 descriptor units the probability of membership in the original CEF level is greater than .985. In other words, based on the twelve ratings, each descriptor unit can be correctly assigned to the level to which it originally belongs.

TABLE 4. Discriminant analysis – summary results

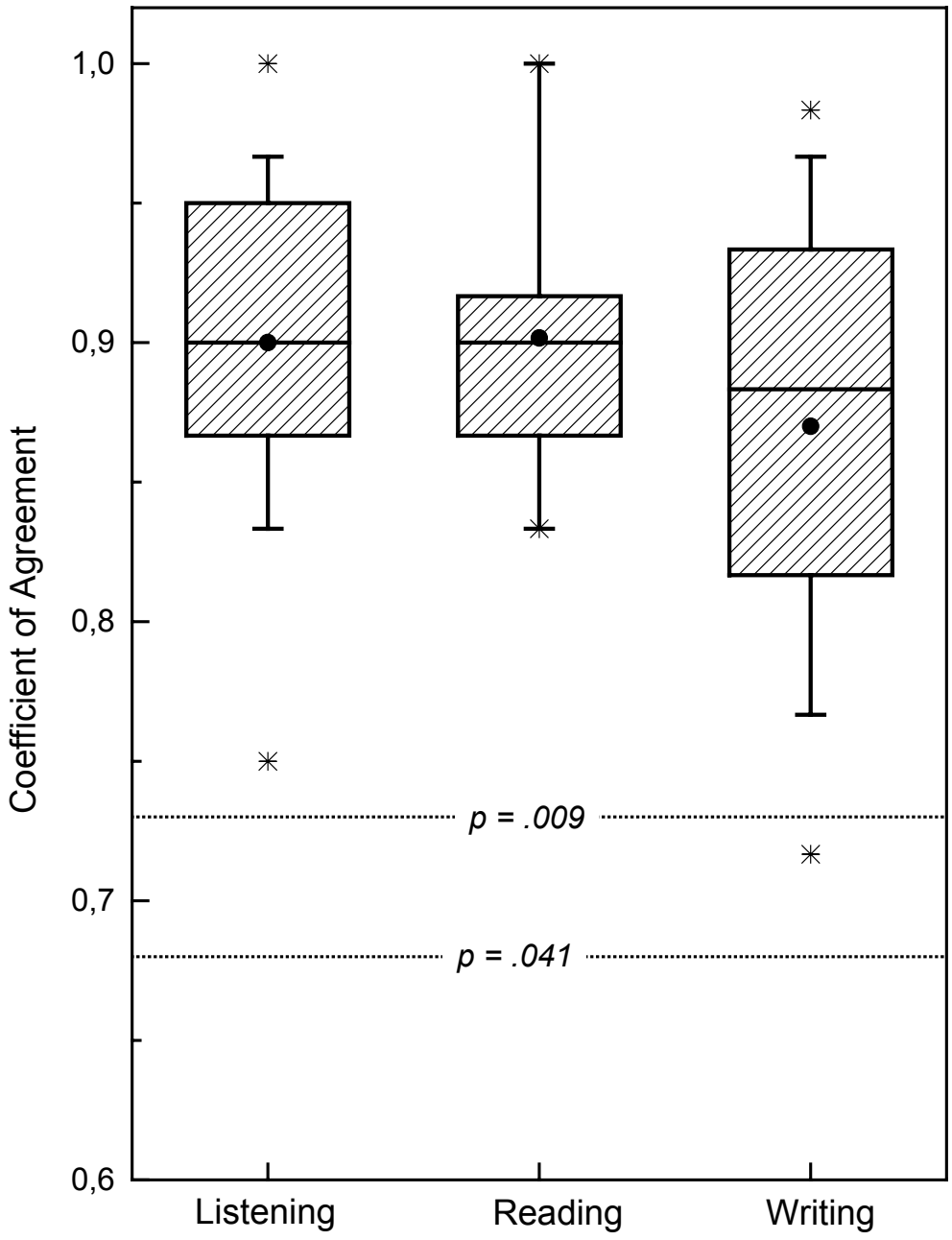
	ID	CoE level	Predicted level	P[a] (A1)	P[a] (A2)	P[a] (B1)	P[a] (B2)	P[a] (C1)	P[a] (C2)
Listening	11	A1	1	1,0000	,0000	,0000	,0000	,0000	,0000
	12	A1	1	1,0000	,0000	,0000	,0000	,0000	,0000
	13	A1	1	1,0000	,0000	,0000	,0000	,0000	,0000
	21	A2	2	,0019	,9981	,0000	,0000	,0000	,0000
	22	A2	2	,0000	1,0000	,0000	,0000	,0000	,0000
	23	A2	2	,0000	1,0000	,0000	,0000	,0000	,0000
	31	B1	3	,0000	,0000	1,0000	,0000	,0000	,0000
	32	B1	3	,0000	,0000	1,0000	,0000	,0000	,0000
	41	B2	4	,0000	,0000	,0000	1,0000	,0000	,0000
	42	B2	4	,0000	,0000	,0000	1,0000	,0000	,0000
	43	B2	4	,0000	,0000	,0000	,9992	,0008	,0000
	44	B2	4	,0000	,0000	,0000	,9992	,0008	,0000
	45	B2	4	,0000	,0000	,0000	1,0000	,0000	,0000
	46	B2	4	,0000	,0000	,0000	,9982	,0018	,0000
	51	C1	5	,0000	,0000	,0000	,0000	1,0000	,0000
	52	C1	5	,0000	,0000	,0000	,0000	1,0000	,0000
	53	C1	5	,0000	,0000	,0000	,0002	,9998	,0000
	54	C1	5	,0000	,0000	,0000	,0000	1,0000	,0000
	61	C2	6	,0000	,0000	,0000	,0000	,0000	1,0000
Reading	11	A1	1	1,0000	,0000	,0000	,0000	,0000	,0000
	12	A1	1	1,0000	,0000	,0000	,0000	,0000	,0000
	13	A1	1	,9852	,0148	,0000	,0000	,0000	,0000
	21	A2	2	,0059	,9941	,0000	,0000	,0000	,0000
	22	A2	2	,0000	1,0000	,0000	,0000	,0000	,0000
	23	A2	2	,0000	1,0000	,0000	,0000	,0000	,0000
	31	B1	3	,0000	,0000	1,0000	,0000	,0000	,0000
	32	B1	3	,0000	,0000	1,0000	,0000	,0000	,0000
	33	B1	3	,0000	,0000	1,0000	,0000	,0000	,0000
	34	B1	3	,0000	,0000	1,0000	,0000	,0000	,0000
	35	B1	3	,0000	,0000	1,0000	,0000	,0000	,0000
	41	B2	4	,0000	,0000	,0000	1,0000	,0000	,0000
	42	B2	4	,0000	,0000	,0000	1,0000	,0000	,0000
	43	B2	4	,0000	,0000	,0000	1,0000	,0000	,0000
	51	C1	5	,0000	,0000	,0000	,0000	1,0000	,0000
	52	C1	5	,0000	,0000	,0000	,0000	1,0000	,0000
	53	C1	5	,0000	,0000	,0000	,0000	1,0000	,0000
	54	C1	5	,0000	,0000	,0000	,0000	1,0000	,0000
	55	C1	5	,0000	,0000	,0000	,0000	1,0000	,0000
	61	C2	6	,0000	,0000	,0000	,0000	,0000	1,0000
Writing	11	1	1	1,0000	,0000	,0000	,0000	,0000	,0000
	12	1	1	1,0000	,0000	,0000	,0000	,0000	,0000
	21	2	2	,0000	1,0000	,0000	,0000	,0000	,0000
	22	2	2	,0007	,9993	,0000	,0000	,0000	,0000
	23	2	2	,0006	,9994	,0000	,0000	,0000	,0000
	31	3	3	,0000	,0000	1,0000	,0000	,0000	,0000
	32	3	3	,0000	,0000	1,0000	,0000	,0000	,0000
	33	3	3	,0000	,0000	1,0000	,0000	,0000	,0000
	34	3	3	,0000	,0000	,9999	,0001	,0000	,0000
	41	4	4	,0000	,0000	,0000	1,0000	,0000	,0000
	42	4	4	,0000	,0000	,0000	1,0000	,0000	,0000
	43	4	4	,0000	,0000	,0000	1,0000	,0000	,0000
	44	4	4	,0000	,0000	,0010	,9990	,0000	,0000
	51	5	5	,0000	,0000	,0000	,0000	1,0000	,0000
	53	5	5	,0000	,0000	,0000	,0000	1,0000	,0000
	53	5	5	,0000	,0000	,0000	,0000	1,0000	,0000
	54	5	5	,0000	,0000	,0000	,0000	1,0000	,0000
	55	5	5	,0000	,0000	,0000	,0000	1,0000	,0000
	61	6	6	,0000	,0000	,0000	,0000	,0000	1,0000
	62	6	6	,0000	,0000	,0000	,0000	,0107	,9893
	63	6	6	,0000	,0000	,0000	,0000	,0000	1,0000
	64	6	6	,0000	,0000	,0000	,0000	,0001	,9999
	65	6	6	,0000	,0000	,0000	,0000	,0041	,9959
	66	6	6	,0000	,0000	,0000	,0000	,0000	1,0000
	67	6	6	,0000	,0000	,0000	,0000	,0000	1,0000

a) Probabilities of membership in Level A1, A2, B1, B2, C1, C2

The Agreement Coefficients (Aiken, 1985) between the CEF levels of scale descriptor units and the rating of the 12 raters also confirm this inference. As can be seen from the box-plots in Fig. 2 all agreement coefficients, even the outliers, are significant at least at level $p < .05$ and the average agreement (given below) is very high:

- Listening 0.90
- Reading 0.90
- Writing 0.80

Fig. 2. Agreement between CoE levels and Ratings: Frequency Distributions



Summarizing the results of all applied methods of analysis (exploratory factor analysis, pattern matching, discriminant analysis and analysis of the agreement), we can conclude there is strong evidence in support of the construct validity of CEF scales of language proficiency.

On the other hand, the comparative analysis of the three analyzed scales of language proficiency indicates that the scales for Reading and Listening are better developed than the scale for Writing, since the discrepancy between ratings and the theoretical construct for Writing is larger than for the other two scales, which was evident for all applied methods for construct validation.

3.3. The Continuum of Developing Language Proficiency

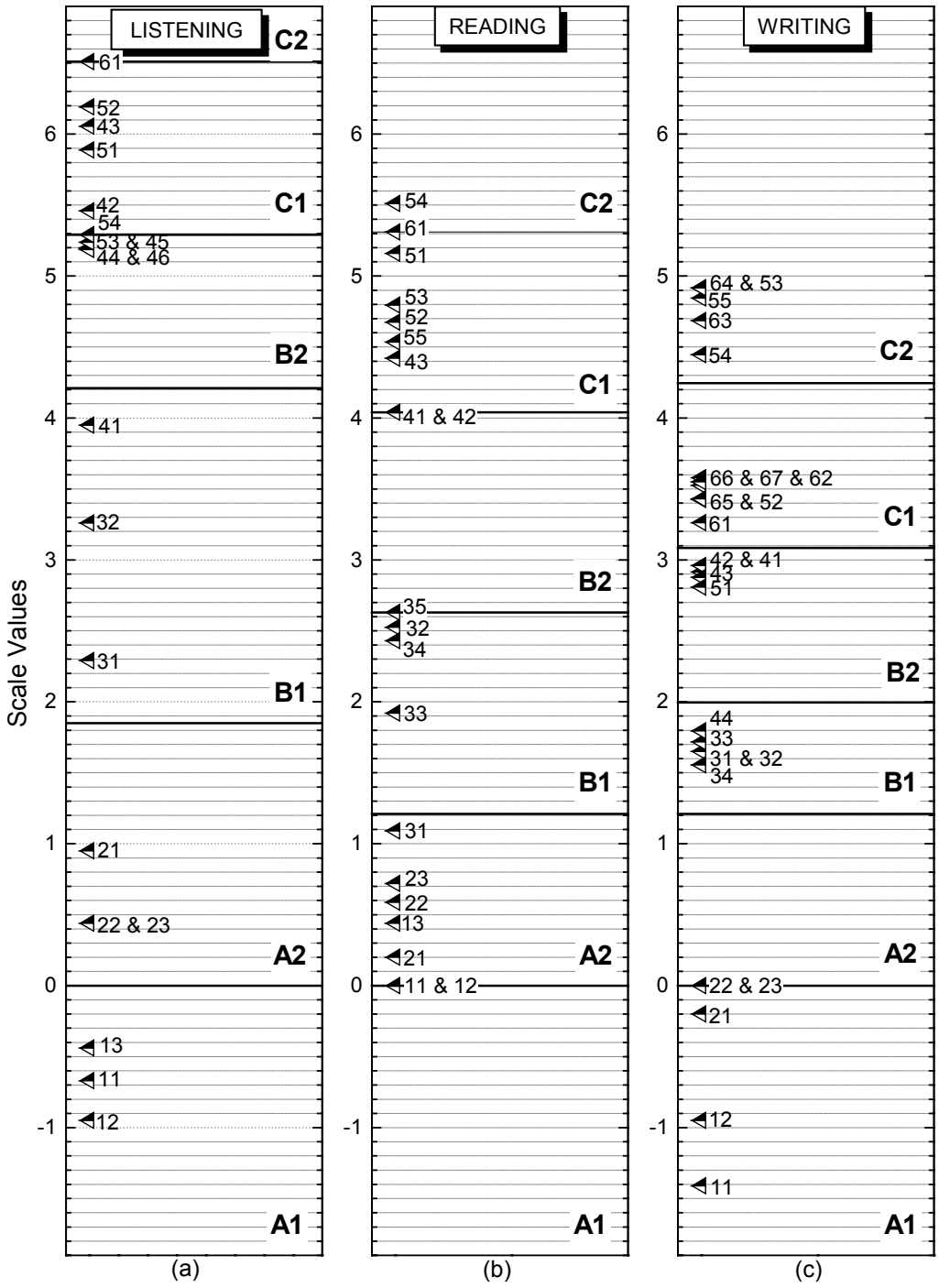
One of the main criticisms of Scales of Language Proficiency in general is that "… *with the vast majority of scales of language proficiency, it is far from clear on what basis it was decided to put certain statements at level 3 and others at Level 4 anyway*" (North & Schneider, 1998). In other words, there are doubts that some of the existing scales represent adequately the continuum of developing language proficiency.

Another problem concerns the choice of the most appropriate level of measurement for a certain scale. The quantification of the CEF Scales of Language Proficiency carried out by North and Schneider (1998) was based on Speaking and Interaction and the applied approach aimed to create an approximately equal-interval scale, assuming interval level of measurement. As Sadler (1987) points out, however, "simply having a set of standards does not necessarily imply an interval scale".

In the present study a different approach - the **method of successive intervals** – was applied (Torgerson, 1958, Dunn-Rankin, 1983) since it does not assume equal intervals and can be applied to comparatively smaller samples.

The results of the present quantification are presented in Fig. 3 and the exact scale value for every descriptor unit is given in the last column of Table 1. The scale values correspond closely to the factor scores – the correlation between them for all three skills is above 0.98 – which is an obvious indicator of their credibility.

Fig. 3. Method of Successive Intervals:
Descriptor Quantification



As can be seen in Fig. 3, the levels of language proficiency in Listening, Reading and Writing are not of equal length. This finding is not surprising, because there are no obvious theoretical or empirical rationales to expect equal length of the language proficiency bands. It may very well take a different amount of time and effort to move to the next level of proficiency for different successive levels and for different skills.

Another important conclusion of the present scale quantification is that there are some cases where no clear distinction between successive bands can be made (for example – between levels A1 and A2 in Reading). Therefore, if the scales are used for assessment purposes, the assessor might easily make classification errors when she has to decide for an examinee whether he or she is on level A1 or A2 in Reading.

If we compare the average scale values for the CEF levels (Table 5) we will notice that although, as a whole, they increase with the growth of language proficiency, there is one case where the average for the higher level descriptor units is smaller than that for the lower level ones (Writing – levels C1 and C2). It suggests that these level descriptors need reconsideration and revision. Comparatively small are also the differences in the averages for levels B2 & C1 for Listening and levels A1 & A2, C1&C2 for Listening. This would lead to more classification errors even though these differences are in the correct direction.

TABLE 5. Scale quantification – summary results

CoE Level	Listening			Reading			Writing		
	Mean	***n***	**Range**	**Mean**	***n***	**Range**	**Mean**	***n***	**Range**
1. A1	–0.69	3	0.51	+0.15	3	0.44	–1.18	2	0.46
2. A2	+0.61	3	0.51	+0.50	3	0.52	–0.07	3	0.20
3. B1	+2.78	2	0.97	+2.12	5	1.54	+1.64	4	0.16
4. B2	+5.17	6	2,11	+4.17	3	0.38	+2.64	4	1.16
5. C1	+5.65	4	0.95	+4.94	5	0.97	+4.09	5	2.11
6. C2	+6.51	1	0.00	+5.31	1	0.00	+3.85	7	1.65

As a whole, there is a high correlation between the scale values of the descriptor units and their corresponding CEF level (0.96 for Listening and Reading and 0.92 for Writing), but at the same time some descriptor units should, according to their scale values, belong to a higher or lower level of language proficiency. A more detailed analysis of these descriptor units will be carried out in the next section.

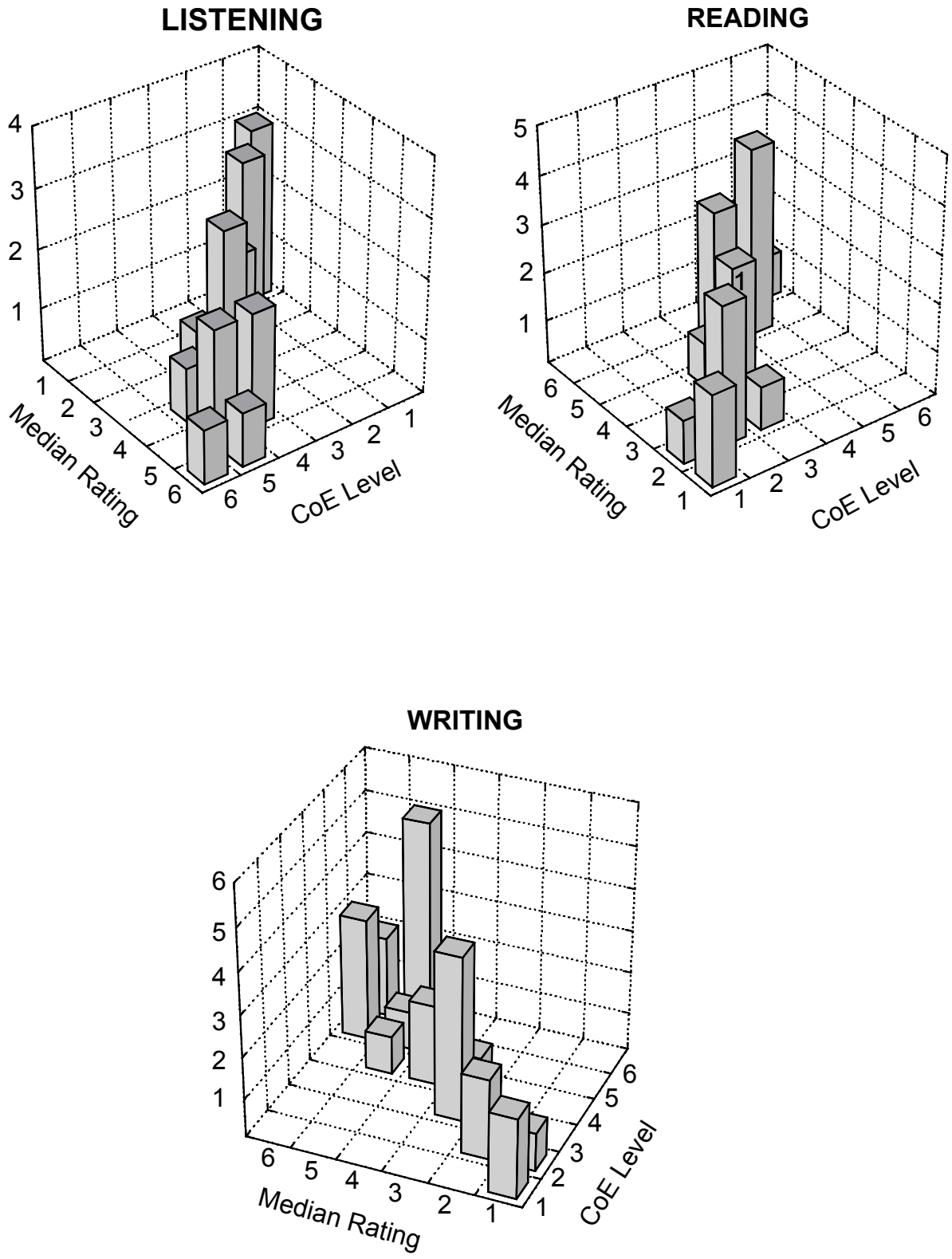
3.4. Descriptor Analysis

As Cronbach (1990, p. 186) pointed out "the job of validation is partly to support an interpretation, and partly to detect anything wrong with it". Therefore, for the purposes of further development of the CEF scales of language proficiency in Listening, Reading and Writing it is useful to have a closer look at those descriptor units for which there is:

- Discrepancy between their original membership in CEF levels and their rating.
- Higher degree of disagreement between raters.
- Misplacement on the scale of language proficiency according to the results of scale quantification.

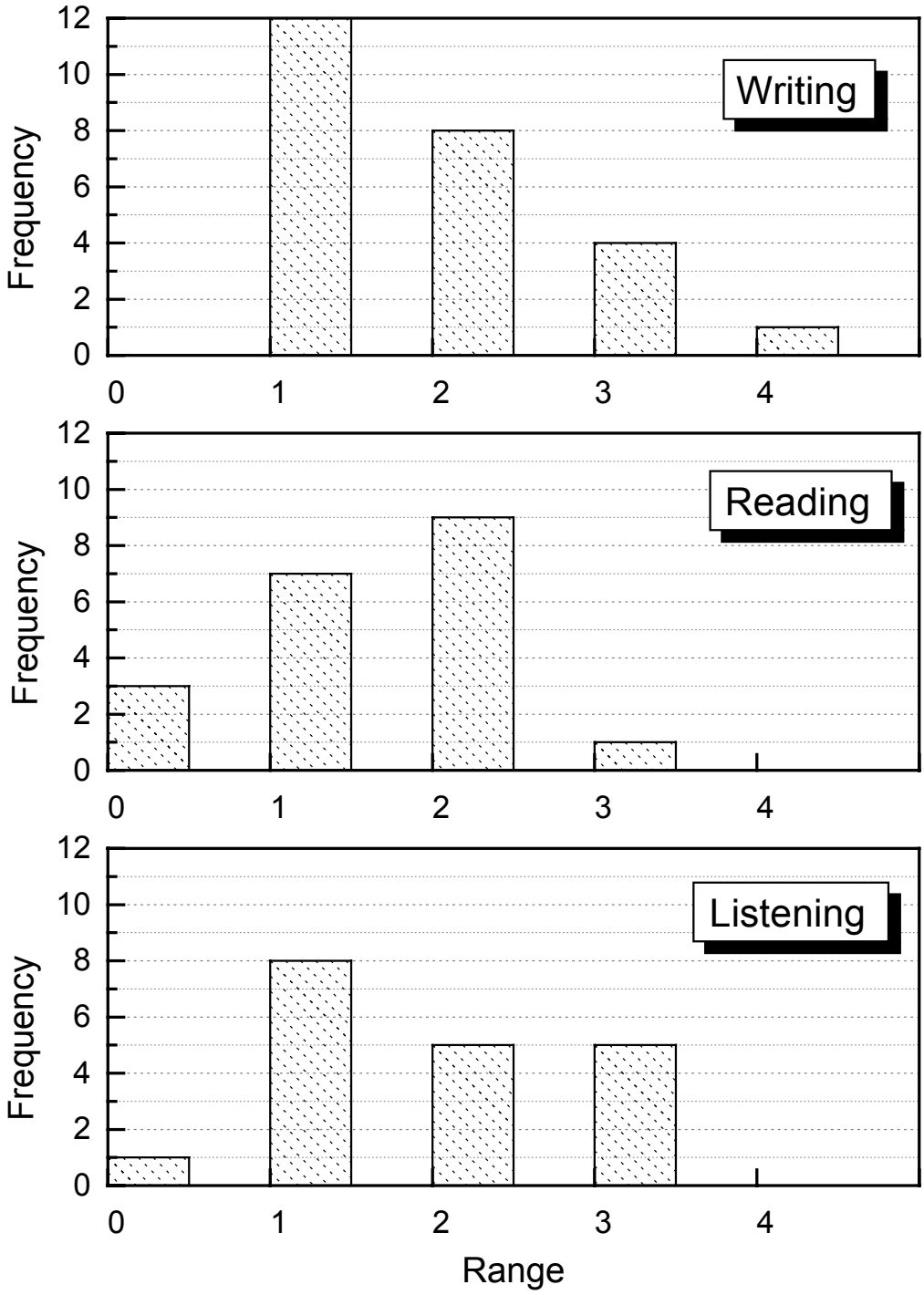
As can be seen in Fig. 4, most of the descriptor units, according to the median of the 12 ratings, were assigned to the same levels of language proficiency as in the original CEF level description (74% for Listening, 65% for Reading and 52% for Writing).

Fig. 4. Bivariate Frequency Distributions for CoE Levels and Median Rating of Descriptor Units



For half of the units, raters differed in their ratings with no more than 1 point (Fig. 5). And only for 17% of the all descriptor units (11 out of 64) was the range of the ratings larger than two points.

Fig. 5. Frequency Distributions of Range of Ratings



For **Listening** there were five cases of discrepancy between the original CEF levels of the descriptor units and the median rating. It occurred only on the higher levels of proficiency (B2 and C1) and the difference was no more than one level up or down. In only two (L43 & L52) of those five cases, can the inconsistency be explained with lower agreement between raters. In the other three cases (L41, L45, L53), there was a high degree of inter-rater agreement and according to their aggregated rating the units should be positioned one level up or down. Whether the raters are right or wrong is not of primary concern here. The more important thing is to explore in greater depth these cases of disagreement and potential reasons for this discrepancy.

Let us consider units L45 and L53:

> *L45. I can understand most current affairs programmes.*
> *L53. I can understand television programs without too much effort.*

According to the raters, the first unit (L45) should belong to level five - C1, while the second one should move one level down to level four - B2. A possible explanation for this inconsistency with the original CEF levels might be that the phrase "understand … without too much effort" is a signal for raters that there will be some (but not too much) effort. On the other hand, the verb "understand" without any other specifications may be interpreted that there will actually be no effort in the process of understanding. It seems that the phrase "without too much effort" has a stronger impact on the raters than the fact that the first unit (L45) concerns only "most current affairs programs" while the second unit (L53) deals with "television programs" as a whole.

There were seven cases (R13, R31, R35, R41, R42, R43, R54) for **Reading**, where the raters in general disagreed with the original level membership of the units. Again as in the case of Listening, most of those occurrences were in the higher levels of proficiency. None of these discrepancies can be explained with lower disagreement between raters. The rater agreement index (RAI – Burry at al., 1996) for all the seven cases was at least 0.89. The other indexes of agreement, calculated on item level - homogeneity coefficient (Aiken, 1985), and PRL (Rust and Cool, 1994) - also confirm that there is a high degree of agreement between raters concerning those seven units.

In five of the seven cases, the aggregated rating suggests a higher level for the descriptor units and for three of these cases the units concern reading articles and reports. This activity probably is quite demanding and less attractive for many people even if the articles and reports were written in their mother tongue, and this might explain to some degree the disagreement between raters and the original order of the descriptor units.

There is, however, no obvious explanation for unit R13, which according to the raters should be on level two (A2) and not on level one (A1) as it is in the original scale. If we compare (Fig. 3) this unit (R13. "*I can understand familiar, very simple sentences, for example on notices and posters or in catalogues."*) with unit R21 (*"I can read very short, simple texts."*) we will notice that the second one (R21) is rated lower than the first one (R13). There is no logical explanation for this raters' order since most texts – even the shortest and simplest one – usually consist of sentences and to understand the text one should be able to understand simple sentences. It is possible, however, that raters might think that posters, notices and catalogues (R13) contain some more technical vocabulary than *"simple texts"* (R21).

The scale for language proficiency in **Writing** is the scale with the greatest discrepancy between raters and the original levels of the descriptor units. Twelve (W21, W42, W44, W51, W53, W54, W55, W61, W62, W65, W66, W67) out of 25 cases, according to the aggregated rating, belong to either a higher or lower level of proficiency. The degree of disagreement is remarkable especially for the two highest levels (C1 and C2). In four of the cases (W51, W61, W65, W67) the discrepancy between the original levels and the median rating might be due, in part, to the high degree of disagreement between raters.

Unit W43 (*I can write a report passing on information or giving reasons in support of or against a particular point of view)* deserves special attention. Although this is a unit where the median rating is the same as the original level (B2), the ratings varied between levels two (A2) and six (C2). The most reasonable explanation would be some accidental mistakes by some of the raters, but the frequency distribution of ratings (Table 6) excludes this explanation.

TABLE 6. Frequency distribution of ratings for descriptor units W43

Levels	Frequency	Percent	Cumulative Percent
1. A1	0	0.0	0.0
2. A2	1	8.3	8.3
3. B1	1	8.3	16.7
4. B2	6	50.0	66.7
5. C1	2	16.7	83.3
6. C2	2	16.7	100.0
Total	**12**	**100.0**	

The discrepancy between the original descriptor levels and the median rating in some of the cases can be easily explained. Let us compare for example unit W55. "*I can write different kinds of texts in an assured personal style appropriate to the reader in mind.*" with unit W62. "*I can write complex letters."* It seems reasonable that if someone can write different kind of texts in an appropriate style he/she can write also complex letters. Therefore, the median rating, which suggests W55 to be on level six (C2) and W62 to be on level five (C1), should be taken into account in the further scale development.

In general, writing reports and articles as well as reading them was considered the highest level of language proficiency, while writing summaries, reviews and letters is something which, according to the raters, requires a lower level of proficiency.

The content of the three units given below differ only in terms of the form of the writing (an essay, a letter or a report), but while the median rating for unit W52 is level five (C1), the median rating for the other two units is level six (C2), and they are placed far above the first one according to their score values (Fig. 3).

W52. *I can write detailed expositions of complex subjects in **an essay** underlining what I consider to be the salient issues.*
W53. *I can write detailed expositions of complex subjects **in a letter** underlining what I consider to be the salient issues.*
W54. *I can write detailed expositions of complex subjects **in a report** underlining what I consider to be the salient issues.*

It is difficult to explain these differences in the ratings of three almost identical descriptor units except if we take into account that the raters are all teachers in Finnish as a foreign (second) language either in the upper secondary school or at the University level and, for all of them, writing an essay is more or less a school-related activity, which in comparison with writing a report or an official (business) letter, may appear much less demanding and, therefore, require a lower level of language proficiency.

These and the other possible explanations for the inconsistency between the ratings and the original scale descriptors require verification and further scale development in order to minimize scale misinterpretation and to ensure population generalisability of the interpretation of the construct.

4. Conclusion

This study aimed to analyse the validity of the CEF scales of language proficiency in Listening, Reading and Writing. It produced a fair amount of evidence about the validity of the analyzed scales, which leads us to the conclusion that they can be successfully used as a framework for foreign language learning, teaching and assessment.

Validity, however, as Messick (1989, p. 13) has observed, is "an evolving property and the validation is a continuing process". The finding that the scales have a high degree of validity does not mean that the scales are perfect and need no improvement. The concrete results of the study show some of the directions of further development and revision.

The comparison between the three analyzed scales demonstrated that the scale for language proficiency in Reading is the best one and that the scale for Writing needs more detailed reconsideration and revision, especially its higher level descriptors (levels C1 and C2).

Looking at the scale validation as a process, this study should be considered only as a small part of the process and its results should be interpreted cautiously, bearing in mind its limitations, which are as follows:

The validity evidence is based on a small sample of raters with a common background – teachers of Finnish as second/foreign language. To be able to generalize the results of the study, it has to be replicated with other samples of experts with a different background.

- The Finnish translation of the CEF scales of language proficiency in Listening, Reading and Writing was used; if their original English version or some other translations were used the results might be different.
- The materials for the sorting task were prepared by splitting the original level descriptors into a larger number of independent units. This split, however, is not the only one possible and if the descriptors were split in a different way, the results of the validation study might differ.

- The sorting task is only one of a number of possible approaches and changing the procedure could also lead to different results.
- The validity evidence provided by this study was based on only one source of information – raters and their interpretations of CEF scales – and therefore the scope of the study is quite limited.

In other words, although the results of the study indicate that the CEF scales of language proficiency are valid enough to be used as a framework for foreign language learning, teaching and assessment, the generalisability of this finding needs further investigation.

References:

Abedi, J., Baker, E. & Herl, H. (1995). Comparing reliability indices obtained by different approaches for performance assessment. CSE Technical Report 401, LA: CRESST.

Aiken, L. (1985). Three coefficients for analyzing the reliability and validity of rating. Educational and Psychological Measurement, 45, 131 – 142.

Brindley, G. (1997). Investigating second language listening ability. In: Brindley, G. & Wigglesworth, G. (Eds.) Access: Issues in language test design and delivery. Sydney: National Centre for English Language Teaching and Research, Macquarie University, 65 – 85.

Burry- Stock, J., Shaw, D., Laurie, C. and Chissom, B. (1996). Rater agreement indexes for performance assessment. Educational and Psychological Measurement, 56, (2), 251-262.

Butler, F. & Stevens, R. (1998). Initial steps in the validation of the second language proficiency descriptors for public high schools, colleges, and universities in California: Writing. CSE Technical Report 497, LA: Center for the Study of Evaluation, University of California.

Cook, T. and Campbell, D. (1979). Quasi-experimentation: Design & analysis issues for field setting. Boston: Houghton Mifflin Company.

Council of Europe, (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press.

Cronbach, L. (1990). Essentials of psychological testing. N.Y.: HarperCollins.

Crooks, T. & Kane, M. (1996). Threats to the validity use of assessment. Assessment in Education: principles, policy & practice, 3 (3), 265 – 286.

Dun-Rankin, P. (1983). Scaling methods. Hillsdale: Erlbaum.

Fleisman, E. and Mumford, M. (1991). Evaluation classification of job behaviour: A construct validation of the ability requirement scales. Personnel Psychology, 44, (3), 523 – 576.

Huttunen, I. & Huttunen, J. (1998). Vieraat kielet: Oppiminen, opettaminen, arviointi yleiseurooppalainen viitekehys luonnos 2 ehdotuksesta viitekehykseksi. Oulu: Oulun Yliopistopaino.

Matthews, M. (1990). The measurement of productive skills: doubts concerning the assessment of criteria of certain public examinations. English Language Teaching Journal, 44, (2), 117 – 121.

McKay, P. (2000). On ESL standards for school-age learners. Language Testing, 17, (2), 185 – 214.

McNamara, T. (1996). Measuring second language performance. London: Longman.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement. Washington, DC: American Council of Education and National Council on Measurement in Education. 13-103.

North, B. and Schneider, G. (1998). Scaling descriptors for language proficiency scales. Language Testing, 15 (2), 217 – 263.
Rust R. & Cool, B. (1994). Reliability measures for qualitative data: theory and applications. Journal of Marketing Research. 31, 1-14.

Sadler, D. (1987). Specifying and promulgating achievement standards. Oxford Review of Education, 13, (2), 191 – 209.

Shohamy, E. (1995). Performance assessment in language testing. Annual Review of Applied Linguistics, 15, 188 – 211.

Standards for Educational and Psychological Testing. (1999). Washington, DC: Educational Research Association.

Thompson, B. and Daniel, L. (1996). Factor analytic evidence for the construct validity of score: A historical overview and some guidelines. Educational and Psychological Measurement, 56, (2), 197 – 208.

Torgerson, W. (1958). Theory and methods of scaling. N.Y.: Wiley.

Trochim, W. (1999). Pattern matching for construct validity. [On-line] Available: http://trochim.human.cornell.edu/kb/pmconval.htm

CHAPTER 9: DIALANG - A DIAGNOSTIC LANGUAGE ASSESSMENT SYSTEM FOR ADULT LEARNERS

by Ari Huhta, Sari Luoma, Mats Oscarson, Kari Sajavaara, Sauli Takala, and Alex Teasdale

Introduction

This chapter describes the way in which the Common European Framework (CEF) was used to develop the DIALANG assessment system. The DIALANG project implemented both the CEF assessment scale from A1 to C2 and key parts of the CEF definition of language ability. We describe the DIALANG Assessment Framework (DAF), which builds on these foundations. . In writing a further document, the DIALANG Assessment Specifications (DAS), the contributions of the CEF were used together with other language-oriented and assessment-oriented frameworks. . The chapter also describes how the CEF concept of self-directed learning has influenced the DIALANG approach to assessment. DIALANG integrates self-assessment with external assessment, and the statements used for operationalizing self-assessment in the system are based on the CEF skill level descriptors. These descriptors also provided valuable input for the feedback component, which learners access upon completing a set of DIALANG tasks. In the development work, the 1996 version of the CEF was used, but for reasons of ease of access, reference is made in the text below to the latest edition (Council of Europe 2001).

1. Introduction to DIALANG

DIALANG is a strongly learner-oriented system for assessing and providing feedback on language proficiency in 14 European languages. The system does not issue formal certificates but gives feedback to the learner upon completion of a set of tasks. The target audience is European adults who are interested in assessing their own skills in one or more of the 14 European languages which the system encompasses. The DIALANG project is carried out with the financial support of the European Commission, Directorate-General for Education and Culture (SOCRATES Programme, LINGUA Action D). The development work reported in this chapter was done during DIALANG Phase 1, when the project was coordinated by the Centre for Applied Language Studies, University of Jyväskylä. The work involved collaboration between 24 European universities and institutions.

The languages of the DIALANG assessment system are Danish, Dutch, English, Finnish, French, German, Greek, Icelandic, Irish, Italian, Norwegian, Portuguese, Spanish, and Swedish. DIALANG includes sections for assessing listening, reading, writing, structures, and vocabulary.

The assessment system draws significantly on the CEF. The CEF model of language ability was adopted as the basis for the DIALANG Assessment Framework and the Assessment Specifications, and the 6-level concept of language proficiency proposed in the CEF was adopted as the DIALANG scale. The CEF was chosen because it had the kind of functional approach to language ability which the DIALANG project sought to implement, because it was possible to apply this approach to all the 14 DIALANG languages, and because the CEF was well known and widely accepted in Europe. Some of the skill level descriptors in the

CEF are applied in both the self-assessment and the feedback sections of the DIALANG system. Learner autonomy, self-direction and learner motivation, which have been strongly promoted by the Council of Europe, constitute the philosophical underpinning of DIALANG.

1.1 The DIALANG assessment procedure

Learners begin a DIALANG session by choosing the language in which they wish all instructions and feedback to be presented, and the language of the test they want to take. They then choose the skill they want to be tested in from five alternatives: reading, writing, listening, vocabulary, and structures.

The assessment system in DIALANG Version 1 is test-level adaptive, which means that on the basis of a placement procedure, the system chooses a test of appropriate level of difficulty from three possible choices. DIALANG Version 2 will be adaptive at the item level, which means that the difficulty of the next item to be presented is estimated on the basis of the learner's responses to previous items. In Version 1, DIALANG uses two procedures to estimate the learner's level: a placement test based on vocabulary knowledge, and a set of self-assessment questions. In Version 2, the placement procedures will rely on self-assessment only, since only a very rough estimate is needed to start the test and items are subsequently presented to learners tailored to their ability level.

The first step in the two-step placement procedure is a Vocabulary Size Placement Test. In this test, developed for DIALANG by Paul Meara, learners decide whether words presented to them are real words in the language or not. In the second step, the learner is presented first with a scale describing the six ability levels and second with a set of *I can* statements, through which they assess their own language skills. This section exists for reading, writing, and listening, and the CEF was used as a source in developing the statements. (There is no self-assessment for vocabulary and structures, because no source scales existed in the version of CEF that was used during the development of the assessment framework for these areas since they obviously vary considerably from language to language[5]. Accommodation of language-specific scales would have been difficult in DIALANG, which has 14 test languages and 14 languages in which the test instructions and feedback are presented.) The learner's VSPT and self-assessment responses are used to select the level of difficulty from which the test should start, and the test begins.

There are about 30 items in each test which a learner takes in DIALANG Version 1. In Version 2, it is likely that fewer items will be presented to each learner, since the system will be designed to home in on the learner's ability level as quickly as possible.

At the end of the test, the system gives feedback to the users. It reports scores on the CEF six-level scale from A1 (Breakthrough) to C2 (Mastery), and gives learners an opportunity to explore their performance. The contents of the feedback are explained in more detail later in this chapter. At the end of the feedback section, the user can choose another skill in the same language, choose another test language, or exit DIALANG.

[5] The current version of the CEF does contain scales for general linguistic range, vocabulary range, vocabulary control, and grammatical accuracy (Council of Europe 2001). In accordance with the functional approach of the CEF, these scales describe the learner's lexical and grammatical resources in productive language use, primarily speaking. Successive levels describe ability to deal with more and more communicative situations with increasing linguistic accuracy. However, not surprisingly since the CEF is intended to be applicable to a wide range of different languages, no examples of typical vocabulary or structures are included in the descriptors.

1.2 Information technology in the service of language learners

The DIALANG project utilizes new technology to make the assessment system accessible to its users and flexible to their needs. DIALANG is programmed in Java, and it available on the Internet (www.dialang.org). Users can access the system whenever they want as long as they have an Internet connection at their disposal.

DIALANG has had to come to grips with the problem of developing a robust assessment system which can be put on-line, using IT for innovative and interactive language assessment. While the standard system is fairly conservative in terms of task types, the project has explored the possibilities for innovation in test methods and the provision of feedback. These include a range of different test methods and proposals for assessing writing and speaking, where learners are asked to perform a task and then compare their own performance to benchmarks with comments.

2. How CEF was used in the DIALANG assessment Framework (DAF)

The DIALANG Assessment Framework (DAF) defines the theoretical framework which the assessment system is intended to operationalise. The DAF is directly based on parts of the CEF, especially some dimensions of the theoretical framework outlined in Chapter 2, and the more concrete inventories of communicative tasks, themes, activities, types of texts, and language functions in Chapter 4.

The contents of the DAF are shown in Figure 1. The main body of the DAF, organised in nine sections from A to I, is only seven pages long. This brevity is explained by two things. Firstly, extensive use is made of CEF definitions, which are not repeated in the body of the text. Secondly, the categories included in the Framework are only briefly defined in the body of the text, and extended lists of examples are given in the appendices to the DAF. These will be discussed below. Appendix 1 to this chapter shows a sample page from Appendix 1 of the DAF.

Figure 1. Contents of the Dialang Assessment Specifications (DAF)

The view of language use which DIALANG adopted for implementation is outlined in CEF Chapter 2. This particular construct was adopted because of its acceptability in Europe and its adaptability to a multilingual system. The CEF calls the view "action-oriented" because language users are seen primarily as social actors who use language in the course of their social tasks and activities. The CEF focuses on describing language use, but the authors make the point that this cannot happen independently of the social context in which language is used, or the individual learner-users who take part in the social activities. DIALANG subscribes to this approach even though it can only implement a narrow range of language use situations. This is because the physical assessment context is the computer, and because the assessment procedure involves interaction with texts and tasks but no real-time interaction with other people.

The theoretical dimensions of people's skills and language use which CEF discusses are on a very high level of abstraction. DIALANG refers to these, but it has not been possible to create research designs and provide data on whether such sub-categories and components actually underlie language use in DIALANG. This particularly concerns the general competences of an individual (knowledge, skills, and existential competence, which influence an individual's ability to learn), and the three main components of communicative language competence (linguistic, sociolinguistic, and pragmatic). Similarly, regarding strategies of language use, the project agrees with the CEF description that learner-users use strategies to accomplish tasks. However, strategies vary depending on the situation and the participants' goals. Strategies are also related to language processes and are not necessarily relatable to features of language products, such as task responses, in a simple and straightforward way. Therefore, the members of the project considered it too big a task to try to create tests of strategy use for DIALANG. The lists of strategies in CEF Chapter 4, especially those for reception strategies

were used as resource material in creating advisory feedback, to be discussed under section 5 below.

The CEF divides language activities into four main categories: reception, production, interaction, and mediation. Of these, the current DIALANG system concentrates on receptive skills in particular, with some limited production activities. Receptive skills are emphasized because the responses are scored automatically rather than by human judgement. However, DIALANG has made an effort to create experimental items for assessing productive language skills, which build on learners comparing their own performances against marked-up benchmarks. It is possible that after further development of these items, they will be incorporated into the mainstream DIALANG system.

The CEF identifies four contextual domains in language use: personal, public, occupational and educational. This division makes it possible to specify the range of domains covered in DIALANG. The personal and public domains are at the centre of what DIALANG is intended to assess. Most of the situations which the CEF mentions under these domains, especially in terms of events and operations, have been used as a basis for DIALANG tasks. The occupational domain is also very relevant for DIALANG, but since the assessment system is non-specialized, the work-related situations have been kept at a general rather than profession-specific level. The educational domain is relevant for DIALANG in a limited sense, because the system is not linked to any particular educational setting and in fact is designed to support informal, self-directed language learning as well as formal courses. Interaction situations with "your teacher", for instance, would thus not be relevant to all. The task developers were therefore advised that only the most generic of the educational events and operations listed in the CEF should be used.

The purpose of the DAF is to describe in relatively concrete detail the kinds of language skills that are to be tested in DIALANG, and much of its content is based on the inventories of communicative tasks, themes, activities, types of texts, and language functions in CEF Chapter 4. However, since these inventories are summaries with narrow exemplification and since the DAF was developed as a resource for test writers, we found it useful to develop links to the more extensive lists of examples in the Council of Europe Waystage, Threshold and Vantage Level documents (Council of Europe 1990, 1991, 1996). The inventories for the three levels were presented side by side to illustrate how the Council of Europe documents visualised the progress of language ability (for a sample page, see Appendix 1). The test writers found this presentation layout useful even if they did not always agree with the descriptions of differences between levels. Without this presentation, they would have found it difficult to compare the Waystage, Threshold, and Vantage definitions. Gaps in the original documents were not filled in during the course of DIALANG, but on the basis of the current DAF, it would be possible for the Council of Europe or other interested parties to take up this work. It would also be possible to make some systematic comparisons between the Waystage, Threshold, and Vantage definitions and the CEF. Most notably, as concerns communicative language activities, the level definitions cover receptive and interactive skills, but do not contain specifications for production or for mediation.

The largest conceptual difference between the current DIALANG system and the CEF is that DIALANG makes a major division between the "macro" skills of reading, listening, and writing, and it also includes separate assessment sections for grammar and vocabulary. Such skill-specificity is much less prominent in the CEF, where types of activities are the main organizing principles for the descriptions of language use.

The division in DIALANG was motivated by strong traditional support for such a division in language teaching and assessment. The project wished to give the learner-users of DIALANG a limited and familiar set of choices of areas in which they wanted to assess their own skills. If the project had implemented the CEF categories of Reception, Production, Interaction, and Mediation, it would have had to explain the abstract words of Reception and Production into the more familiar Reading, Listening, Writing, and Speaking anyway to guarantee comprehensibility. Furthermore, the project would have had to make further distinctions in Writing and Speaking to accommodate Production, Interaction and Mediation. Since the current form of DIALANG is unable to deal with the productive modality to the same extent that interactive learning environments can do, the distinctions are less salient for this system.

Although no overall grade is given in DIALANG, no claim is made that the sections would be unrelated. It is even possible that DIALANG may implement the further divisions in the CEF into activities such as, for aural reception (listening): listening to public announcements, listening to media, listening as a member of a live audience, and listening to overheard conversations etc . Such a change would require the investigation of empirical relationships between the sections and items, which would require a large set of performance data and complex conceptual and statistical analyses. This means that changes of this kind are not imminent. However, it is a possible avenue for the future, especially if other projects related to teaching or assessment are able to apply the CEF in full and can thus provide examples for operationalisation.

3. How CEF contributed to the DIALANG assessment specifications (DAS)

The DIALANG Assessment Specifications (DAS) give both generic and detailed guidelines for the writing of DIALANG items. They give an overview of the assessment process, introduce the construct of reading which the tasks aim to assess, define the nature of the task pool which the members of a DIALANG language team were to jointly create, give advice for selecting task materials, and provide detailed instructions for item types together with examples. A sample contents page of the DAS for Reading is given in Figure 2.

In writing the DAS, the CEF was used as one of the resources in defining the construct. Thus, the CEF helped define the dimensions of the DIALANG task pools. Further sources which the project used when writing the DAS included Lynch and Davidson's (1994) procedural approach to writing and revising specifications, and Alderson, Clapham and Wall's (1995) and Bachman and Palmer's (1996) structuring of specification contents. The DAS were written and revised a number of times during the first three years of DIALANG development work in active collaboration between the project co-ordinators and the 14 language-specific teams which developed items for DIALANG.

Figure 2. The contents of the DIALANG Assessment Specifications for Reading

The CEF definitions of the language user's functional competence (such as giving and seeking factual information, expressing and finding out attitudes, suasion) and roles (defined by social status, novice-expert knowledge of the topic, and so on) were adopted for the DAS. The task writers were asked to check that the language use situations covered in their item pools included variation on these dimensions, even if informational purposes and topics might be emphasized more than the others. Variation was desirable because different demands on the users' skills are made by different purposes of language use. Informational purposes and topics are emphasized more than others because the potential exponents of the informational functions, such as reporting, asking, or answering form much wider classes than the exponents of, say, the exponents for socialising such as attracting attention, addressing, or greetings.

Item difficulty is discussed in all the specifications for the five assessment sections of DIALANG. The CEF points out that item difficulty cannot be predicted with certainty because of the complexity of interactions involved. However, the CEF also points out that the idea of difficulty helps grade tasks, and thus potentially facilitates learning. The areas that the CEF discusses as dimensions of task difficulty are learner characteristics, task conditions and constraints, text characteristics, and types of response. Of these, the most relevant for DIALANG were the last two. As an assessment system, DIALANG cannot manipulate learner features, although it does strive to use task types, frameworks, and schemata which would be familiar to language learners. The CEF discussion of task conditions concentrates on interaction and spoken production, which are not covered in the DIALANG system. In addition to the text and task definitions in the CEF, we used frameworks such as Rost (1990) and Pihko (1996), which attempt to organise textual features into hierarchies or combine a

number of textual or task features into dimensions of difficulty. None of the existing frameworks combine all the dimensions, but attempts at synthesis were found more helpful for the creation of a multi-faceted set of tasks than plain lists of features, such as in the CEF.

The CEF discussion of text types lists different *sources* for texts, for instance books, magazines, leaflets, or personal letters for written texts, and announcements, lectures, entertainment shows, or telephone conversations for spoken discourse. While texts can be so categorised, this classification does not help characterise the nature of language use in the text. Yet, if discourse form and typical patterns of textual organization are taken into account, texts can be grouped into categories which share common textual features. This was the type of textual categorisation which was needed in DIALANG, and instead of the CEF, we chose to use the typology of Werlich (1976, 1988) as the primary source for classifying text forms and types. The primary division in his framework is into text forms, of which there are five: descriptive, narrative, expository, argumentative, and instructive texts. These are further concretised at two levels. For instance, descriptions can be impressionistic or technical, and narrative texts creative or factual. Concrete examples for these are texts like travel accounts (impressionistic description), reference books (technical description), stories or jokes (creative narration), and biographical notes or news (factual narration). Werlich's framework helps identify forms of texts that are structured differently, and hence contain different challenges for comprehension. The three levels of concreteness help define the construct assessed. When selecting texts for reading comprehension, the DIALANG task writers were thus asked to check that the five main categories in Werlich's classification were covered in each of the initial DIALANG task pools.

Like the CEF, the DIALANG section for grammar uses the traditional division of grammatical phenomena into morphological and syntactic features. However, the specifications also encouraged the task writers to consider contextual and functional variation in contextualising the items. A joke, a dictionary definition or a brief news clip can provide the context for a grammar item equally well as a short statement. This should provide variety of language use in the grammar section even if the categories tested are traditional.

The CEF definition of lexical competence identifies two main categories: lexical and grammatical elements, or open and closed word classes. Within lexical elements, a distinction is made between fixed expressions of several words and single word forms. The closed word classes include articles, quantifiers, demonstratives, personal pronouns, question words and relatives, possessives, prepositions, auxiliary verbs, conjunctions, and particles. Of these, the fixed expressions and single word forms were considered most clearly "Vocabulary" in DIALANG. Regarding the closed word classes, the DIALANG language groups were allowed to choose which section, vocabulary or grammar, these classes should belong to. Most teams placed them in the Grammar section, following traditional textbooks in their language as a foreign language.

However, the DIALANG project did not consider the division into fixed expressions and single word forms detailed enough for describing the nature of a learner's vocabulary knowledge. The assessment of fixed expressions was important, but the category of "single word forms" could include anything. The CEF categorisation was thus amended with the help of the assessment development teams' joint experience in teaching foreign languages. The different kinds of word knowledge which they distinguished were: word meaning, semantic relations, combinations, and word formation; each with further sub-divisions. Such a categorisation did not pre-exist in any single source, but was a synthesis from several sources.

The categories were considered to assess not only breadth of vocabulary but also depth of word knowledge. Furthermore, potential for feedback was sought. If a learner had done poorly on the category for word formation, for instance, the identification of this category as a problem area might encourage him or her to pay attention to compounding and affixation mechanisms for the language which he or she is learning.

4. CEF and the DIALANG self-assessment statements

A central feature of the DIALANG system is its capacity to function as more than a conventional proficiency test. The provision of feedback to test-takers was identified as an essential component of an effective diagnostic system. Additionally, in view of the increased recognition of the role of self-awareness and the contribution of meta-cognitive dimensions in language learning (e.g. O'Malley & Chamot, 1990; Oxford 1990; Oscarson, 1997), self assessment was identified both as a vital element and an innovative feature of the system. A review of research findings, undertaken as part of project activities, confirmed the assumptions about self-assessment on which the diagnostic system had been planned. However, integrating self-assessment into a multi-lingual objective measurement system was expected to be somewhat problematic as it would inevitably involve complex equating procedures.

The decision to use the CEF provided a firm grounding for the project, since the work of North (1995) contained an extensive bank of descriptive proficiency statements in English which had been subject to extensive trialling, validation and calibration. A key question was how to select from and convert these descriptive statements into "I can" type statements without changing the essential properties of the original formulations; and further how to render them across the other thirteen languages covered by the project. It was acknowledged that very careful attention needed to be devoted to assuring comparability between the fourteen language-specific sets of statements. There were also some additional conceptual problems of how to integrate and use self-assessment in the materials clients would meet and how to report the results of both the self-assessment and the objective measurement parts of the test.

It was decided that the element of self-assessment should play the following roles in the DIALANG assessment procedure:

- As a placement tool, along with the Vocabulary Size Placement Test, as described in Section 1 above: these instruments together would provide an initial estimate of person ability and allow for the presentation of objective items at an appropriate level of difficulty
- As a means of stimulating clients to reflect on their own strengths and weaknesses
- As a means of providing feedback to clients about the accuracy of their own self-assessment in relation to the information provided by the objective test tasks administered

However, in order to put these principles into practice a number of development stages were required. Firstly, statements needed to be selected, edited and adapted from those calibrated in North's research to fit the "I can" format required for self-assessment.

The selection of statements was based on piloting and calibration data. Only items of a certain model fit value were selected for adaptation and new trialling. A further requirement was that

they were judged to be clear and concrete enough for self-assessment by the DIALANG target group (including inexperienced learners at low levels of proficiency). In a few cases new statements had to be designed in order to cover the full range of skill areas and levels in the assessment system.

The editing and adaptation procedure also entailed breaking down some of the longer statements into two shorter statements and, where necessary, simplifying the language used. At the A2 level (Waystage), for instance, the statement for Reading:

A2: *Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday language or job-related language* (CEF, Chapter 4.4.2.2).

was split into two simpler statements for trialling:

A2: *I can understand short, simple texts written in common everyday language*.
A2: *I can understand short simple texts related to my job*.

And the statement for Listening:

A1 level (Breakthrough), Listening:
Can understand everyday expressions aimed at the satisfaction of simple needs of a concrete type, delivered directly to him or her in clear, slow and repeated speech by a sympathetic speaker

was adapted for DIALANG trialling to:

A1: *I can understand everyday expressions dealing with simple and concrete everyday needs, in clear, slow and repeated speech.*

Standard formulas such as *directly to me* were removed in order to keep the statements as short as possible.

Secondly, translation of the statements was required. The adapted English language statements were audited by Brian North, who found that they were, in general, closely matched to the original descriptors. The statements then needed to be rendered in equivalent versions across all the other thirteen languages involved. An initial plan to use the technique of translation and back translation to help ensure equivalence was abandoned in favour of an approach in which each national team leader and two native experts produced independent translations. These were then checked for correspondence with the original English version and merged into a master manuscript. Further, the national translations were compared line by line and checked for quality across languages leading to recommendations for changes where necessary. Finally each language team produced the final versions of the statements to be piloted.

Thirdly, the statements were allocated to separate pilot test forms, ensuring some overlap of statements between forms. This stage aimed to provide data to enable the equating of self-assessment and objective measurement instruments. A range of standard setting procedures were also implemented in order to anchor the instruments to the CEF levels. (For some details of the standard setting procedure in DIALANG, see the chapter by Kaftandjieva and Takala in this volume, and Kaftandjieva and Verhelst, 2000.).

The results of a first round of trialling using Swedish, English and Finnish language statements covering three skills (reading, listening, and writing) proved very successful (see the chapter by Kaftandjieva and Takala in this volume). Most of the more than a hundred statements employed were calibrated successfully, and the analyses show that the self-assessment results are reliable enough to enable the pre-estimation of learner proficiency. They can thus be used in deciding the approximate level of difficulty of the first DIALANG items to be presented to learners.

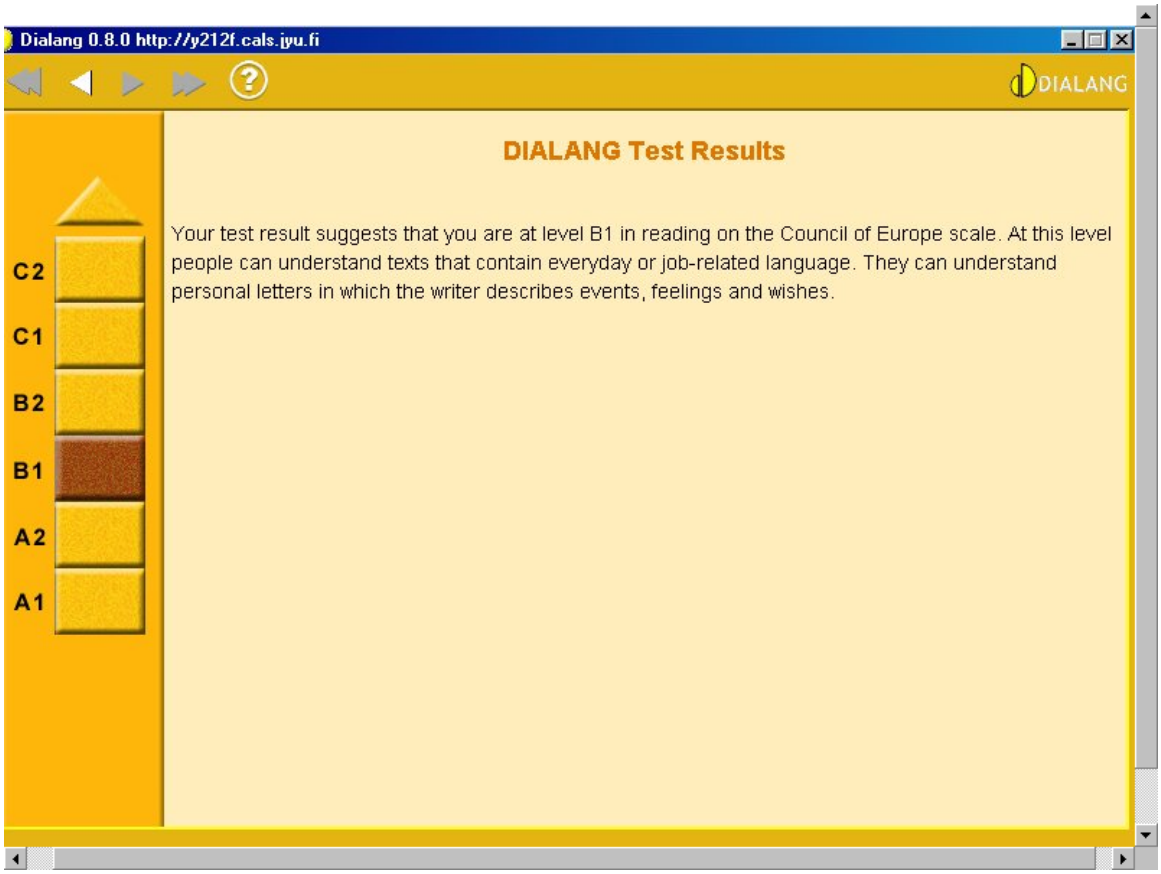
5. CEF input for the feedback available from DIALANG

Feedback in DIALANG consists of several sections. The users get the test result, i.e. an estimate of their level in the skill they have selected. They are also informed if their self-assessment in the skill differed from the test result and are given the option of viewing explanatory feedback about possible reasons for the mismatch between self-assessment and test result. Furthermore, the users can review all the items they have taken. The screen shows which items they answered correctly and which they got wrong, and they can access each item to see how they answered and what the correct responses were. They also get an estimate of their vocabulary size based on their VSPT responses. Finally, the users can view advisory feedback which contains further information on their level of proficiency and advice on how to improve their skills further.

The Common European Framework was used mainly in the test result section and in the advisory feedback, but the explanatory feedback was also influenced by the CEF to some degree.

The test result is reported to the user on the Council of Europe six-level scale from A1 to C2. The result screen consists of two parts: firstly, a picture of the scale showing the six levels and the level assigned to the user, and secondly, a verbal description of what learners can typically do with the language at that level of proficiency (see Figure 3 which gives the description which a learner estimated to be at level B1 in Reading would get. Note that the layout and colours of this sample screen from the DIALANG Prototype may not exactly match those in the operational system. The descriptions are based on the CEF scale descriptions (see CEF Chapter 3).

Figure 3. DIALANG Test Results Screen (from the Prototype)



The other section of feedback that is closely based on the CEF is the advisory feedback which is available for reading, writing and listening. (It is not available for grammar and vocabulary because the CEF does not contain enough material on those aspects of language to draw on.) The advisory feedback offers the users descriptions of what typical learners at the different levels of proficiency, in the skill in question, can do with the language (op. cit.). Compared with the descriptions on the test result screen, these accounts of proficiency are more detailed and are based on the scales in the CEF and also on certain specific sections in it (e.g. Chapter 4.1.3). To illustrate the descriptions, an example (level A1 for Reading) is provided below

	A1
What types of text I understand	Very short, simple texts, typically short, simple descriptions, especially if they contain pictures. Short, simple written instructions e.g. short simple postcards, simple notices.
What I understand	Familiar names, words, basic phrases.
Conditions and Limitations	Single phrase at a time, re-reading parts of text.

Figure 4

These extended descriptions of proficiency levels divide into three parts. The first part lists the kinds of texts that a typical learner at the particular level can understand (or produce, in

the case of writing), the second characterises the nature of comprehension (or writing) at that level, and the third explains the conditions and limitations under which the learners can typically use language. The users can see not only the description of the level assigned to them but also the level immediately below and above that level. Thus, the users can easily compare the different levels and may be able to see more clearly where they are on the ability continuum.

In addition to the descriptions of proficiency levels, advisory feedback offers concrete advice to the users on how to improve their reading, writing or listening from their level to the next level up. Sections in the CEF on language learning and teaching methodology (section 6.4 in particular) were used as a basis of some of the advice. This advice relates to the development of grammatical, sociolinguistic and pragmatic competences, to study skills and heuristic skills, to communicative language processes, and to language learning and language use strategies). The more general advice includes exhortations to engage in communicative activities with authentic (or graded) written materials, radio and TV broadcasts and with native speakers of the language studied. The more specific suggestions advise the learners to make use of, for example, the visual elements of texts in order to understand, to take notes or to keep a diary to learn more effectively, or to pay attention to how politeness is expressed in particular situations.

Another component of DIALANG feedback, the explanatory feedback, is also based on the CEF but at a more general level and with less concrete material drawn directly from the CEF. The purpose of the explanatory feedback is to give the users some possible reasons if there has been a mismatch between their self-assessments and test results. It also includes sections which explain what information language tests and other assessment instruments can give to language learners and how language tests differ from real life language use.

The CEF encourages the development of awareness in language learners about different aspects of their learning and it is with this aim that the explanatory feedback is constructed: it aims at increasing the learners' awareness of their language proficiency, including their own view of its level and adequacy for different purposes, and of a number of factors that relate to language use and learning. Such factors include, for example, the context and purpose of language use, the current and target level of proficiency, and a balanced vs. imbalanced use of language skills. Some of these factors are also described in the CEF, although much of the content is derived from other sources. Similarly, there are points in the section about language tests in the explanatory feedback which make some use of the CEF – mainly the points about the factors which affect performance in different situations and contexts (e.g. time pressure, possibility to prepare for the task, physical conditions) (see CEF, Chapter 7.3.2).

The current DIALANG system does not "know" why a particular user's self-assessment differs from his or her test result. In the future, however, it might be feasible to develop e.g. a questionnaire charting the users' language learning background, which might enable the system to direct the users towards the most likely reasons for the mismatch. The current system simply presents the user with a number of potential reasons.

Finally, the Vocabulary Size Placement Test result is reported as a score with 1,000 as the maximum. The scores are divided into six levels, each with a brief description of what a learner at that level can typically do in the language.

6. Summary and conclusions

Above all, the Council of Europe Framework influenced the DIALANG project in terms of its overall approach to assessment. The system is intended for self-directed language learners who are interested in knowing about their language ability. It encourages users to assess their own skills, and provides feedback on the accuracy of the learners' self assessment. It also provides a range of feedback which learners hopefully find useful in planning their future learning, but they can also choose not to view detailed feedback if they so wish. Thus, the rationale for the DIALANG assessment system supports the Council of Europe ideals of self-directed language learning.

The concrete definitions in the text of the CEF were used in DIALANG in three main ways. Firstly, the CEF concepts of language ability and language use, defined primarily in Chapter 2, formed a theoretical starting point for the DIALANG project. All tests need a theoretical framework, and the CEF provided DIALANG with one. Secondly, the inventories of communicative tasks, themes, activities, types of texts, and language functions in the CEF were drawn on to specify the concrete form of the DIALANG Assessment Framework (DAF). Reference to the CEF was also made in the DIALANG Assessment Specifications (DAS). Thirdly, the CEF concept of language proficiency levels was adopted as the DIALANG scale. This means that the DIALANG system reports its scores in terms of the CEF scale from A1 (Breakthrough) to C2 (Mastery), and the self-assessment statements and the verbal feedback from the system are largely based on descriptors from the CEF.

Chapter 9 of the CEF proposes that the parts of the CEF which can be particularly useful for assessment are the definitions of contexts of language use, conditions and constraints, communicative tasks and purposes, communicative themes, and text dimensions in Chapter 4, and the scale definitions and scales regarding communicative activities, communicative language competences, and strategies in Chapter 3, 4 and 5. DIALANG used all of these and operationalised significant proportions of the concept of language ability defined in the CEF.

It is envisaged that both individual language learners and educational institutions will find uses for the information and feedback that DIALANG provides. For example, language teachers at adult education institutions often find it difficult to place new students onto courses of an appropriate level. Initial feedback from teachers who have been informed about DIALANG indicates that there is a considerable amount of interest in the system in these institutions.

The DIALANG assessment system will continue to develop in several ways, some of which may also shed some further light on the validity and usefulness of the CEF. One of the key areas for development in the future is the feedback system. Which feedback is most useful, and for whom? Which forms of presenting feedback are intelligible and effective? Research on how learners interact with and react to the feedback offered by DIALANG has already begun.

References

Alderson, J.C. (2000) Technology in testing: the present and the future System, Volume 28, pages 593-603

Alderson, J.C., Clapham, C. and Wall, D. (1995). Language Test Construction and Evaluation. Cambridge: Cambridge University Press.

Bachman, L. and Palmer, A. (1996). Language Testing in Practice. Oxford: Oxford University Press.

Council of Europe (1996) : *Modern languages: learning, teaching, assessment. A Common European Framework of reference*. Draft 2 of a framework proposal. CC-LANG (95) 5 rev.IV, Strasbourg.

Frase, L. (1997). Technology for Language Assessment and Learning. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (Eds.) Current Developments and Alternatives in Language Assessment. Proceedings of LTRC 96. Jyväskylä: CALS/University of Jyväskylä. Pp. 519-543.

Lynch, B.K. and Davidson, F. 1994. Criterion-Referenced Language Test Development: Linking Curricula, Teachers, and Tests. TESOL Quarterly, 28,4:727-734.

North, B. (1995). The Development of a Common Framework Scale of Language Proficiency. Based on a Theory of Measurement. PhD Thesis. Thames Valley University, England

O'Malley J.M., & Chamot A.U. (1990). Learning Strategies in Second Language Acquisition. Cambridge: Cambridge University Press.

Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In *The Encyclopaedia of Language and Education,* Vol. 7 (Language Testing and Assessment). Dordrecht: Kluwer Academic Publishers, 175-187.

Oxford R.L. (1990). Language learning strategies: what every teacher should know. Boston, Mass: Heinle and Heinle.

Appendix 1. Sample page from the DIALANG Assessment Framework

DIALANG Assessment Framework Appendix 1

SPECIFICATIONS RELATED TO COMMUNICATIVE TASKS AND PURPOSES IN THE COUNCIL OF EUROPE'S THRESHOLD AND VANTAGE LEVEL PUBLICATIONS

August 1997 / November 1999

This appendix relates to the section of communicative functions and purposes (chapter 4.2,) in the Council of Europe Common Framework, draft 2, 1996. It lists the corresponding sections in the Threshold and Vantage documents. For ease of comparison, the same categories are presented side by side. The main differences between the specifications for the two levels are highlighted by using **bolding**. (N.B. Waystage does not specify communicative tasks and purposes in detail so this document includes the specifications found in the Threshold and Vantage level documents only.)

THRESHOLD LEVEL (1990)	**VANTAGE LEVEL (1996)**
1. Practical transactions	1. Practical transactions.
Learners will be able to cope with transactional situations of everyday life requiring **a largely predictable use of language.**	Learners will be able to cope with transactional situations in everyday life. At Vantage Level, learners will be able to deal more flexibly with these situations than at Threshold Level when they are **problematic or take an unexpected turn.** With enriched language resources (especially a wider vocabulary), learners will be able to express their needs and intentions **more precisely**, with less (though still some) need for compensatory strategies.
1.1. Contacts with officials	1.1. Contacts with officials
In all contacts with officials learners would be able to ask for repetition, clarification and explanation etc. of any information, questions or documents not understood, and should be able to ask for the services of an interpreter and/or legal adviser in case of serious difficulty (chapter 12).	N.B. In all contacts with officials learners would be able to ask for repetition, clarification and explanation etc. of any information, questions or documents not understood, and should be able to ask for the services of an interpreter and/or legal adviser in case of serious difficulty.

CHAPTER 10: A CEF-BASED SELF-ASSESSMENT TOOL FOR UNIVERSITY ENTRANCE

Brian North, Eurocentres Foundation[6]

Introduction

This chapter reports on the development for the University of Basle of a self-assessment instrument related to the Common European Framework (CEF). This project is one of several follow-ups to a Swiss Science Research Foundation project which developed the illustrative descriptors for the CEF Common Reference Levels (North this volume, 1996; 2000; North & Schneider 1998). The main follow-up project in Switzerland is the development and trialling of a European Language Portfolio (See Lenz and Schneider, this volume). The Portfolio is an instrument for learners to use in order to self-assess their language proficiency, reflect on language learning experiences, and, if they wish, present samples of work. The project reported on in this chapter can be interpreted as an investigation of the feasibility of a cut-down Portfolio designed for self-placement at university entrance. The full Swiss Portfolio contains checklists for each CEF level, which learners are invited to use for a more detailed self-assessment than that offered by the more holistic grid which appears in the Passport section of the Portfolio (Table 2 from Chapter 3 of the CEF). This project aimed to exploit the CEF descriptors to produce a single checklist focussed around the level deemed suitable for university entrance in Switzerland (Level B2).

1. Aims of the project

The project in fact had several aims in two different areas:

a) Operational: provision of a practical self-assessment tool for the University of Basle;
b) Research: validation of the CEF scale of descriptors in one specific context, and illustration of the way in which the set of CEF descriptors can be further developed to suit local circumstances.

1.1 Operational: The need for a self-assessment tool

The lack of detailed information about the language level and language needs of in-coming students at Swiss universities is a matter of some concern. Possession of the "Matura" qualification gives the right to attend any Swiss university, regardless of the language region of the university in relation to the mother tongue of the individual. In addition, English is only now becoming a compulsory subject in some of the cantons since national languages have traditionally been given preference. Therefore there is no way of knowing what English language proficiency individual in-coming students have, though English may be vital to their

[6] The Eurocentres Foundation (www.eurocentres.com) is a Swiss-based not-for-profit educational foundation subsidised by the Migros Federation of Cooperatives. Eurocentres offer language learning stays abroad for 7 languages in regions where the languages concerned are spoken. Eurocentres offer language learning stays abroad in a quality, acquisition-rich environment and have acted as consultants to the Council of Europe since 1968.

studies. It would also be seen as unconstitutional to give in-coming students a compulsory test, as they have a right to attend if they have a "Matura."

The main aim of the study reported in this chapter was to provide the university faculties, the students and the university language service departments with a picture of how incoming students see their communicative language ability. This could then be used to discuss and negotiate language support if necessary. Richterich and Schneider (1992) pointed out that the provision of a Common Framework defining different categories of communicative language competence at a series of ascending levels offered the opportunity to profile achievement in terms meaningful to all the different partners in or users of an educational service. and thus achieve "transparency." In the current project, the "partners" are the students arriving at university, the university language service departments, and the university faculties. Clarity of communication between partners achieved by verbal definition in this way is referred to in Switzerland as "transparency."

A related aim was to provide the means to report back to the upper secondary sector the perception their ex-pupils have of their practical language ability. The main reason Switzerland has been so interested in the development of the CEF and Portfolio is the issue referred to here as the "Schnittstelle" problem (cross-over problem). Each educational sector tends to have a narrow perspective focused on the part of the proficiency spectrum that is of immediate relevance. Within that relevant range of proficiency, practitioners tend to focus on certain competences and activities at the expense of others. Upper secondary language education still tends to be somewhat traditional, with a focus on grammar and literature at the expense of fluency and practical language skills. As part of an effort to address this problem, Switzerland is in the process of adopting a Gesamtsprachenkonzept (Babylonia N4/1998), in which Common Reference Levels from the CEF for different receptive, interactive and productive language activities defined in the CEF will be adopted as targets for different sectors. Aggregated results from skills profiles self-assessed at university entrance offers a way for universities to inform the upper secondary sector about the gap between what in-coming students felt they could actually *do* in terms of practical language skills, and what the university felt they should be able to do. This might inform dialogue, and, over time, help the language learning provision in different sectors to fit better together. The quality that parts of a system fit smoothly together is referred to in Switzerland as "coherence."

The operational purpose of the instrument, therefore, was to contribute to improving transparency and coherence in the Swiss educational system (Richterich and Schneider 1992) as regards language needs and language skills at university entrance.

To achieve this end, the instrument had to be simple and practical. The European Language Portfolio, being trialled in Switzerland as in other countries at the time, was felt to be impractical for this purpose. The Portfolio is intended to be consulted and completed over a period of time. What was needed here was a questionnaire which could be explained, and then completed for several languages in the space of less than an hour. The model was a self-assessment instrument produced in 1992 by the author for Credit Suisse. That instrument had an initial holistic self-assessment onto a scale of 5 levels, followed by self-profiling on some twenty sub-scales describing different work activities.

Although funds were available for the initial development and validation of the instrument reported on in this chapter, it was important that the questionnaire would have a simple key

transforming scores into a reporting scale, and the CEF levels. It would be this key that would be used on future occasions.

The provision of such a key or "transformation table" to convert raw scores on a reporting scale is a common feature of tests which are drawn from a validated item bank. The point about an item bank is that the "difficulty value" of every single item is known. That means that scores obtained from any selection of the items presented in a test can be converted to a score on a reporting scale. For example, Eurocentres have small item banks for each language taught, which directors of studies can use to create tests (Jones 1993; North 1994). These tests come with a "transformation table" converting scores on the particular test to scores on the common reporting scale. This reporting scale can be bands of proficiency, like the CEF Common Reference Levels, or it can be a scale of numbers like 1-1,000. Since the CEF descriptors each have an exact scale value in the item bank, exactly the same principle can be applied to them. In other words, scores from responses to "I can ….." statements presented as questionnaire items to learners can be converted to a reporting scale of 1-1,000. The questionnaire can be provided with a "transformation table," saying what 12 out of 20 (for example) means in terms of the reporting scale. That can give a precise answer to the question: "What does it mean if I tick 12 of the 20 items on this questionnaire?"

A fundamental aim of the project was to produce and validate such a transformation table so that the questionnaire could be used operationally in future years.

1.2 Research

As well as the "real-world" aim, the project was also an opportunity to contribute to the validation of the CEF descriptors. Would the scale values of the CEF descriptors produced through teacher assessment of foreign language competence in the lower secondary, upper secondary and adult sectors (North 1996; 2000) prove stable when applied to self-assessment at one particular "Schnittstelle" (upper-secondary/university)?

An ancillary aim of the study was see how mother tongue competence would be rated onto the CEF Common Reference Levels. Level C2 of the CEF is not intended to imply native speaker, "educated native-speaker" or the "well-educated native-speaker" of the American Interagency Language Roundtable (ILR) scale (WENS: see Lowe 1985; 1986). It has been accepted for two decades in the language testing literature that native-speakers vary widely in their language knowledge and in their ability to use that knowledge (Carroll, J.B. 1979:22-23). In addition to reinforcing this lack of uniformity, research in Australia and Great Britain during the development of IELTS (International English Language Testing Service) pointed out that native-speakers in upper secondary colleges seemed to achieve only the Level 6 or Level 6.5 required by British universities from overseas students (C1 in CEF terms). In reading tests, for example, from three groups, (postgraduates, lecturers, junior lawyers) the only people who reached the IELTS "Expert User" - a thinly disguised WENS - were the lawyers (Hamilton et al 1993). Native-speaker post-graduates and university lecturers had a better, more consistent performance than the upper secondary students - but were not "Expert Users" How would German-speaking Swiss upper secondary school leavers rate their mother tongue competence, using the descriptors adapted from the CEF? Would they rate themselves C2?

Finally, a further aim of the project was to describe categories of communicative activity particularly relevant to university study which had not been included in the CEF descriptor

bank, and to adapt formulations on existing scales to the particular context. Since the CEF descriptors were developed as an item bank (North, this volume), the existing descriptor bank could be extended in the same way that a test item bank can be extended. The project described in this chapter is a case study of the way in which this can be undertaken.

2. Self-assessment context

In a Swiss context, self-assessment of language ability at university entrance is attractive because the administration of a placement test to Swiss citizens would be unconstitutional. French-speaking, Italian-speaking and Romansch-speaking citizens have a legal right to study at a German-speaking university like Basle, provided they have the Matura qualification.

Self-assessment has also been implemented successfully in other contexts in Switzerland. The instrument produced for Credit Suisse which was referred to above has been used successfully with new employees in combination with tests since 1992. The resulting profile is input to a discussion on language training needs between new employees and the language unit in the personnel department. In another project involving self-assessment, Eurocentres sent young long-term unemployed Swiss on sponsored 12-week stays abroad at Eurocentres schools between 1994 and 1999, in collaboration with the Federal Office for Industry, Trade and Work. Before the course, these learners self-assessed their level on a holistic scale of 10 defined levels. Correlations in the range of 0.70-0.78 were consistently obtained between these self-assessments and placement assessment averaged from a test and an interview. In another study in a Swiss setting, Wilson and Lindsey (1999) report a correlation of 0.746 (n=937) between self-assessment on an adapted version of the American ILR (Interagency Language Roundtable) scale and results on the TOEIC (Test of English for International Communication: Educational Testing Service).

Other surveys of self-assessment also suggest that well-developed instruments can achieve respectable correlations to language tests and teacher assessments (Oscarson 1978; 1984, 1988; Le Blanc and Painchaud 1985; Blanche 1990; Ross 1998). Spolsky summarises the situation by saying that it has been shown that learners are able to assess their own language proficiency with reasonable accuracy if two basic conditions are met. Firstly it is important that there is no special encouragement to give inaccurate answers, as would be caused by a desire to please, or by rewards and penalties. Secondly, the responses required must be about aspects of language proficiency which are within the experience of the answerer (Spolsky 1992: 36). Ross concludes after a meta-analysis of 60 reported studies:

> "The results of the meta-analysis concur with the Blanche and Merino (1989) conclusion that self-assessment typically provides robust concurrent validity with criterion variables. The close-up examination of the process of self-assessment, mediated by variation in direct experience in language learning tasks, suggests that the degree of experience learners bring to the self-assessment context influences the product. We might assume that when the criterion is one that does not invoke episodic memory, the self-assessor may have to rely on a recollection of his or her own general proficiency to make the assessment. It is perhaps at this point that the methodological artefacts of self-assessment are most likely to interfere with the process. Subjects may also resort to relativity, or be influenced by self-flattery." (Ross 1998: 16)

Spolsky's and Ross's comments raise two important issues. Firstly, even if one were able to identify with any degree of accuracy specific activities which students of different faculties

would be called upon to perform, their very lack of experience of such specialised activities would lead them to guess—or extrapolate from their view of their overall proficiency—and thus to give inaccurate, inconsistent responses. Secondly, as Spolsky points out, self-assessment brings the danger of response effects. Self-assessments at university entrance will be inaccurate if respondents are being asked to volunteer as candidates for compulsory, time-consuming pre-sessional training which does not count towards their grades.

The University of Ottawa gives an example of this. In 1984 the university introduced Le Blanc and Painchaud's (1985) self-assessment instrument as the sole means of placement in pre-sessional language training courses to prepare for the university's language admission test. The instrument is a questionnaire on which learners use a five-point scale from:

I cannot do the task at all to *I can do the task all the time*

to respond to descriptors of listening and reading ability presented in rank order of difficulty. One example is:

I can take a French newspaper and understand the gist of the stories on the front page.

Ready-Morfitt (1991) describes how initial success with this approach changed after the self-directed enrolment was replaced in 1987 by placement into a specific class at one of 4 levels—of which only the top two levels gave academic unit-credits. Comparison with standardised tests and with tasks linked to the wording of the self-assessment statements led to the conclusion that there was a substantial degree of over-assessment. The students tended to rate themselves as good enough for the credit-awarding language courses at the higher levels. To avoid such response effects in this study, students remained anonymous when they completed the self-assessment instrument.

3. The measurement context

The descriptors in the CEF have precise scale values in an "item bank" of descriptors. These "difficulty values" were obtained through a statistical analysis of teachers' ratings of students in their classes (See North, this volume). As with the Ottawa questionnaire mentioned above, the teachers in the CEF study rated on a five-point scale for each descriptor. The same five point (0-4) rating scale used in the CEF study (North this volume) was again used in the study reported here. As in the CEF study, the rating scale was explained on the first page of the questionnaire, and repeated in shortened form at the top of each page of the questionnaire. . The statistical methodology used in the study here was also identical to that used in the CEF study: the Rasch model, from the Item Response Theory (IRT) family of measurement models. Since approximately half the descriptors used on the questionnaire in this study came from the CEF, and since the same methodology was used for the analysis, the results from this study could be reported on the scale produced in the original CEF study. That is to say, results for learners were reported onto the CEF scale, and new descriptors were scaled in relation to the existing CEF scale.

The great advantage of IRT (Item Response Theory) over CTT (Classical Test Theory) is that the results reported are not restricted to the original group of candidates and the original test administration, provided new groups can be considered part of the same test population. Therefore, once the questionnaire has been developed and validated, the "transformation

tables" will be valid for new groups of students at Basle, provided there is no radical change in the student population.

4. Development and trialling of the instrument

The development of the instrument took place between March and July 1999 and is described in detail in Beranek (1999). The instrument was designed to have two forms of self-assessment:

- Section 1: Holistic self-rating: This involved a holistic self-assessment onto the CEF Common Reference Levels. Originally the CEF self-assessment grid of levels by skills (Table 2) was going to be used for this purpose. This self-assessment grid is also used in the Passport section of the European Language Portfolio. The particular version of the grid had six rather than the traditional four skills, since Reception, Interaction and Production were each divided into spoken and written. With holistic rating, the user decided which level (A1 – C2) most matches what they consider their ability to be, in this case for each skill.

- Section 2: Analytic self-rating: This section comprised a series of 3-level mini-scales, for what was thought to be the most relevant range of level on the proficiency spectrum: Level B1+, Level B2 and Level B2+. Each of the 5 skills had three categories taken from the CEF. For example, for spoken production: Describing Experience, Sustaining a Point of View, and Addressing an Audience. 18 language activities (6 skills x 3 categories) with descriptors at 3 levels each would give 54 descriptors. Each descriptor was to be rated separately, using the 0-4 rating scale from the Swiss study which produced the CEF. The result would be a score out of 216, which could be converted to a position in a CEF level, plus a profile across skills in terms of scores and CEF levels.

4.1 General or Specific?

There was considerable discussion on the question of whether the second section should be general or specific. Should it ask the students to profile their current language competence on more general tasks, or should it ask them to diagnose their future faculty-specific needs by rating themselves on more specialised descriptors of tasks which were particularly relevant to different faculties? It was thought that if specialised versions were produced, there would be at least two (Humanities / Natural Sciences) and at most four or five (e.g. Humanities, Law, Natural Sciences, Medicine, Language & Literature).

A major problem with a specific approach was that it would imply knowing what tasks students in different faculties needed to be able to perform for their studies. The lack of information about the language needs of university students studying different disciplines was referred to earlier. Waters (1996: 52) comments that even in relation to the need for English at English-speaking universities, no systematic co-ordinated programme of enquiry has been conducted in North America to complement the British research of the early 1980s (e.g. Hawkey 1982, Weir 1983). The same can apparently be said for universities in the German-speaking world (Lüdi, personal communication). A study is underway at Basle to remedy this deficiency. The study has adopted an ethnographic methodology to try and get over the problem that what university faculty *say* the students need appears to match at most tangentially to the activities they are observed to undertake (Mondada et al. forthcoming). The

methodology involves observing and recording classes in different faculties over a considerable period of time to obtain a balanced sample of the kind of behaviour that occurs in each faculty concerned. The data collected is then analysed, and categories of activity identified, defined and illustrated.

Since this ethnographic study had not reported at the time of this project, a conventional survey of expert opinion was conducted, since as Waters (1996: 54) recommends: "valuable information about EAP needs, at all levels, can also be obtained by surveys of expert opinion." In this survey, Beranek (1999) interviewed two professors and two students from five faculties: Medicine, Physics, Economics, Law and French. Each interviewee was asked to:

- (Section 1) use the CEF Table 2 self-assessment grid to profile their own competence (students) or to profile the current and the desired competence profile of incoming students (professors);
- (Section 2) rank the list of 18 language activities (three for each of the 6 skills) in order of priority for their faculty; then suggest specialised versions of those 18 activities and/or other language activities not on the list which would be particularly relevant to students in the faculty.

The interviews were recorded and analysed in some detail as Beranek (1999) reports. The assessments with the CEF self-assessment grid confirmed that the range of level from B1+ to C1 was the relevant one. Results in relation to Sections 2 were disappointing, however, perhaps due to the small number of respondents involved. Whilst it was sometimes possible to see trends in the ranking exercise, there were also serious contradictions between what two professors said, or between what two students said, or between what the pair of professors and the pair of students said. Furthermore, even when there appeared to be a consensus, this seemed counter-intuitive or tended to be immediately contradicted by other faculty consulted less formally. These fundamental contradictions in the opinions of the experts it had been possible to consult made the development of specialised checklists unfeasible.

This result should perhaps not be seen only as a disappointment. The idea that language training is most efficient when it is subject-specific is a demand-led view held by many companies, but has little basis in research or expert opinion in at least the English Language Teaching profession. In the UK, there has been a marked move in both pre-sessional training and in language testing away from an LSP approach (language for specific purposes) towards a more general skill-based LAP approach (language for academic purposes) (Waters, 1996 48-50). What was once an academic argument in relation to the limited effect of subject-specific texts in the IELTS revision project (Clapham 1996) has now reached the general English teaching public in LAP supplements to the EL Gazette and Guardian Weekly.

There is also the proviso defined by Spolsky for effective self-assessment. The responses required must be about aspects of language proficiency which are within the experience of the answerer (Spolsky 1992: 36). Since these students were only just arriving at university, they had not yet had such experience.

It was therefore decided not to develop subject-specific descriptors for Section 2. Instead, CEF descriptors were reworded where necessary in order to slant them towards the academic context (e.g. listening to lectures rather than presentations; taking part in seminars rather than meetings, writing papers rather than reports). In addition, descriptors were developed for

relevant areas not covered in the CEF, such as Note-taking, Presenting to an audience, and Essay-writing..

4.2 Scale or Grid?

The second major problem encountered in development concerned the form of the holistic assessment in the first section. Trialling the draft questionnaire with a small group of students showed that the CEF self-assessment grid (6 levels x 6 categories = 36 boxes) was just too complex for the task at hand. Most students would be rating themselves for three languages (English, German and French); some for more. Completing the whole questionnaire for all languages was supposed to take an hour. Some students took the whole hour just working on the self-assessment grid.

The grid was therefore replaced with the holistic scale of Common Reference Levels (Table 1 in the CEF), but expanded to include the "plus levels" between the Common Reference Levels, which had been defined in the Swiss research project (North 1996; 2000).

Replacement of the self-assessment grid (CEF Table 2) by a holistic scale (CEF Table 1) now meant that the concepts Reception, Interaction and Production (spoken / written) had not been introduced in the first section. The grid had offered an overview of what these categories meant. It seemed inappropriate to confront students in the second section with an unfamiliar form of categorisation. Therefore it was decided to base Section 2 on the 4 skills, which would be familiar to the students, plus interactional competence (innovative in an upper secondary context) and qualitative factors like grammatical accuracy and socio-cultural aspects (traditional in an upper secondary context). After some discussion, it was also decided to move *Fluency* from the qualitative factors to Speaking, and to move *Information Exchange* from Interaction to Speaking (previously Spoken Production) in order to replace *Describing*, which had been given a very low priority in the interviews.

4.3 Final Version

Therefore the revised instrument finally consisted of two parts defined as follows:

Part 1: Language Biography: information about languages spoken, and how they were learnt, together with a holistic self-assessment for all these languages, using the holistic CEF scale (Table 1), expanded to include "plus levels."
Part Two: Language Profile: self-assessment for High German, French, English and any language which was the subject of study, on 17 sub-categories of language use, grouped into 6 language areas as shown in Table 1 below:

LISTENING SKILLS:	Listening in an audience Understanding TV and media
READING SKILLS:	Reading for orientation Reading for detail Reading to process information
SPEAKING SKILLS:	Fluency Information exchange Putting a case *Presenting to an audience
WRITING SKILLS:	*Note-taking *Essay writing
INTERACTIVE SKILLS:	Exchanging opinions in discussion Informal conversation Taking the floor (Turn-taking strategies) Working with a group (Co-operative strategies)
LANGUAGE KNOWLEDGE:	Grammatical accuracy Situational appropriacy

Table 1: Language areas and Categories of language use for analytic rating in Part Two:

There were descriptors for levels B1+, B2 and B2+ for each of the 17 categories presented in order of difficulty, giving a total of 51 descriptors. 27 of these 51 descriptors were self-assessment versions of CEF descriptors, the others were either newly formulated or so heavily edited that they could no longer be assumed to share the statistical properties of the originals. The three categories above which are starred (Presenting to an audience; Note-taking; Essay writing) all had new descriptors. The three descriptors for Note-taking, for example, are given in Table 2.

German original	English translation	CEF Level	Difficulty Value
Ich kann während eines klar strukturierten Vortrages über ein mir vertrautes Thema einfache Notizen zu auffallenden Punkten schreiben, auch wenn ich die Tendenz habe, mich auf die sprachliche Ebene zu konzentrieren und deshalb gewisse Informationen nicht erfasse.	I can understand a clearly structured lecture on a familiar subject, and can take simple notes on points which strike me as important, even though I tend to concentrate on the words themselves and therefore miss some information.	B1+	0.35
Ich kann während eines Vortrages ausreichend präzise Notizen für den Eigengebrauch machen, sofern das Thema mit meinen Fach- und Interessengebiet zu tun hat und die Rede klar und strukturiert ist.	I can take notes during a lecture which are precise enough for my own use, provided the topic is within my field of interest and the talk is clear and well-structured.	B2	1.04
Ich kann während eines Vortrages detaillierte Notizen über Themen in meinem Fach- und Interessengebiet schreiben und darin Informationen so originalgetreu festhalten, dass sie auch für andere nutzbar sein könnten.	I can take detailed notes during a lecture on topics in my field of interest, recording the information so accurately and so close to the original that the notes could also be useful to other people.	B2+ (high)	2.69

Table 2: Descriptors for Note-taking

The third column in Table 2 gives the level the descriptors were intended to be. The fourth column shows the difficulty values they turned out to have on the mathematical scale underlying the CEF levels. These descriptors turned out to be well targeted since they were found to have difficulty values within the range of the scale for the levels intended.

It was possible to calibrate these new descriptors to the CEF scale because approximately 50% of the 51 descriptors (27 in fact) were used in a formulation virtually identical to that used in the original Swiss project which created the scale (North this volume). These 27 CEF descriptors were used to "anchor" the two data sets together.

5. Subject

Over 700 students completed the questionnaire in supervised sessions a few weeks after arrival at the university. All were asked to rate themselves for High German, English and French, as well as any other languages they were studying. Data from the questionnaires was entered into an Excel sheet. Languages other than German, French and English were formed into groups (e.g. Nordic) for analysis. Each language rated by each learner was a separate line of data. Thus, a learner rating themselves for High German, English, French and Italian produced 4 lines of data, henceforth referred to as "person/languages." Since each of the 782 students completed the questionnaire fully for between 3 and 6 languages, this produced a total of 2,447 person/languages, each able to be separately identified in the analysis.

6. Data analysis

Analysis was undertaken with the same Rasch rating scale model used in the original Swiss project (North 1996; 2000). During the first few analysis runs, 126 person/languages (approximately 5%) were temporarily removed from the data because of inconsistent,

improbable responses. This is a routine step in a Rasch analysis to improve the quality of the measurement. Their removal left 2321 person/languages in the analysis.

Two separate sets of analyses were then done:

a) global – with all the data
b) separate analyses for each of the 6 language areas (Listening, Reading, Speaking, Writing, Interactional Competence, Language Knowledge).

The global results were at the borderline of acceptability for this kind of analysis, applying conventional quality criteria used in a Rasch analysis to check that the statistical model is being applied appropriately. 11.1% of the data showed statistical "noise," as opposed to the conventional criteria of 10%. However, only 1.7% had "standardised residuals[7]" equivalent to .05 significance, suggesting very "plausible" results despite the amount of "noise." The "noise" referred to above is caused by an element of multi-dimensionality: dimensions are being measured together which might be better analysed separately. Not surprisingly, therefore, the "noise" virtually disappeared when the six areas were analysed independently. This suggests that student profiles across the six areas can be expected to be slightly more accurately measured than the aggregate result being reported in the global analysis.

7. Reporting scales

A Rasch analysis produces ability estimates presented on a "logit" scale. A logit scale traditionally has zero in the middle, with lower results expressed as minuses. In this study, since half of the items had been anchored to their values on the logit scale underlying the CEF Common Reference Levels, results were reported on that CEF logit scale. To make this scale more comprehensible to users, it was converted into a scale of whole numbers so that the maximum score on the full questionnaire would give 1000 on the reporting scale. Cut-offs on the CEF Common Reference Levels are then as shown in Table 3.:

[7] Standardised residuals are an expression of factors in the measurement situation that have not been accounted for in the analysis: what is left over. Residuals are reported for each item and for each person. A high number of items or persons with a high residual would mean that the statistical model was not appropriate for the data. Residuals are reported in standard deviations away from what the model expects. Thus a residual of 2.0 is at approximately 0.05 significance – and considered an implausible result.

1	**2**	**3**	**4**	**5**	**6**
	Score on Full Questionnaire	**CEF Logit Score**	**CEF Level**	**Band**	**Basle Reporting Scale**
Proficient User	*204*	*7.37*	***Max***	11th	*1000*
	190	6.36	Ambilingual		932
	152	3.90	**C2**	10th	**766**
	133	2.80	**C1**	9th	**691**
Independent User	113	1.73	B2+	8th	616
	94	0.72	**B2**	7th	547
	76	-0.26	B1+	6th	482
	58	-1.23	**B1**	5th	**416**
Basic User	41	-2.21	A2+	4th	349
	25	-3.23	**A2**	3rd	**278**
	8	-4.29	**A1**	2nd	**208**
	1	-5.40	Tourist	1st	130

Table 3: Global transformation table: Cut-scores to convert questionnaire raw scores to reporting scales

The second column shows raw scores on the questionnaire, the maximum being 204 (51 items x 4). Each raw score is reported as a score on the logit scale underlying the CEF levels. The cut-scores between CEF levels are shown in Column 3. The "plus levels" reported in the study in which the CEF scale was developed (North 1996; 2000) are used since this is a linear scale of bands with approximately equal intervals. There are actually 11 bands (Column 5). The cut-score on the Basle reporting scale out of 1,000 (Column 6) is an arithmetical conversion from the logit score.

The highest scorers on the questionnaire are situated in the 11th band. The vast majority of these scores are from Swiss-German native-speakers rating their High German. However, some logit scores in this 11th band were recorded for non-native speakers in this study and in the original Swiss project (North 1996; 2000). This level of proficiency represents a very high level of ambilingual competence, not necessarily native-speaker (mother tongue) competence. Speakers in this score range recorded on video in the original Swiss project are very articulate and fluent, basically bilingual speakers - but not native-speakers.

Scores for individuals should not be interpreted too literally. Any test or questionnaire survey involves standard error, which is a factor of the length of the instrument and the targeting and quality of the items. On this questionnaire, standard error is about 0.25 logits, approximately 15 points on the 1-1000 scale. Therefore the reporting score "932" really implies a score between 907 and 959.

A further complication is that very high or very low scores are actually known to produce distorted results with this statistical model (Camilli 1988; Warm 1989; Jones 1993; North 1996; 2000). This distortion affects results over 80% or under 20%. On this particular questionnaire, distortion of results for high scorers does not really matter because they are already well into Level C2, the top CEF level, before this distortion starts. The distortion just means that some of the people reported as being above C2 are really "only" C2. However, for low scorers, results reported onto CEF levels may be underestimated. Accurate, undistorted reporting starts at Level A2+. Below that level, the most one can really say is that people are

either A1 or A2. But this is not a very serious problem in an instrument designed for assessing language ability at university entrance.

Similar charts transforming raw scores into scores on a reporting scale and onto the CEF levels were produced for each of the 6 language areas. Because the number of items involved for one area is much smaller, however, the standard error for any individual (factor of test length and item quality) is considerably higher when they are assessed on just one of the 6 areas.

8. Results

The project had both operational and research aims, as described at the beginning of the chapter. Results are discussed in relation to each below.

8.1 Operational

The main aim of the project was to provide an operational instrument to self-profile the language skills of incoming students in order to inform faculties and language departments, and give feedback to the feeder sector: upper secondary. The two main aspects involved here were (a) the credibility of the picture of the proficiency of incoming students provided, and (b) the feasibility of the questionnaire as an operational instrument.

Credibility of the self-assessment: The picture provided by the questionnaire came from two sources: (a) a simple record of self-assessment of overall ability for each language, using the holistic overview of the CEF levels (CEF Table 1) presented in Part One, and (b) the results produced by the statistical analysis of the self-profiling in relation to the more detailed descriptors for the 6 areas and 17 sub-categories presented in Part Two.

The correlation between these two very different forms of self-assessment is interesting. The overall correlation between all results on Part One (holistic) and all responses on Part Two (analytic, by areas) was **0.843** (n=2417). However, this overall correlation masks quite different correlations for different languages, as summarised in Table 4:

Language	Correlation	Subjects
High German	0.49	n=778
English	0.74	n=781
French	0.76	n=760
Italian	0.94	n=43
Spanish and Portuguese	0.43	n=25
Slavonic	0.98	n=8
Nordic	0.4	n=6

Table 4: Correlation between holistic rating onto CEF Table 1 in Part One, and the analytic rating on separate tasks in Part Two

Let us leave aside High German for the moment, as this concerns assessment of mother-tongue. The results for groups of languages with small numbers of respondents (Slavonic; Nordic, Spanish/Portuguese and Italian) are curiously balanced but not very significant. The correlations for the two main foreign languages English and French, for which a response was obligatory, on the other hand, are both around 0.75. This is a level of correlation one comes across quite frequently in a comparison of two different tests. Both the studies involving a

comparison of self-assessments to test results with Swiss learners which were reported at the beginning of this chapter reported correlations of this size. A recent comparison of the TOEIC and the TOEFL based on a sample of around 300 learners also reported a correlation of 0.75 between the two tests (TOEIC Service International. The Chauncey Group International 1999). North (1991) also reported repeated correlations around 0.75 between UCLES First Certificate results and global assessment for English on the Eurocentres scale. In other words, this level of correlation is not uncommon between two assessments with different perspectives on a similar construct.

In this case, the two perspectives *are* very different. Presented with a vertical scale of defined levels (A1-C2) in the holistic self-assessment in Part One, respondents are invited to say whether their competence is good, weak or middling. The wording helps them to refine this feeling in choosing one particular band rather than another, but the question they are answering is essentially: "How good is your French/English/etc?" They will be influenced by how long they have learnt French or English, what kind of school grades they got, what qualifications they have. Presented with a series of very focussed descriptors defining performance in practical tasks in French/English/etc, as with the 51 descriptors in Part Two, respondents are asked to say whether they can do these things with ease, or only in favourable circumstances. The questions being asked were: "What can you actually *do* with your French/English/ etc. in different circumstances? Can you do this? Can you do that?"

It is therefore not surprising that the results for many people from two different forms of self-assessment should be different. The relationship between the two self-assessments for the vast majority of learners seems plausible, but with some interesting differences in the distribution. Figure 1 gives the number of respondents at different bands of the reporting scale for French. The analysis of the questionnaire (left) produced the 11 bands shown previously in Table 3. These 11 bands are the numbers 1-11 on the horizontal axis. The number of respondents in a particular band is reflected in the height of the bar to the left of that number. The holistic assessment (right) has only 9 bands, starting with 2 (A1), because the expanded version of CEF Table 1 which presented the holistic scale did not offer the bottom or top bands produced in this analysis and in the original research project: North 1996; 2000). C2 is the 10th and highest band possible.

French: Questionnaire **French: Holistic**

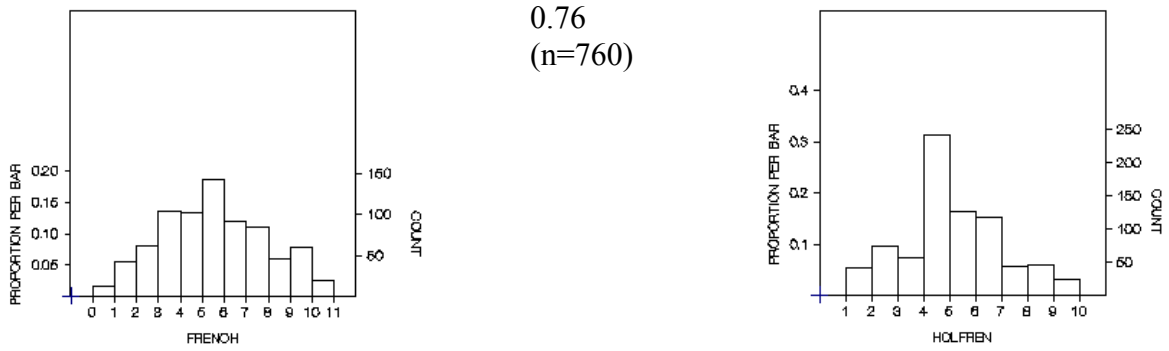


Figure 1: Analytical (Questionnaire) and holistic (CEF Table 1) self-assessments for French

On the right (holistic), one can see that about 250 or approximately one third rated themselves as being in the 5th band (B1: Threshold Level) on the holistic CEF scale. About 50 considered themselves in the 4th band: A2+, and about 130 rated themselves as the 6th band: B1+ and as the 7th band: B2. In the analysis shown on the left, 150 are placed at B1+ (6th band), which is slightly more than in the holistic assessment, and just under 100 are placed at both B2 and B2+ (7th and 8th bands). About 100 are also placed in both B1 and A2+ (4th and 5th bands).

In other words, students faced with a holistic scale are tending to select B1, or B1+ and B2, whereas the analysis, based on their responses to descriptors at B1+, B2 and B2+, is spreading them out more, with the median at B1+. To put this another way, the analysis of self-ratings by language areas produces a very regular bell curve, whilst the holistic self-assessments suggest a pronounced peak at B1 (Threshold Level).

The picture for English (Figure 2) is similar. The analysis gives a more differentiated picture than the holistic impressions, and again suggests students are slightly under-estimating themselves when presented with a holistic scale of levels.

Based on their holistic impression (right), the vast majority of students claim an overall level of English above Threshold Level. The holistic results shown a marked tendency to select the 5th-7th bands (B1-B2) with most selecting B1. However, the analysis (left) based on what respondents say in relation to tasks at B1+, B2 and B2+, shows some 550 of the 780 spread reasonably evenly between the 6th and 10th bands (B1+, B2, B2+, C1 and C2). The largest number, just under 150, are at B2+. This level could be said to represent a good pass at Cambridge First Certificate, or about a 6.0 on the IELTS (International English Language Testing System) test used by British, Australian and American universities.

English: Questionnaire **English: Holistic**

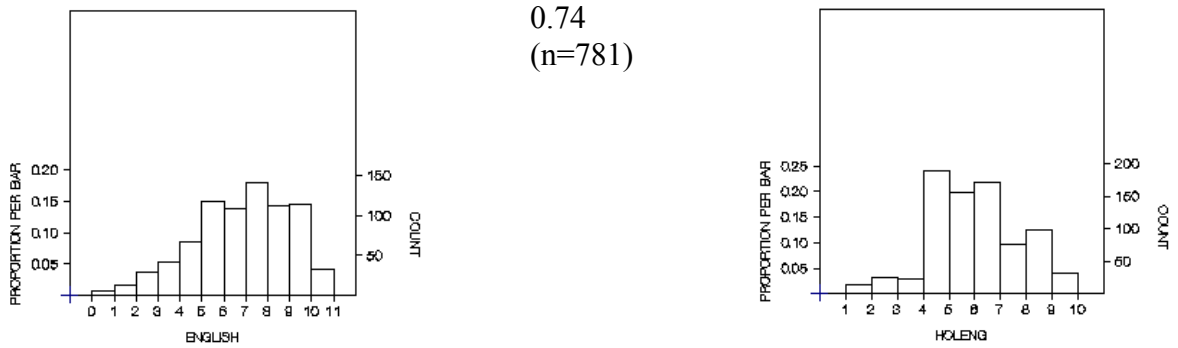


Figure 2: Analytical (Questionnaire) and holistic (CEF Table 1) self-assessments for English

Unlike responses for French, which produced a fairly regular curve, responses for English are skewed to higher levels. In both the holistic and the analytic assessment, students in general are more confident about their English than about their French.

One should add that there are some individual cases which just do not make sense at all. How can someone rate themselves initially as A1 or A2 and then claim to be able to fully understand lectures, summarise articles and develop arguments in discussion? Yet there are cases in which this happens. Which is then "correct?" There is no simple answer. For the context in which the self-assessment is being used, namely profiling language ability relevant at university entrance, an objective aggregate across the different skills (Part Two) assessed carefully can perhaps be considered more relevant than an initial impression.

As regards High German, the fact that the correlation between the two forms of assessment is so much lower at 0.49 is a reflection of the fact that, since these judgements are nearly all about mother tongue, we are dealing with a truncated range of scores around C2. The score ranges are actually very similar, as shown by Figure 3.

German: Questionnaire **German: Holistic**

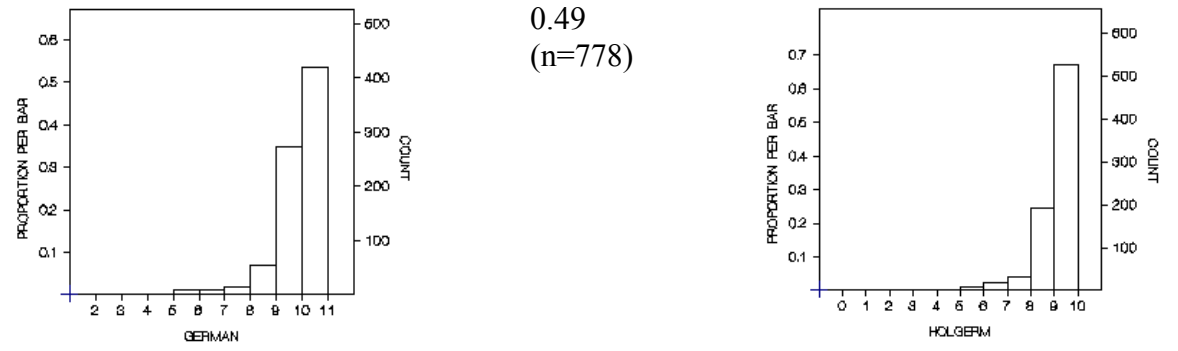


Figure 3: Analytical (Questionnaire) and holistic (CEF Table 1) self-assessments for High German

Conversely, once all the mother tongue ratings are added in to the foreign languages, this produces a longer range of scores, which boosts a correlation. Therefore, although the correlation of the two assessments for High German is only 0.49, adding responses for High German into the analysis increases the overall correlation from around 0.75 to 0.843.

What happens in the analysis of High German, as illustrated in Figure 3, is that many of the native-speakers are identified as being above Level C2, which was their highest choice in the holistic rating, and many who had avoided the top category, C2 and put C1, are identified as C2.

Operational Feasibility: The administration of the instrument proved feasible, in that some 780 students could be processed in batches as part of their intake procedure. The question was whether the questionnaire would be able to be used in future without a statistical analysis each time. As mentioned at the beginning of the chapter, one of the advantages of a Rasch analysis is the capability of reporting results with "transformation tables" whose values will remain stable as long as the test population remains stable. Unlike in classic test analysis, validity of results is not restricted to the group of subjects in the analysis.

The availability of a small set of independent data from a different source offered the opportunity to establish that the reporting table would prove valid for different but similar groups. A group of 178 learners arriving at institutes of higher education in Basle other than the university also completed questionnaires. Their data was not included in the results reported above.

The correlation between the two forms of self-assessment (initial holistic self-assessment in Part One; aggregate reported from Part Two) for these 178 students was 0.852, very similar to the 0.843 reported above for the main data set of 780. The proportion of students at each level for different languages was similar to that discussed above in relation to the main set.

To all intents and purposes the results for these 178 students mirrored those of the students in the original data set of 780. This would suggest that the questionnaire can be used as an operational instrument without the need for further statistical analysis. But collating and calculating the results using the transformation tables took only two hours rather than three days. This is a dramatic increase in the feasibility of the instrument.

8.2 Research Results

One ancillary aim of the study was to see if the difficulty value of CEF descriptors would remain stable when applied in a separate analysis to one specific educational sector. In fact the scale values for the 27 items close enough to their CEF formulations to be regarded as the same items correlated 0.899 to their scale values in the original research project (North 1996; 2000). This is a high correlation when one considers that:

- the descriptors were used for self-assessment in this project rather than for teacher assessment as in the CEF project;
- the descriptors were used for rating mother tongue ability as well as foreign language ability;
- the descriptors represented one short range of proficiency (B1+ to B2+).

A second research aim was to see if the CEF descriptors were suitable for people rating their mother tongue competence. As mentioned above, the scale values of the CEF descriptors remained remarkably stable, and this is despite the fact that approximately one third of the ratings concerned mother-tongue competence. The fact that not all of the students identified in the analysis as Level C2 or above chose C2 for their holistic impression is almost certainly an example of a well-documented rater error called "central tendency." When faced with a scale, untrained raters show a marked trend to avoid the top and bottom categories. This is a common human failing and a problem with holistic scales, rather than a problem with the CEF descriptors.

Finally, the project aimed to develop some new descriptors which might describe areas that had not yet been scaled in the CEF. As mentioned earlier, the three descriptors for note-taking turned out to be within the proficiency bands intended, and this mini-scale formed the basis for the scale for this area included in the revised CEF. Similarly, the descriptors for situational appropriacy formed the basis of a scale for Socio-linguistic appropriacy.

9. Conclusions

The replication of the results in the main data set used in the analysis (780 students), with those from the secondary data set interpreted with the transformation tables (178 students) suggests that the questionnaire is a reliable operational instrument. The correlation of around 0.85 between the two forms of self assessment in both sets of data is in the region accepted as adequate for concurrent validity (Criper and Davies 1988). At the time of writing, there has not yet been an opportunity to repeat the use of the questionnaire.. It is equally premature to discuss the impact of the instrument, since the profiles produced are still being studied.

What is perhaps of wider interest is the high stability in the values of CEF descriptors in an instrument consisting of 50% new formulations. In addition, the fact that all the new formulations landed in the bands intended, is an indication that *after* a detailed common reference framework has been established, well-crafted descriptors defining other aspects in terms of CEF levels can be successfully targeted. It was a design feature of the first Swiss checklists for the European Language Portfolio (Schneider and North 2000) that they should be open-ended. At the end of each section, space was left for teachers to add their own locally relevant descriptors. It is a feature of the CEF itself that it is open-ended.

The experience of this project suggests that provided a reasonable proportion of the descriptors on a checklist are derived from CEF descriptors, and provided the authors of new descriptors internalise the CEF levels, further development of descriptors can "work" from a measurement perspective as well as from an educational perspective.

References

Beranek, M. (1999): Travaux préliminaires pour un Concept Général des Langues pour l'Université de Bâle. Lizentiatsarbeit eingereicht an der Philosophisch-Historischen Fakultät, im Dezember 1999.

Babylonia (N4/1998): *Babylonia, A Journal of Language Teaching and Learning*: Die Zukunft der Sprachen in der Schweiz: das Gesamtsprachenkonzept. ISBN1420-0007. CH-6949 Comano, Fondazione Lingue e Culture.

Blanche, P. and Merino, B. (1989): Self-assessment of foreign language skills: implications for teachers and researchers. *Language Learning* 39, 313-40.

Blanche, P. (1990): Using standardized achievement and oral proficiency tests for self-assessment purposes: the DLIFLC study. In: *Language Testing* 7, 2, 202-229.

Camilli, G. (1988): Scale shrinkage and the estimation of latent distribution parameters. *Journal of Educational Statistics* 13, 227-41.

Carroll, J.B 1983: Psychometric theory and language testing. In Oller, J.W. (ed.) *Issues in language testing research*. Rowley (Mass.), Newbury House: 80-107.

Clapham, C. (1996): *The development of IELTS. A study of the effect of Background on reading comprehension*. Studies in Language Testing 4. Cambridge, Cambridge University Press.

Council of Europe (1996): *Modern languages: learning, teaching, assessment. A common European framework of reference. Draft 2 of a framework proposal*. CC-LANG (95) 5 rev IV, Strasbourg..

TOEIC Service International. The Chauncey Group International (1999): The TOEIC test and the TOEFL Institutional Testing Program (ITP). Information sheet.

Hamilton J, Lopes M., McNamara T., and Sheridan E. (1993): Rating scales and native speaker performance on a communicatively-orientated EAP test. *Melbourne Papers in Applied Linguistics* 2, 1, 1-24

Hawkey, R. (1982): An investigation into inter-relationships between cognitive/affective and social factors and language learning. Unpublished doctoral thesis, University of London.

Henning, G. (1984): Advantages of latent trait measurement in language testing. *Language Testing* 1, 2, 123-33.

Jones, Neil (1993): *An item bank for testing English language proficiency: using the Rasch model to construct an objective measure*. PhD thesis, University of Edinburgh.

LeBlanc, R. and Painchaud, G. (1985): Self-assessment as a second language placement instrument. In: *TESOL Quarterly* 19, 4, 673-687.

Linacre J.M. (1988): *FACETS: A computer program for the analysis of multi-faceted data.* Chicago, MESA Press.

Lowe, P. (1985): The ILR proficiency scale as a synthesising research principle: the view from the mountain. In James C.J., ed. *Foreign language proficiency in the classroom and beyond* National Textbook Co. Lincolnwood Illinois: 9-53.

Lowe, P. (1986): Proficiency: Panacea, framework, process? A Reply to Kramsch, Schulz, and Particularly, to Bachman and Savignon. *Modern Language Journal*; 70/4: 391-97 Win 1986

Mondada, L., Fünfschilling, J., Müller, K. and Pieth, C. (forthcoming): Die interaktive Konstruktion des Wissenschaftlichen Diskurses in mehrsprachigen Situationen. Basel. Romanisches Seminar.

North, B. (1991): Standardisation of continuous assessment grades. Alderson, J.C. and North, B. 1991: (eds.): *Language testing in the 1990s:* Modern English Publications/British Council, London, Macmillan: 167-177.

North, B. (1994): Itembanker: A Testing Tool for Language Teachers. *Language Testing Update* 16, Autumn 1994: 85-97.

North, B. (1996)*: The Development of a common framework scale of descriptors of language proficiency based on a theory of measurement*. PhD thesis, Thames Valley University.

North, B. (2000)*: The Development of a common framework scale of language proficiency*. New York, Peter Lang.

North, B. and Schneider, G. (1998): Scaling descriptors for language proficiency scales. *Language Testing* 15, 2, 217-262.

Oskarsson, M. (1978): *Approaches to self-assessment in foreign language learning*. Strasbourg: Council of Europe.

Oskarsson, M. (1984): *Self-assessment of foreign language skills: a survey of research and development work*. Strasbourg: Council of Europe.

Oskarsson, M. (1988): *Self-assessment of communicative proficiency*. In: Trim, John L.M. (ed.) (1988): *Evaluation and testing in the learning and teaching of languages for communication*. Strasbourg: Council of Europe (= Council for Cultural Co-operation), 46-58.

Ready-Morfitt, D. (1991): *The role and limitations of self-assessment in testing and research. Paper presented at the Language Testing Research Colloquium held in Princeton N.J., March 1991*. University Press

Richterich, R. and Schneider, G. (1992): Transparency and coherence: Why and for whom? In Council of Europe 1992: 43-49

Ross, S. 1998: Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing* 15, 1, 1-20.

Schneider, G. and North, B. 2000: *Fremdsprachen können - was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit.* Chur, Zürich: Verlag Rüegger.

Spolsky, B. (1992): The gentle art of diagnostic testing revisited. In Shohamy, Elana / Walton, A. R. (ed.) (1992): *Language assessment for feedback: testing and other strategies*. Dubuque: National Foreign Language Center, Kendall/Hunt, 29-41.

Warm, T. A. (1989): Weighted likelihood estimation of ability in Item Response Theory. In: *Psychometrika* 54, 3, 427-450.

Waters, A. (1996): A review into research into the needs in English for Academic Purposes of relevance to the North American Higher Education context. Princeton, New Jersey. *TOEFL Monograph Series*, MS-6, November 1996

Weir, C. (1983): Identifying the language problems of overseas students in tertiary education in the United Kingdom. Unpublished doctoral thesis, University of London.

Wilson, K. and Lindsey, R. (1999): Validity of global self-ratings of ESL speaking proficiency based on an FSI/ILR-referenced scale. *TOEFL Research Report* RR-99-13, Princeton, New Jersey, Educational Testing Service.

CHAPTER 11: RELATING THE ALTE FRAMEWORK TO THE COMMON EUROPEAN FRAMEWORK OF REFERENCE

Neil Jones, University of Cambridge Local Examinations Syndicate, EFL Division

Introduction

The ALTE Framework is a descriptive system of levels of language proficiency developed to promote the transnational recognition of certification in Europe. This purpose is clearly served by aligning the ALTE Framework with the Common European Framework. The value of the CEF is that it represents a serious attempt to provide a system of detailed level descriptions underpinned by an empirical measurement approach. The level descriptions provided by the ALTE Framework are in many ways complementary to those of the CEF, and to some extent both systems have made use of a similar statistical approach.

Aligning the two frameworks adds meaning and interpretability to both of them. At the same time, attempting to align the frameworks enables us to assess the methodology itself, and thus to identify issues arising in the empirical construction and comparison of descriptive proficiency scales, and possibly in the use that is made of them.

This chapter focuses on the latter topic. After providing necessary background to the ALTE Framework, it describes the empirical approach taken to aligning it to the Common European Framework, and discusses issues that arise.

1. The ALTE Framework and the Council of Europe Framework

ALTE – The Association of Language Testers in Europe – is an association of providers of European foreign language examinations. Registered in 1992 as a European Economic Interest Group (EEIG), it provides a context for transnational collaboration between some of the major international providers in the field of language testing.

ALTE recognizes the importance of the Common European Framework for establishing common levels of proficiency. From the outset, one of ALTE's main aims has been to establish such common levels, in order to promote the transnational recognition of certification in Europe, and much work has been invested in this in the context of the ALTE Framework Project.

Through the ALTE Framework Project ALTE members have classified their examinations within a common system of levels. (An up-to-date list of exams can be found on the ALTE website: *www.alte.org.*) Beginning with general language proficiency exams, the Framework is being extended to cover exams for specific purposes, such as business. The Framework is currently a five-level system, but work is proceeding to define a lower level (Breakthrough or A1 in Council of Europe terms).

Members' exams have been classified within the Framework through an extended process of comparison, content and task analysis, including reference to external specifications such as the Council of Europe Waystage and Threshold levels (Milanovic 1993, Milanovic and

Saville 1995, Milanovic *et al* 1997). In the new area of computer-based testing a group of ALTE members are collaborating on projects which have already produced tests in several languages that report results in ALTE levels. This work involves linking each test to a common measurement scale, and thus complements the existing analytical definition of the ALTE Framework.

The ALTE Framework aids interpretation. Relating examinations to a single common framework makes it easier to explain to end users of exam certificates, such as employers, exactly how to interpret that certificate. This makes qualifications more usable and thus increases people's potential mobility.

The ALTE "Can do" Project, which is introduced in greater detail below, is an important part of the work on the Framework. The "Can do" statements provide a comprehensive description of what language users can typically do with the language at different levels, in the various language skills and in a range of contexts (Social & Tourist, Work, Study). The "Can do" Project has a dual purpose: to help end users to understand the meaning of exam certificates at particular levels, and to contribute to the development of the CEF itself by providing a cross-language frame of reference.

The ALTE Framework and the Council of Europe Framework have much in common, and there are obvious benefits in relating them to each other in as simple and transparent a way as possible. This chapter presents an account of work that has been done to relate the ALTE and the Council of Europe Frameworks to each other, which involves the alignment of three scales:

1. The ALTE "Can do" scale, that is, a language proficiency scale defined through descriptions of typical abilities at each ALTE level;
2. The ALTE exam scale, that is, a language proficiency scale defined by performance in exams at each ALTE level. Relating the "Can do" and exam scales allows us to say: "If you pass an ALTE exam at Level 3 you will typically be able to do *x*, *y* and *z*."
3. The Council of Europe Framework.

Frameworks by their nature aim to summarize and simplify, identifying those features which are common to all language users and all situations of use, in order to provide a point of reference for particular users and particular situations of use. There is a balance to be sought here: a framework which is too high-level, abstract or general will be difficult to relate to any real-world situations, and will thus tend to be interpreted differently by anybody who seeks to make use of it. On the other hand, a framework which is too low-level, concrete and particular will run the risk of simply not corresponding to any real-world situation, because of the many ways in which these may vary.

The approach taken in this chapter, of taking a heterogeneous range of data and attempting to construct a single interpretative framework by fitting it to a particular statistical model, can be expected to throw some light on the limits of generalisability in building a descriptive framework for language proficiency.

2. The "can do" project

The "Can do" Project is a long-term ALTE development which has received European Union Lingua funding for several stages since its inception in 1992.

The aim of the "Can do" Project is to develop and validate a set of performance-related scales, describing what learners can actually do in the foreign language. In terms of Alderson's (1991) distinction between constructor, assessor and user oriented scales, the ALTE "Can do" statements in their original conception are user-oriented.

The "Can do" statements are multilingual, having been translated so far into thirteen of the languages represented in ALTE. These languages are: Catalan, Danish, Dutch, English, Finnish, French, German, Greek, Italian, Norwegian, Portuguese, Spanish, and Swedish. They constitute a frame of reference to which different language exams at different levels can potentially be related. They offer the chance to demonstrate equivalences between the exam systems of ALTE members, in meaningful terms relating to the real-world language skills of people achieving a pass in these exams.

2.1 Structure of the "Can do" statements

The "Can do" scales consist currently of about 400 statements, organised into three general areas: *Social and Tourist*, *Work*, and *Study*. These were judged to be the three main areas of interest for most language learners. Each includes a number of more particular areas, e.g. the Social and Tourist area has sections on *Shopping*, *Eating out*, *Accommodation* etc. Each of these includes up to three scales, for the skills of *Listening/Speaking*, *Reading* and *Writing*.

Each such scale includes statements covering a range of levels. Some scales cover only a part of the proficiency range, as of course there are many situations of use which require only basic proficiency to deal with successfully.

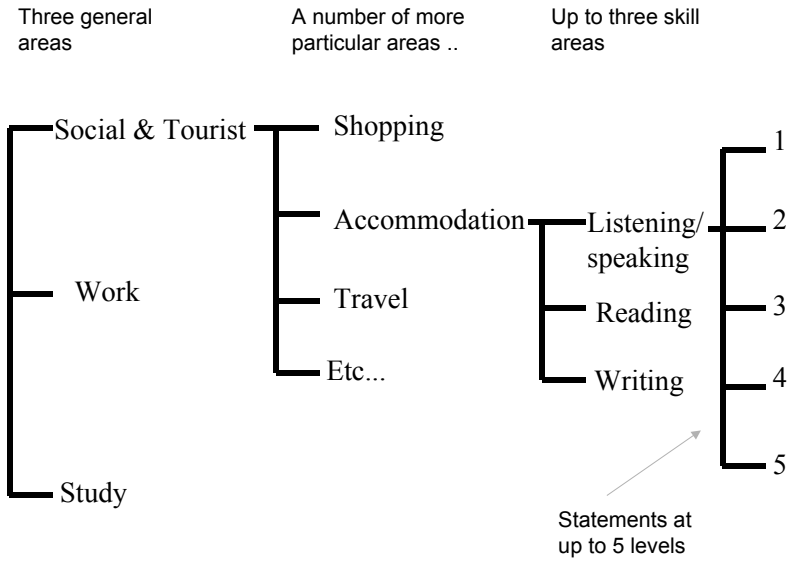


Figure 1: Structure of the "Can do" statements

The Social and Tourist area in particular draws on the Threshold level (1990). Table 1 gives an example of a "Can do" scale from the Work area.

Area:	Work
Activity	Requesting work-related services
Environment	Workplace (Office, factory etc)
Language skill	Listening/Speaking
1	CAN state simple requirements within own job area, for example "I want to order 25 of ….".
2	CAN ask questions of a fact-finding nature, for example establishing what is wrong with a machine, and understand simple replies.
3	CAN put her/his point across persuasively when talking, for example about a familiar product.
4	CAN give detailed information and state detailed requirements within familiar area of work.
5	CAN argue his/her case effectively, justifying, if necessary, a need for service and specifying needs precisely.

Table 1: Selected statements at Levels 1 – 5 from an example "Can do" scale

2.2 Development of the "Can do" statements

The original development process went through a number of stages (Milanovic and Saville 1995). Studies of ALTE users led to the specification of a range of candidate needs and major concerns. Draft statements were then trialled with teachers and students to evaluate relevance and transparency, and were moderated in an iterative process.

The scales as developed above were then empirically validated, in order to transform the "Can do" statements into a calibrated measuring instrument.

It is worth stressing that this is a major, long-term research undertaking. The construction and validation of such a large multilingual set of task descriptors requires a great deal of work, firstly of a qualitative nature, at the writing and translating stage. Empirical validation then requires the collection of a large amount of data. The aim is, after all, to construct a detailed, language-neutral description of functional levels of proficiency, which should apply equally well to the learning and use of foreign languages in all European countries, irrespective of which language the statements are expressed in, and which target language they are used to describe. This implies a two-dimensional matrix of first language by target language, with data available for each of the cells in the matrix. Evidently, some cells will be easier to fill than others, and it is not practical to collect equal amounts of data for each language; however, it is necessary to collect as wide a range of language data as possible. The project will also incorporate a growing number of languages, as new members and associate members translate and make use of the "Can do" statements.

A subsequent stage is to link the language exams of each ALTE member to the scale described by the "Can do" statements. Again, this process is expected to take a number of years, depending on the resources available to each ALTE member and the size, in terms of candidature, of their exams.

2.3 Data collection using questionnaires

So far data collection has been based chiefly on self-report, the "Can do" scales being presented to respondents as a set of linked questionnaires. Nearly ten thousand respondents have completed questionnaires. For about 4000 of these respondents, additional data are

available in the form of language exam results. These are mainly for Cambridge EFL exams. A limited amount of data has been collected for other exams, but is not discussed in this chapter.

The decision to use questionnaires to collect data on the "Can do" statements had several important consequences. Firstly, the text of the "Can do" statements had to be organised appropriately. This meant de-constructing the original paragraph-length level descriptions into a larger number of shorter statements, of generally no more than one sentence. Then different forms of response were trialled. The form finally chosen elicited a Yes/No response to each statement, using the instruction:

> Put ONE cross next to each statement. Tick YES if the statement describes your level, or if you "Can do" BETTER than this. Tick NO if you CAN'T do what is described because it is TOO DIFFICULT for you.

In other words, respondents were asked to respond negatively to those statements which described a task beyond their capacity, and respond positively to everything else.

By choosing a Yes/No form of response, we placed the onus squarely on the respondents to decide whether or not they could do the task in question. We might have used a scalar form of response, allowing respondents to express degrees of confidence in their ability to do something, but we were not convinced that this would provide more information, or indeed solve the problem of deciding what "Can do" actually means.

In the end the meaning of "Can do" is something we will need to define explicitly, in terms of the probability of someone having a certain proficiency level being able to succeed on certain tasks. This topic is discussed in section 3.3.2 below.

The use of self-report has implications for the relative difficulty of the "Can do" statements that make up the description of a level. This reflects entirely the perception of the respondents themselves – how they view their ability to manage in different situations. This goes, for example, for the relative difficulty of the language skills – speaking, reading, writing – or the relative difficulty of operating in a professional or a social setting.

The description of a level which emerges from combining all the statements that fall within that level will reflect this perception. If respondents find reading relatively easy, for example, it may be that the highest level description contains very few statements describing reading.

The descriptions of proficiency which result from this approach must thus be seen as normative and descriptive of the typical pattern of abilities of a particular group of respondents. They are not intended to be absolute, prescriptive statements of what a level "should" mean in terms of some set of criteria developed elsewhere. This has advantages if one considers how the ALTE Framework is intended to be used. A common use is to describe proficiency globally. An employer will advertise a post requiring, say, ALTE Level 3, because he expects that a person at that level will typically have the range of skills he needs. In designing general-purpose language proficiency exams, too, it makes sense to target typical profiles of ability – that is, ensure that the subtests testing different language skills are experienced as being at a similar level of difficulty by the "typical" candidate.

2.4 Comparison of "Can do" and CEF statements

It is possible to attempt empirically to link the ALTE and the Council of Europe Frameworks, because the CEF is itself originally based on empirical work (North 1996/2000, North and Schneider 1998, Council of Europe 2001).

The analysis reported here includes data from versions of questionnaires which contained statements taken from the first draft of the Council of Europe Framework document (Council of Europe 1996). Statements were taken from:

1. The self-assessment grid prepared for the European Language Portfolio, identified as "Table 7" in Council of Europe 1996,133, and as Table 2 *Common Reference Levels: Profile by language skills* in Council of Europe 2001. Hereafter referred to in this chapter as "Table 2"
2. 16 statements from scales relating to spoken interaction ("Fluency"), identified as "excellent items" in North 1996/2000, 405-9.

"Table 2" consists of five scales for Listening, Reading, Spoken Interaction, Spoken Production and Writing. It was chosen because of its wide use in the context of the European Language Portfolio. The "Fluency" statements had been found to have stable difficulty estimates when measured in different contexts in the Swiss project. It was expected that they would provide a good link between the ALTE "Can do" statements and the CEF.

These CEF statements represent the end product of an extended process of development, with each level description being a composite of carefully selected typical elements. Here is an example statement from "Table 2":

> I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works. (Writing, Level C2)

Here is an example "Fluency" statement:

> I can express myself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language. (Fluency & Flexibility, Level C1)

As noted above, the "Can do" statements in their present form are relatively short, atomic statements. This is an important difference which can be expected to complicate comparison.

3. Method and findings

3.1 Validation of the "Can do" statements

3.1.1 Structure of the response data

Questionnaires have generally been administered in the subjects' own first language, and so far mainly in European countries. Respondents have as far as possible been matched to appropriate questionnaires – the Work scales given to people using a foreign language

professionally, the Study scales to respondents engaged in a course of study through the medium of a foreign language, or preparing to do so. The Social and Tourist scales are given to other respondents, while selected scales from this area have also been included in the Work and Study questionnaires as an anchor.

The use of anchor statements is necessary to enable the relative difficulty of the areas of use, and particular scales, to be established (for more on this see6.4 below). The use of Social and Tourist scales as an anchor was based on the assumption that these areas call upon a common core of language proficiency and can be expected to provide a valid point of reference for equating the Work and Study scales. In the later data collection discussed in this chapter the role of the anchor was taken over by statements from the CEF illustrative scales.

The basic analytic approach is thus to bring all the data together into a single dataset, irrespective of the language in which the questionnaire was completed, and the target language which the respondent was describing (after initial analyses of the individual datasets to identify particular problems). This reflects the view that our best idea of the difficulty of any particular statement comes from averaging across all L1 and L2s in the data. Having found this best estimate, we can then seek to identify those statements that show differential item function – i.e. that particular groups respond to in a significantly different way. These group effects can be interpreted in relation to the global estimate of statement difficulty.

3.1.2 Initial validation

Empirical work looked initially at the internal consistency of the "Can do" scales themselves, the aims being to verify how individual statements function within each "Can do" scale, to equate the different "Can do" scales to each other, i.e. establish the relative difficulty of the scales, and to investigate the neutrality of the "Can do" scales with respect to language (Jones forthcoming). Main findings of this part of the work are summarized below.

The difficulty of individual statements was found to agree quite well with the original assignment to levels which was made when the statements were originally written ($r = .78$).

Where statements moved in difficulty this could often be explained by certain features of the statement. Features which generally made statements more difficult included the use of very specific exemplification, reference to a situation involving stress or heightened responsibility, and reference to a difficult channel of communication (e.g. the telephone). Features which generally made statements easier included generality or brevity.

Statements with explicit negative qualification ("CANNOT do …") were found to perform badly, in that higher-level respondents reacted in the opposite way to that intended. The CANNOT qualification describes a lower level of proficiency – that is, an easier task – but higher-level respondents were reluctant to endorse such statements, with the result that they were found to be relatively more difficult than intended for these respondents. This was not unexpected, and reflected a problem specifically with using negatively-oriented statements in a self-report questionnaire format.

Effects connected with first or target language, or other group effects, have been investigated as far as the present range of data permits.

It appears that a number of statements vary in difficulty according to the respondent's first language (or the language of the questionnaire). Some of these effects have been linked to problems of translation, but others are less readily explained.

There also appear to be significant differences between certain target language groups, where learners of certain target languages may be relatively more or less confident of their communicative as opposed to receptive skills. If confirmed, this would indicate that people learn languages for different reasons, not all connected with transactional use, and also that approaches to teaching particular languages may tend to place different degrees of stress on communication.

Work in this area will continue as more data become available.

An outcome of the first phase was a textual revision of the "Can do" statements, removing, in particular, statements with negative orientation, and adding new statements where a need was identified.

3.2 Relating the "Can do" statements to the CEF

In 1999 responses were collected in which the original anchor statements taken from the Social and Tourist area were replaced by statements taken from the Council of Europe Framework document (1996 edition). Two sets of anchor items were used:

1. the grid of major categories of language use by level identified as "Table 2";
2. 16 statements from scales relating to spoken interaction ("Fluency").

3.2.1 The anchor design

The anchor design is shown in Figure 2.

The diagonal sets of boxes in Figure 2 are the seven "Can do" questionnaire forms. They appear twice, once for the June and once for the December data collection. In June they each contained anchor statements taken from "Table 2"; in December these were replaced by the CEF "Fluency" statements.

In both June and December, some of the respondents were also candidates for Cambridge EFL exams, and their exam grades could thus be compared with their responses.

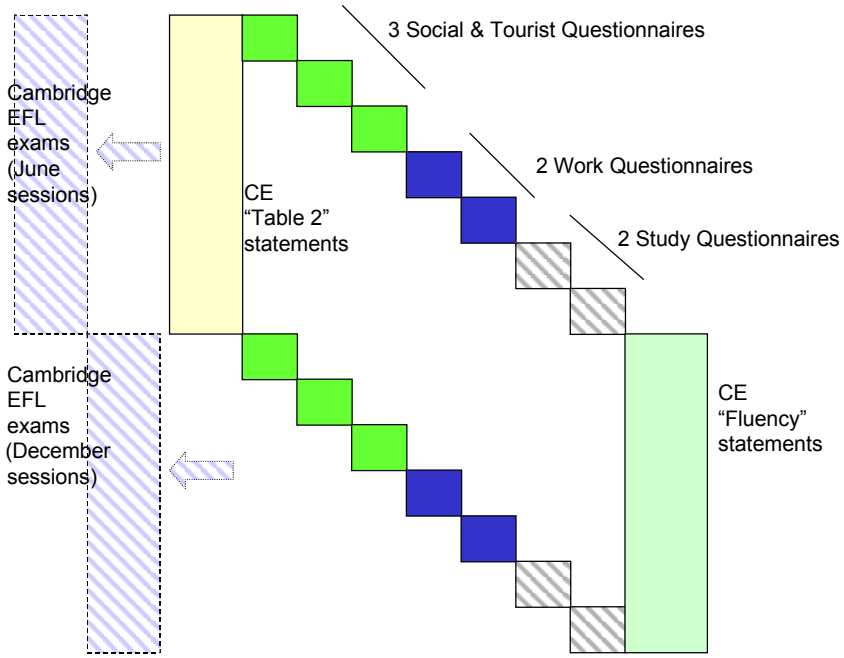


Figure 2: Anchor design for linking ALTE "Can do" questionnaires, CEF statements and EFL exams

Figure 2 shows that the links in the data run vertically and horizontally. Thus the "Can do" questionnaires are linked to each other only indirectly, by the horizontal links to the accompanying CEF statements. The CEF "Table 2" and "Fluency" statements are linked to each other indirectly, via the "Can do" questionnaires, which are the same in both December and June.

The anchor design illustrated here includes only a subset of the available data, as the earlier questionnaires with their Social & Tourist anchors are left out . Although all of these data are relevant to working out a final equating of the "Can do" scales, this chapter focuses on the more recent data and the link to exam grades and the CEF statements.

3.2.2 Cleaning the data: removing people and items

In Rasch analysis it is usual to try and improve the coherence of the data, and consequently the definition of the measurement scale, by removing misfitting responses. From separate analyses of the Fluency, "Can do" and "Table 2" sets about 350 misfitting respondents from the total 3,000 were identified and removed.
An investigation into the reason for misfit found that the only systematic effect related to age. As Figure 3 shows, it is respondents below the age of 18 that are most likely to misfit.

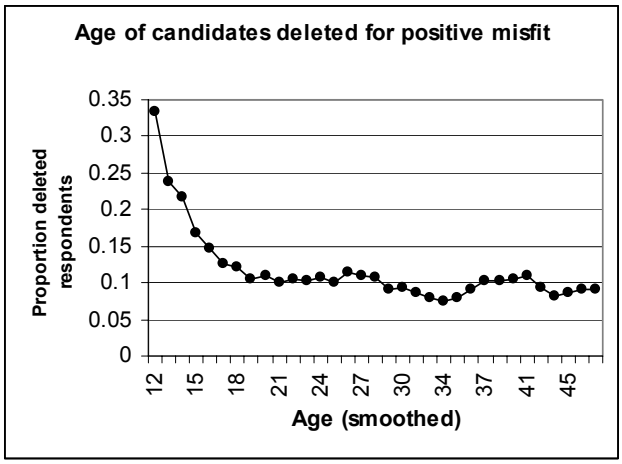


Figure 3: Age of deleted respondents (5-year moving average)

3.2.3 Linking "Table 2" and the "Fluency" statements

A separate analysis of the "Fluency" statements was run to check the correlation of the difficulties found with those given by North (North 1996/2000). The correlation was high (r= 0.97). Difficulties were also found to be fairly stable across respondent groups. Thus these statements should constitute a good anchor between the "Can do" statements and the scales used to illustrate the Council of Europe Framework. One problem is that in the current analysis the highest level (C2) statements are not well distinguished from the level below (C1). This may be a general effect in these data, based on self-ratings.

An analysis of the whole data set was then run, to estimate difficulties of the "Table 2" and "Fluency" statements on the same scale. The difficulty estimated for the fluency statements was compared with the difficulty given by North in order to derive parameters for equating to the CEF scale. The "Table 2" statements were then plotted against the CEF band cutoffs, as thus estimated, to compare the levels found with those intended.

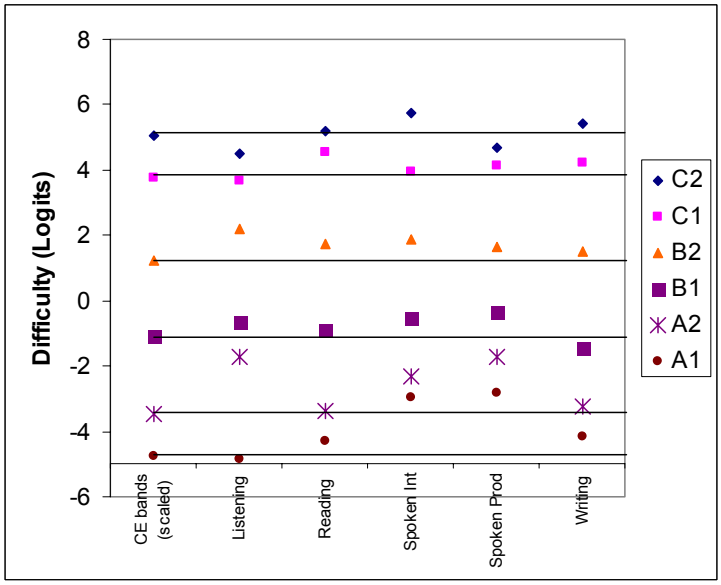


Figure 4: "Table 2" statements equated to the CEF via "Fluency" statements

Figure 4: shows the "Table 2" statements against the CEF level thresholds A1 to C2, as derived from the equating via the "Fluency" statements. The logit values on the vertical axis are shown as scaled to the metric of the present analysis.

"Table 2" contains five scales, each with one statement describing each CE level. It can be seen that there is generally very good correspondence between the intended level of these statements and that found in the analysis. Between A2 and C1, nearly all the statements fall in the intended level. At the highest level, respondents do not distinguish C2 so clearly from C1. At the lower end, two of the scales (Spoken Interaction and Spoken Production) do not have a clearly distinguished level A1.

3.2.4 Linking "Can do" statements to the CE scale

Having verified that the two sets of CEF statements agreed quite well with each other, the difficulties estimated for the "Can do" statements were then used to group them by CEF level. This tended to bunch them into the middle bands, indicating that the "Can do" statements in their present form operate on a shorter scale than the CEF statements. The "Can do" difficulties were therefore scaled to spread them more widely. The approach used was to find the spread of person abilities as estimated from the "Can do" statements alone, and also from the CEF statements alone. The ratio of these indicates the difference in discrimination of the two sets of statements, and can be used to re-scale the "Can do" statements. This spread out the "Can do" statements, approximating more closely the original assignation of "Can do" statements to ALTE levels. However, there remained fewer statements at the highest levels than in the original, analytical assignation of "Can do" statements to levels.

3.2.5 Linking the ALTE and CEF levels

An assumption was made that Levels A2 and B2 (Waystage and Vantage) could be taken to correspond to ALTE Levels 1 and 3. These points were chosen because the ALTE Level 1 exams are explicitly modelled on the CEF Waystage specification, and Vantage represents that rather well-understood intermediate level tested in English by the FCE exam.

The scale and the band cutoffs used in ALTE computer-adaptive testing projects were used as a working definition of the ALTE levels. The origins of this scale lie in work done over several years at Cambridge to establish the relative difficulty of UCLES EFL exams. The difficulty threshold of the passing grade on each exam (KET, PET, FCE, CAE, CPE) is taken as a provisional definition of the ALTE five-level system.

The ALTE and CEF levels were equated by scaling to the two reference points (A2 = ALTE 1, B2 = ALTE 3) identified above. This allowed the relative position of the other bands to be compared, and allowed the "Can do" statements to be grouped by ALTE level.

The two groupings of "Can do" statements (by ALTE level and by CEF level) were found to be very similar.

3.3 Relating the "Can do" statements and the CEF to ALTE exams

3.3.1 Relation between exam grade and self-ratings

The data contain a link between "Can do" self-ratings and grades achieved in UCLES EFL exams at different levels. This analysis focussed on the data collected in December 1999, which is linked to the versions of the "Can do" questionnaires which contained the CE "Fluency" statements. 478 candidates completed questionnaires, almost half (213) being candidates for FCE.

Figure 5 shows the mean self-rating of candidates grouped by the exam grade which they achieved. The exams are ordered by level (KET = ALTE Level 1, CPE = ALTE Level 5). The grades shown for KET and PET are: P = Pass, and M = Pass with Merit. For the other exams the grades run from A down to C, with C being a pass.

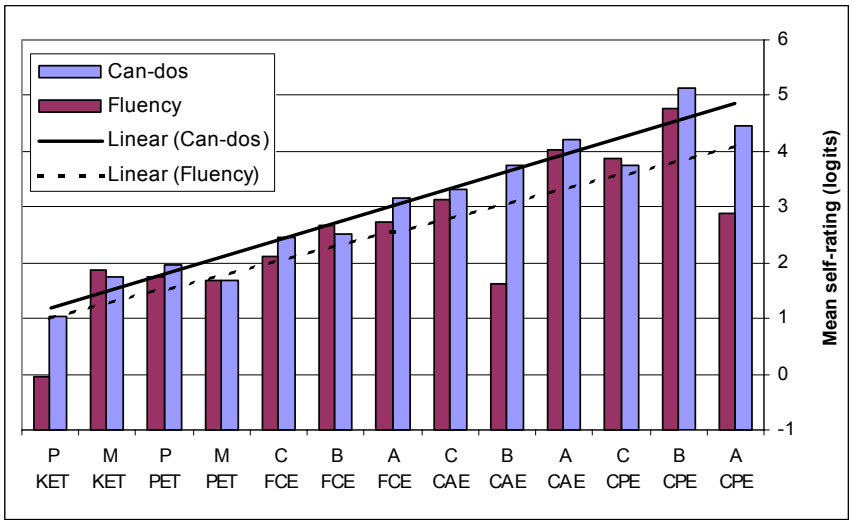


Figure 5: Mean self-ratings ("Can do" statements, Fluency) by exam grade (December 1999 data)

Figure 5 shows self-ratings on the "Can do" statements and on the "Fluency" statements separately estimated. A clear relationship is evident between self-rating and exam grade achieved (the odder values are due to very small numbers of candidates in particular groups).

Grouping on exam grade, a high correlation was found between mean self-ratings and exam grade achieved. Table 2 shows that the "Can do" ratings bear a slightly closer relation to exam grade achieved than do the "Fluency" statements.

	Cando	*Fluency*
Fluency	0.86	
exam level	0.91	0.79

Table 2: Correlations between exam level, "Can do" and Fluency self-ratings, grouping by exam level achieved

Summarised by exam group, the strength of the relationship between exam grade and self-rating of ability is clear. None the less, there is considerable variability in self-rating at the level of the individual respondent, and this weakens the individual correlation with exam grade. A simple prediction of one from the other is not supported by the present data.

3.3.2 Using a mastery criterion to define the meaning of "Can do"

The variability between self-rating and exam grade suggests that individuals understand "Can do" in different ways. To relate "Can do" statements to levels of proficiency it is necessary to define mastery of a level in terms of a specific probability of being able to perform particular tasks which describe that level. If self-ratings accurately reflect the true difficulty of tasks, then the probability of respondents at a given level endorsing statements which describe that level should be constant across all levels.

This idea was tested against the present data, grouping respondents by exam level (Pass grade candidates only), and grouping the "Can do" statements by ALTE level, as found in the equating described above. Figure 6 shows that rather than remaining constant across levels, the probability of endorsing statements at the level of the exam drops steadily from 90% at Level 1 to less than 60% at Level 5. In this data, there is a mismatch between the notion of a criterion mastery level and the bandings of the "Can do" statements.

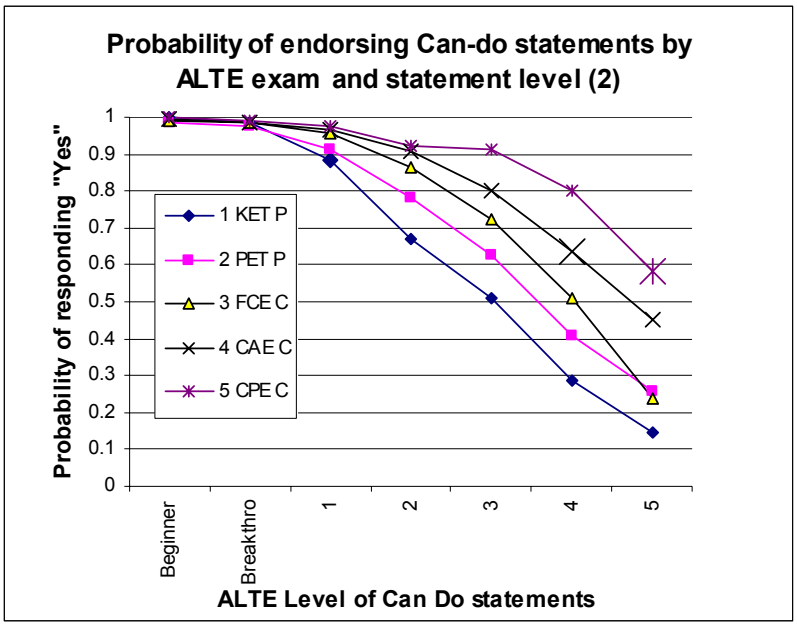


Figure 6: Probability of candidates endorsing "Can do" statements as assigned to ALTE levels by the equating process

4. Discussion

4.1 Effects related to respondent groups

The analysis reported here found that responses to both the ALTE and CEF statements can vary across groups of respondents.

As noted above, it was found that age was a factor: the responses of young people were more likely to misfit. The "Can do" statements chiefly concern ability to operate in an adult world, and refer to tasks which children of school age would probably have had no experience of. Thus we can expect that their responses might be inconsistent, or differ from the pattern observed in the adult majority. We had not actually intended to elicit responses from this age group; these respondents were mostly candidates for Level 1 (KET) exams. This effect is therefore explicable, but it does indicate that a set of language proficiency level descriptors designed for adults may not be applicable without modification for use in schools. Group effects were also found when respondents were grouped by area: Social and Tourist, Work, and Study. Respondents to the Social and Tourist questionnaires tended to discriminate levels more finely than respondents to the Study questionnaires. This was true of responses to CEF statements as well as ALTE "Can do" statements. No demographic features of these groups, or features related to their level or spread of language proficiency, as measured by exam grades, could be found to explain this.

As noted above, considerable variability was observed in self-ratings when compared to exam grade achieved. Some, though not a large part, of this variability appeared to be related to first language background, in that particular L1 groups tended to rate themselves relatively higher or lower.

Evidence was found that respondents with a lower language proficiency level, as measured by exam grades, tend to rate themselves too generously. Candidates at higher levels, perhaps from the vantage point of a better understanding of the range of the foreign language and the difficulty of achieving perfection, seem prone to rate themselves more modestly. This would account for the pattern of responses found in 0 above and illustrated in Figure 6, where exam candidates achieving a pass at Level 1 have a 90% probability of endorsing Level 1 statements, while Level 5 candidates have a less than 60% probability of endorsing Level 5 statements.

4.2 Comparing the "Can do" and CE statements

In 0 above it was noted that the "Can do" statements in their present form operate on a shorter scale, that is, are less discriminating, than the CEF statements. This is not surprising, because the CEF statements tend to be longer, and are composed of pre-calibrated selected and grouped elements. This produces a very coherent pattern of responses, with a consequently longer scale. The "Can do" statements in the form used in the questionnaires are by contrast relatively short, atomic, statements, deconstructed from the original paragraph-length level descriptions. They do not "epitomize" levels in the same way. A subsequent stage of development will group them again into more rounded, holistic descriptions of levels.

This difference means that to equate the "Can do" and CEF scales a linear transformation is needed, to compensate for this difference in discrimination. The scaling adopted distributes the "Can do" statements more as expected, so that they cover every level. They range from the very easy, pre-Breakthrough level (A1), e.g.:

> CAN ask very simple questions for information, such as "What is this?". CAN understand 1 or 2 word answers.

to Level 5 (CEF C2), e.g.:

CAN scan texts for relevant information, and grasp main topic of text, reading almost as quickly as a native speaker.

It is still the case that there are more “Can do” statements at the middle levels. This generally reflects the range of tasks and situations included in the “Can do” statements. However, the relatively low number of tasks at ALTE levels 4 and 5 is not wholly consistent with the intention of the writers of the “Can do” statements.

As noted in 0 above, the highest-level “Fluency” statements were also not well distinguished in the analysis reported here. The problem of identifying “Can do” statements to describe high levels of proficiency may be specific to these response data, which are based on self-ratings.

4.3 Comparing self-ratings and exam grades

There is a close relationship between exam grade achieved and self-rating, both on the “Can do” and the CEF “Fluency” statements. Despite the fact that the “Fluency” statements are more highly discriminating than the “Can do” statements, the latter actually correlated more highly with exam grade. This effect is possibly due to the fact that the “Can do” statements, like the exams, embrace a broad range of language skills. “Fluency” has a narrower focus.

At the level of the individual candidate the relationship between exam grade and self-rating is weaker, due to the considerable variability in respondents' overall perception of their own abilities. That is, different people tend to understand “Can do” somewhat differently.

It is worth stressing that this is not a problem with the coherence of the “Can do” scales themselves (or the CEF “Fluency” scales, which behave in a generally similar way): people agree well about the relative difficulty of statements, and in consequence the measurement scales are clearly defined. The problem is probably a particular feature of the present data, based on self-report. Where ratings are supplied by experienced raters, using standardised procedures, then we can expect much higher correlations with exam grades. Further research using experienced raters will probably be necessary to fully characterise the relationship between exam grades and typical “Can do” profiles of ability.

5. Conclusions

The Council of Europe and ALTE have invested much effort in the development of Framework systems for describing language proficiency because they see that such systems meet the needs of language learners, language teachers, and end users of language qualifications.

This chapter has discussed issues that arise in developing a framework, and more particularly, in relating different frameworks to each other. How concrete can a description of language levels be, while still being generalisable?

Implicit in this is the idea of context, of a target population of language users for whom the framework is intended to be relevant. The context of the work discussed in this chapter is clearly European. This does not necessarily make it irrelevant to other contexts.

The work described here has adopted a measurement viewpoint: we have used Latent Trait methods to attempt to construct measurement scales, on the understanding that features of

interest to us can be adequately characterised as relating to unidimensional scales. By attempting to fit into this useful simplifying framework data from different instruments (CEF statements, "Can do" statements, language exams) relating to different target areas of use, and collected from respondents from a variety of backgrounds, we are able to explore the limits of generalisability.

In this chapter we have focussed on a statistical approach to equating different assessment scales. This approach is useful. It enables us to verify and improve the internal consistency of each scale, and to see how the scales relate to each other, and whether they appear to measure the same general aspect of proficiency.

But the statistical approach is also useful precisely because it identifies problems with how people understand and use assessment scales. In this chapter we have found evidence that a range of effects relating to groups of respondents – their age, first language, proficiency level, area of language use – may affect their understanding of a scale and of the meaning of level descriptors expressed in "Can do" terms. This complicates the task of relating scales to each other, but also has implications for the use of such descriptive frameworks.

The statistical approach must be complemented by a qualitative, analytical one. In the present case, an equating of the ALTE "Can do" scale to the ALTE Framework as defined by language exams, or to the CEF, will be successful to the extent that the scaled statements accord with the judgement of qualified assessors, and the picture of language proficiency which well-constructed, communicatively-oriented language exams provide.

The work described in this chapter thus suggests that in order to construct and work with frameworks of language proficiency the descriptive approach – even where based on empirical evidence – needs to be complemented by an approach based on assessment through formal systems of language examinations.

References

Alderson, J. C. 1991: Bands and scores. In: Alderson, J.C and North, B. (eds.): Language testing in the 1990s London: British Council / Macmillan, Developments in ELT, 71–86.

Council of Europe 1996: *Modern Languages: Learning, teaching, Assessment. A Common European Framework of Reference*. CC-LANG (95) 5 rev IV, Strasbourg, Council of Europe.

Council of Europe 2001: Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Hambleton, R.K., Swaminathan, H. and Jane Rogers H. 1991. Fundamentals of Item Response Theory - Newbury Park: Sage Publications

Henning, G. 1987. A Guide to Language Testing Development Evaluation Research. Cambridge, Mass.: Newbury House

Jones, N. 2000: The Association of Language Testers in Europe (ALTE) Framework is a standard for European language exams. Paper present at TESOL conference

Jones, N.2001: The ALTE “Can do” Statements. In Council of Europe.

Milanovic, M. 1993: The development of European language certificates. Paper delivered at conference: New needs in international communication and foreign language teaching, Tokyo 1993

Milanovic, M. 1995: Comparing language qualifications in different languages: A framework and code of practice. System 23: 467-479

Milanovic, M. and Saville, N. 1995: Anything you “Can do”, I “Can do” better! The ALTE “Can do” Project. Paper presented at the TESOL conference

Milanovic, M., Saville, N. and Taylor, L. B. 1997: The ALTE User's Guide for Examiners. In Huhta, A. et al (eds.)1997: Current developments and alternatives in language assessment. Proceedings of Language Testing Research Colloquium 96. University of Jyväskylä: 465-481
North, B. 1996/2000: The development of a common framework scale of language proficiency. PhD thesis, Thames Valley University. Reprinted 2000, New York, Peter Lang.

North, B. and Schneider, G. 1998: Scaling descriptors for language proficiency scales. Language Testing 15/2: 217-262.

Wright, B.D. and Stone M.H. 1979. Best Test Design. Chicago: MESA Press

CHAPTER 12: THE COMMON EUROPEAN FRAMEWORK AND THE EUROPEAN LEVEL DESCRIPTIONS FOR GERMAN AS A FOREIGN LANGUAGE

Lukas Wertenschlag, Martin Müller, Helen Schmitz

Introduction

ENDaF is an adaptation and expansion of the CEF for German as a Foreign Language. The European Level Descriptions for German as a Foreign Language (i.e. ENDaF - . Europäische Niveaubeschreibungen Deutsch als Fremdsprache) are based on the Council of Europe Framework of Reference and have largely adopted its models and categories in order to describe the attainment levels A1 - B2 more precisely. This chapter focuses on the descriptions of competence (in the form of "Can do" statements) for the four types of language activity: reception, production, interaction and mediation. It describes for each level what a speaker can do linguistically and what linguistic means (i.e. vocabulary and structures) he or she can employ for that purpose. However, language competences alone are not sufficient to enable adequate action in interpersonal encounters. In addition, language users require sociocultural competence to enable them to choose appropriate language activities. The ENDaF project thus also describes sociocultural competence for each level with "Can do" statements. In addition, ENDaF offers open-ended lists of strategies and techniques which are not tied to any level. These new level descriptions are published on a CD-ROM and allow the user ease of access and use by offering multiple search and combination facilities. The main purpose of the ENDaF level descriptions (together with the linguistic tools for each level) is to facilitate the planning, implementing and assessing of activity- and task-oriented foreign language teaching and learning.

1. Background

The Council of Europe's language policy in the 1970s and 1980s concentrated on the democratisation of language learning, on needs-oriented learning, learner autonomy, and the protection of minority languages. One of the major results of this emphasis was the Threshold levels, which were created for various languages following the example of English. In the 1990s, with the slogan "Transparency and Coherence", the Council of Europe opened a new chapter for European language policy; its effects are now seen with the publication of the CEF and the development of various language portfolios. The ENDaF project also uses the CEF as a framework and aims to add to, to extend and to adapt to current needs the existing level descriptions, curriculum, syllabus or test specifications in the area of German as a foreign language.

2. The ENDaF project

2.1 Project Aims

The project aims to test whether the CEF can be used in creating descriptions of levels of competence, for German as a Foreign Language. The main objective is to develop consistent level descriptions based upon the Framework's ideas for the levels Breakthrough (A1), Waystage (A2), Threshold (B1) and Vantage (B2). Thus far these have either not been

described - for instance “Breakthrough for German” - or they are outdated, as, for example, the “Kontaktschwelle” published in 1980 (Baldegger et al). The main focus of ENDaF is to describe in detail the levels suggested by the CEF and the Portfolio (Swiss version), together with related “Can do” statements. A second aim is to exemplify concepts such as “task”, “text” or “strategies” and thereby make them usable for practical applications. A third aim of the project is the reformulation or updating of existing lists for levels A1 to B2, i.e. an adaptation of the lexicon and structures to current and future standards. One of the main concerns of the Framework is the issue of learning to learn. The ENDaF project aims to define this concept more precisely and presents concrete examples of some of the strategies and techniques that “learning to learn” might involve.

This will ensure that the various stakeholders (curriculum planners, materials writers, examiners and teachers) have a reference point in order to develop, for instance, new tests and diplomas which can be fitted into the Council of Europe scales. The ENDaF level descriptions will be published in 2002.

2.2 Project Brief

In 1998 a workshop was organised by the Goethe-Institut to reach agreement on the major aims of the project: to apply and, if necessary, adapt the “Can do”-statements from the CEF and to develop specific lists of vocabulary and speech acts for each level. The project would be led by a team with representatives from Germany, Austria and Switzerland, the main German-speaking countries. The results should be user-friendly and, consequently, it was planned from the outset to make the project results available as a database on CD-ROM. It was decided that ENDaF should be oriented as far as possible along the lines of the CEF and that it should contribute to the continuing development of the Council of Europe's learning objectives by elaborating on and differentiating between the categories of the Framework. The challenge consisted in illustrating the categories from the CEF with concrete language material and in defining the quality and quantity of linguistic means for four out of the six levels of competences . Up to now this kind of description – based on the CEF - does not exist for German as a foreign language.

The project is funded mainly by the Goethe-Institut, the Austrian Language Diploma (ÖSD) and the Austrian Ministry of Education and Science in co-operation with the Council of Europe.

3. CEF and ENDaF

In our work, we have concentrated on the following aspects of the CEF: the “action-oriented” approach (Chapter 2 in CEF), the competences of the user/learner (Chapter 5), language use and the language user/learner (Chapter 4), the processes of language learning and teaching (Chapter 6), the role of tasks in language learning and teaching (Chapter 7) as well as scaling and levels (Chapter 3). The sets of examples and the concrete level descriptions in Chapter 4 and the appendices A – D also proved useful . Where descriptions were missing, for instance for mediating activities and strategies (CEF 4.4.4), we endeavoured to complete them.

The CEF has encouraged us to concentrate on those aspects where research is still needed, but which are crucial in language use. Thus, we have taken the CEF concepts of learning strategies, communication strategies and sociocultural competence and have sought concrete examples of these.

Other aspects of the CEF such as assessment (Chapter 9) are less important for the ENDaF project and were therefore not considered..

4. Main focus of ENDaF

4.1 Introduction

The "Kontaktschwelle", (the German version of Threshold) with its notional-functional approach, is very familiar to foreign language teachers, material writers and curriculum developers. So we have in large part adopted this approach for ENDaF. Thus, the project presents lists of specific language exponents and a system of communicative activities, general and specific notions as well as a functional description of grammar (Figure 1).

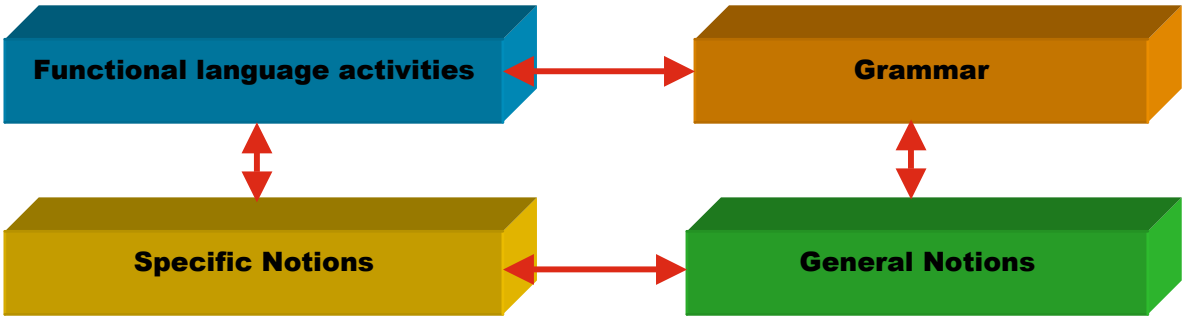


Figure 1 Linguistic Components

Where it is appropriate, the different components are linked; for example to express a possibility in German you can use the form *wenn*. When you look up *wenn* on the CD, you find something like Figure 2:

List: Functional language activities

	communicative activity	Example
wenn ... (dann)	speak about eventuality	Wenn Sie das bis morgen machen könnten, wäre ich sehr froh.

↓

Link to General notion	Link to communicative activity	Link to Grammar
Condition (with other linguistic means)	to describe something as possible (with other linguistic means)	Konjunktiv II (with explanations about form and function)

Figure2: The relationship between the linguistic components

By means of such links the user of the planned CD-ROM can "surf" until she has found suitable material .

In the ENDaf project some other components, inspired by the CEF have been added to these four familiar parts of the Threshold levels (Figure 3).

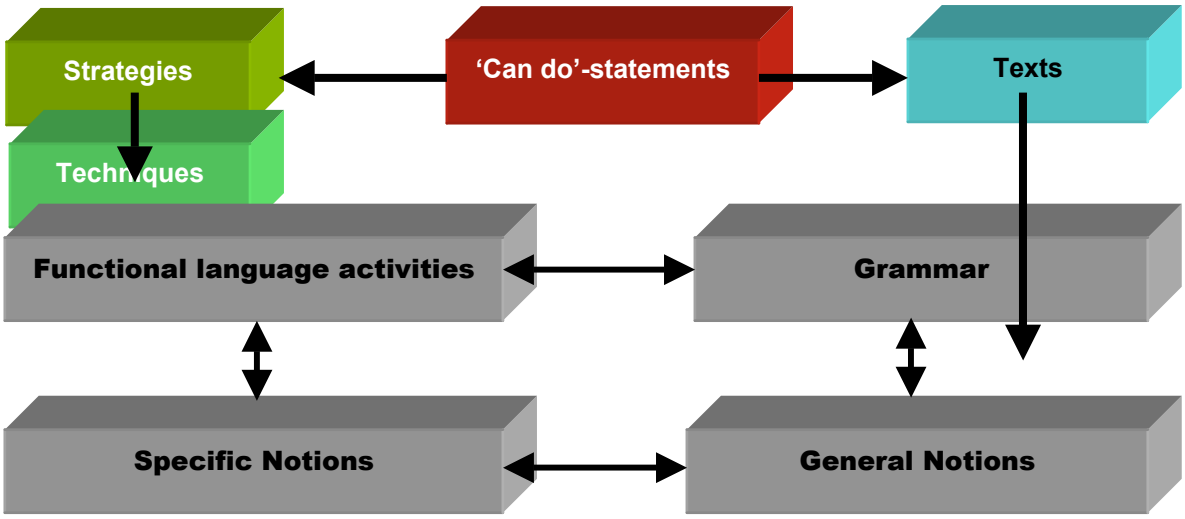


Figure 3: The main components of ENDaF

The "Can do" statements are crucial for the descriptions of the different levels. We link a specific "Can do" statement with a number of texts which are the input or output of communicative activities.

We build on the following aspects of the CEF:

- The action-oriented or task-based approach: When formulating "Can do" statements, we begin with language activities (e.g. mediation), which were not as explicitly defined in the CEF level descriptions.
- The domains of language activities: If the first generation Threshold level is compared with more recent products, it is striking that the earlier descriptions were often based on the assumption that language contact happened predominantly in the domain of leisure activities. With the exception of the French "Niveau Seuil", language learning was seen through the eyes of the tourist. In recent years, the emphasis has shifted under the influence of occupational mobility and educational exchange programmes. The four domains of the CEF - the personal, the public, the occupational and the educational domains - therefore assume importance as follows:
- In the description of the "Can do"s' .
- In vocabulary and structures.
- To differentiate speaker profiles depending on the user's needs. Thus, a learner could aim for level A2 for personal purposes, and level B1 in the occupational domain, and so on.
- In the description of strategies and techniques in the context of tasks and texts.
- Finally, we included socio- and intercultural competence, taken from the CEF, in our descriptions of attainment levels. For each level we describe the sociolinguistic elements for different situations.

In summary, ENDaF has adopted and adapted the CEF in the following way:

There is a certain task to be solved→
in one of four domains →
in a concrete situation and a concrete operational context
→ In order to solve the task successfully, the person must have available certain linguistic (and sociocultural) competences,
→ allowing him or her
if necessary with the help of strategic decisions and using specific techniques to receive a previously produced text or to (re-)produce a new text (possibly together with a partner).

Below we explain how these ideas and models are being implemented.

4.2 Task

The language user carries out tasks using strategies and linguistic, sociocultural and communicative competences. In other words, the task is a meeting point for several strands of an attainment level description (general competence, sociolinguistic competence, linguistic competence, strategic competence and so on) The task also links language learning and language use, and it represents the interface between pedagogic and real-world fields of operation. (Tasks need not be exclusively linguistic).

The action a person takes depends on various factors such as personality and an assessment of his or her own competences. In addition, the investment of time may also determine the choice of action.

The descriptive system of the ENDaF with its different components should enable the user to formulate tasks at different levels and for diverse speaker profiles in such a way as to combine the competences – in the form of "Can do"s' - necessary for carrying out the task with the requisite linguistic means – structures and vocabulary.

4.3 "Can do" Statements

The "Can do" statements are an essential addition to the "Kontaktschwelle" and a key component of the ENDaF Project. They describe what a language user can achieve at a specific level of attainment by a given (language) activity. We distinguish between two types of "Can do" statements:

- **"Can do"s' relating to language activities**

Here we have adopted the CEF division of language activities into four sets of "Can do" statements: reception, production, interaction and mediation.

- **sociocultural "Can do" statements**

They describe socioculturally determined situations in which the language user can act in socially appropriate ways. Such situations are often mastered with ritualised patterns which differ from one culture to another. They include activities such as greeting, leave-taking, expressing thanks, inviting someone, and agreeing or disagreeing with someone. The major

basis for these “Can do” statements is the section on Sociolinguistic Competence in the CEF (5.2.2).

The “Can do” statements relating to language activities

The basis for the “Can do” statements in the ENDaF project are the scaled “Can do” descriptors in the CEF and the “Can do”s' in the Language Portfolio (Swiss version). We have adopted practically all the “Can do”s' from the Language Portfolio since they have been field-tested, and we have compared them with those of the CEF and additional “Can do”s’ stemming from various curricula and catalogues of learning objectives (Goethe-Institut curricula; Certificate of German; curriculum of the Austria Institute; ALTE descriptions).

Following the CEF, we subdivided the language activities into the following four groups:

RECEPTION	spoken written	INTERACTION	spoken written
PRODUCTION	spoken written	MEDIATION	spoken written

By adding three examples from the four domains for each “Can do” statement, we hope to concretise what is meant by the rather general description of competence in the CEF.

Figure 4 presents a short extract from the “Can do” statements at level B1 adapted from the CEF as an example:

‘Can do’ (Interaction - spoken):

CAN EXPRESS HIS/HER OPINION WHEN A PROBLEM NEEDS TO BE SOLVED OR A PRACTICAL DECISION HAS TO BE TAKEN.

Example 1:

Can negotiate with a friend on the telephone where to meet in the evening.

Example 2:

Can express his/her opinion about whether a new piece of *equipment (e.g. a workplace computer) should be purchased*

Example 3:

Can make suggestions in a group of learners how a given task can be solved.

Figure 4 “Can do” statements at level B1

These examples are intended as stimuli for materials writers, curriculum and test developers for further specifications of the “Can do” statements and to illustrate how they can be adapted for a particular target group. The examples can therefore also be seen as an invitation to users of the level descriptions to write their own concrete exemplars for their own groups of learners. We have taken care to select examples from three different domains for each “Can do”, the first example in Figure 4 concerns the personal domain, the second example relates to the occupational domain and Example 3 to the educational domain.

The “Can do” statements are linked to a list of text types, which enables, for example, a search for “Can do”s' which relate to newspaper texts (Figure 5).

“Can do” (Reception written and spoken):		*Texts:*
CAN UNDERSTAND IMPORTANT INDIVIDUAL STATEMENTS AND CONCLUSIONS IN A CLEARLY STRUCTURED TEXT	⇨	Presentation (of broadcast) Terms of business Announcements Inquiry Instructions given by teacher Instructions given in course book Operating instructions Instruction leaflet (medical) ... Summary
Example 1: Can understand the results and information given about a game in a sports report.		
Example 2: Can understand viewpoints and decisions laid down in the minutes of a meeting.		
Example 3: Can understand the main goals and the most important conclusions in a project presentation.		

Figure 5 “Can do” statements and texts

The list of texts includes more than 170 types of text, each of which is assigned to various categories. One of the essential categories is the medium through which a text is transmitted (see section 4.4 below).

For instance, an *inquiry* can be assigned to one of four categories: direct transmission (as in a face-to-face encounter), loose leaf form (a letter or facsimile), new media (e-mail) or telephone and answering machine (as an inquiry by phone). This cross-assignment allows a targeted search e.g. for newspaper texts or new-media texts.

Linking the text list with the “Can do” statements enables the user to search for “Can do”s' which may relate to a certain text, e.g. a presentation, as shown in Figure 5..

The sociocultural “Can do” statements

Based upon sociocultural competence as described in the CEF we have created “Can do”'s for this area and have displayed them in a similar way to the CEF.

Domains of sociocultural aspects of communication		**"Can do"-statement**
ESTABLISHING AND TERMINATING CONTACT	Establishing contact	Can address a person appropriately and can react when he is addressed
		(...)
	Terminating contact	Can use leave-taking routines appropriately.
SOCIAL CO-OPERATION	express one's thanks	(...)

Figure 6: Extract from sociocultural "Can do"'s at level B1

In Figure 6, the left column displays the aspects according to which we have grouped the "Can do"'s in the right column. The right column is linked with appropriate expressions and structures in the list of functional language activities. For the first "Can do"-statement in Figure 6 one can find expressions like "*Entschuldigung,… Bitte, … Hallo!*" and so on. Obviously, the number of "Can do"'s differs from one level to another, increasing with rising level. For instance, the "Can do" statement "Can use exclamations appropriately ...", for instance, appears for the first time at Level B1. We have grouped the "Can do" statements according to the following aspects: *Establishing and terminating contact, social co-operation, emotions* and *body language*.

We have presented the "Can do" statements across the various levels. The purpose of such a presentation is to enable the user of ENDaF, as in the CEF, to compare the "Can do"s' directly with their corresponding entries on different levels. If, for instance, there is a "Can do" for Level A1 which is continued through Level A2 as far as Level B1, the user of the CD-ROM can move from one level to another and compare the respective "Can do"s'. In the same way, he or she can find out if, for a given "Can do" at Level B1, there is a corresponding statement at Level A2, . To take an example from the area of Interaction, the comparison would look like Figure 7.

A1	CAN WRITE A SIMPLE POSTCARD.
A2	CAN WRITE A VERY SIMPLE PERSONAL LETTER IN ORDER TO INVITE SOMEONE OR TO THANK SOMEONE FOR SOMETHING.
B1	CAN WRITE PERSONAL LETTERS TO FRIENDS OR ACQUAINTANCES ASKING FOR OR GIVING NEWS AND TELLING ABOUT EVENTS.
B2	CAN EXPRESS DIFFERENT MOODS AND FEELINGS IN PERSONAL LETTERS AND TELL ABOUT THE EVENTS OF THE DAY, POINTING OUT WHAT IS IMPORTANT IN A PARTICULAR EVENT.

Figure 7: Corresponding "Can do" statements across four levels

If there are corresponding descriptors on each level, the user can follow the development, as it were, of a "Can do" statement from one level to the next. However, there may not be descriptors for all subcategories for each level as some activities are not thought to be feasible

until a certain level of competence has been reached, and others may cease to be an objective at higher levels.

4.4 Texts

In line with the CEF, we have added a special chapter on the topic of texts to the new descriptions of attainment levels for German as a foreign language .We have also put together a comprehensive list of text types, which is useful because they are linked to the "Can do" statements, as described above.

Section 4.6. (Texts) of the CEF offers the following definition: "... "text" is used to cover any piece of language, whether spoken utterance or a piece of writing, which users/learners receive, produce or exchange. There can thus be no act of communication through language without a text." We have based our approach on this definition and describe texts as results of linguistic action which may be spoken or written Text characteristics include not only include formal design aspects (e.g. the subject line in a business letter) but also formal linguistic aspects (certain turns of phrases as, for instance, salutations in a letter) and structural and content aspects (e.g. minutes do not contain any personal opinions of the minute-taker).

The classification of text types

Figure 8 illustrates the classification of texts in the ENDaF project .

	TEXT	Broadcast announcement	General terms of business	Inquiry	Animated film	Announcement	Giving credits after a broadcast	Application	Instruction by teacher	Instructions in course materials	Advertisement	Article	Essay	Scientific paper
Channel	Spoken	x		x	x	x	x	x	x					
	Written	x	x	x	x	x		x	x	x	x	x	x	x
1 Interaction	1.1 simultaneous	x		x		x		x	x					
	1.2 delayed	x		x		x		x						
2 Medium	2.1 direct (no additional transmission medium)	x		x		x	x	x	x					
	2.2 Newspaper and magazine					x					x	x		x
	2.3 Book (and reference work)									x		x		
	2.4 Loose leaf	x	x	x		x		x					x	x
	2.5 Other carriers of writing (not paper-based)								x					
	2.6 Television, cinema or theatre	x			x	x	x		x					
	2.7 Radio and audio recordings					x	x		x	x				
	2.8 New media	x		x	x	x			x	x	x	x		x
	2.9 Telephone/Answering	x		x		x	x							

	machine													
	2.10 Public amplifier system					x	x							
3 Purpose	3.1 Official/legal obligation		x		x			x			x			
	3.2 General information	x	x			x	x	x			x	x	x	x
	3.3 Specialist information		x					x				x	x	x
	3.4 Advertising					x					x			
	3.5 Instruction		x				x		x	x	x			
	3.6 Teaching								x	x			x	
	3.7 Entertainment				x							x	x	
	3.8 Arts				x								x	
	3.9 Religion													
4 Form of - representation	4.1 Graphics or table										x			x
	4.2 with pictures				x									
	4.3 with music				x									
	4.4 in verse													
5 Domain	5.1 Public	x	x	x	x	x	x	x			x	x	x	
	5.2 Private	x		x	x			x						
	5.3 Occupational		x	x		x								x
	5.4 Educational					x			x	x			x	x

Figure 8: Classification of texts

Following the CEF, we first distinguished between spoken and written texts. Then we subdivided the texts into five separate groups.

The first group concerns exclusively texts emerging in interaction and characterises them according to whether they are the result of interaction in which all the partners communicate with one another simultaneously (e.g. in a face-to-face situation or an on-line chat) or whether communication takes place with a degree of delay (e.g. in letter correspondence). The second group refers to the medium used to transmit the text. In contrast to the CEF, where some media are given as examples only, we have opted for a comprehensive representation and assigned each text to at least one of ten media. The third set describes various purposes which texts are intended to accomplish. The fourth group offers the option of assigning special forms of representation to a text, for example graphics and tables; pictures; music; verse. The final set concerns the domain in which the text is produced (public, private, occupation, education). All texts can appear in virtually all of the four domains. However, we have restricted ourselves to assigning the texts to domains in which they occur most commonly.

Putting the text classification to use

Apart from using the list of texts in conjunction with the "Can do" statements, the classification offers an added value: the user can pick particular text types from the CD-ROM which fulfil certain criteria. For instance, a materials writer can search for texts which match a given situation such as *Telephoning in the workplace*. He or she then looks for texts which can be transmitted over the telephone, which provide general information and which are of relevance for the occupational domain. The result of this search would be the text types *Announcement, Application, Informational exchange, Advisory conversation, Report, Complaint, Invitation, Conversation (in the workplace), Comment, Notification, Offer* and

Product information. The result of this search can then be used to select specific text types and convert concrete examples into learning materials.

4.5 Strategies

As we have seen, communication and language-learning processes frequently involve solving tasks. In order to complete tasks as comprehensively, economically and successfully as possible, the individual deploys specific behaviours which can be called *strategies* in the sense of the CEF: "the adoption of a particular line of action in order to maximise effectiveness." Strategies can, of course, be employed for various purposes. We distinguish *strategies of language use,* which are intended to achieve successful communication) from *language learning strategies*.

Strategies of language use

Figure 9 illustrates how an individual decides on a strategy depending on the task he or she is confronted with. If the task is of a linguistic nature, language activity will ensue which results in texts. Figure 9 also demonstrates that the text which is produced, received or jointly created in interaction or mediation in a linguistic activity is dependent on the task. The strategy used to complete a task in turn depends on the text to which the task is related.

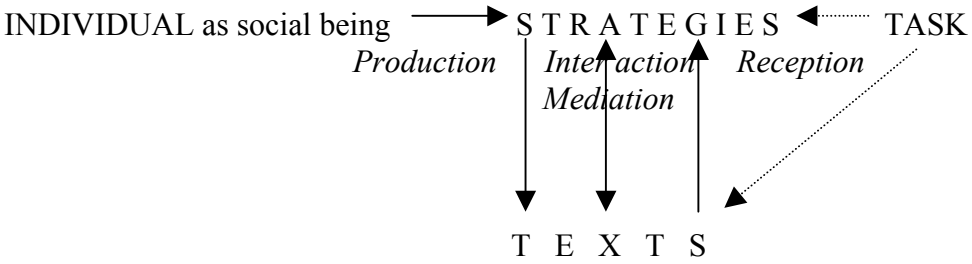


Figure 9: Interrelatedness of task and strategy

On the other hand, the choice of strategy is also determined by individual preferences and the repertoire of strategies and techniques a person may have. In the ENDaF Project we have described, for the four main language activities, various strategies with examples of how they may be operationalised.

Learning strategies

The CEF distinguishes several important factors which can affect learning: language and communication awareness, general phonetic skills, study skills and heuristic skills.

It is important to note that the ability to learn how to learn has no direct relationship with language competence and skills. This means that competence descriptions in the learning strategy domain have to be separated from the language competence descriptions.

The ENDaF Project aims to describe profiles of language competence at certain levels of attainment, rather than to describe learning skills at distinct levels. Therefore, learning strategies are not assigned to any level and are not expressed as "Can do" statements. Instead, we describe the strategies and concomitant techniques with examples based on the work of Bimmel and Rampillon, using a different categorisation, however, which employs the notion

of learning process and strategy use as introduced by the CEF, that is Planning - Execution - Evaluation - Correction. By taking account of the ten study skills mentioned in the CEF we developed a practical system for classifying strategies and techniques. This approach yields the following subdivision: Before learning -- while learning -- after learning.

Different strategies may be subsumed under a common goal. Thus, the learner can achieve the aim of “planning and preparing learning” by means of the strategy “determining forms and times of learning” (Figure 10).

Strategies for self-regulation before the learning process

AIM	STRATEGIES
planning and preparing learning	*establishing personal needs and objectives* *determining appropriate forms of learning* *organising learning times* *equipping the place of study*
organising and adapting the subject matter to be learnt	*activating background knowledge about the subject*
(...)	

Figure 10: Extract from the structuring of learning strategies in the ENDaF Project

A strategy can be deployed by using one technique from a set of different techniques at the disposal of the learner. It is impossible to determine all conceivable techniques. Equally, the choice of technique depends on the individual learner's experiences and the actual learning situation. Therefore, we have decided to present possible techniques as concrete examples hoping that users will be encouraged to add techniques. The techniques usually comprise a very concrete means used to reach an aim (Figure 11).

Strategies for self-regulation before the learning process

AIM	**STRATEGIES**
planning and preparing learning	*establishing personal needs and objectives*

Technique **1**
On the occasion of course enrolment to lay down in a learning agreement which diploma the learner wants to achieve

Technique **3**
To ask a colleague about the language skills needed in a new job

Technique **3**
To reflect on short and medium term objectives (what do I want to be able to do by the end of the semester or before the next test?) and to note them down in a learner journal

Technique **4**
To display on a piece of notepaper in the workplace what should be accomplished within the next week (e.g. to speak German for at least 3 minutes with a colleague in the coffee break)

Figure 11: Learning strategy and techniques

4.6 Pluricentrism

One feature of the ENDaF Project - and an addition to the CEF - is its pluricentrism, i.e. the inclusion of the standard varieties of the German language as spoken in Germany, Austria and Switzerland. For instance, for *Saturday*, in Germany the words *Sonnabend* and *Samstag* are used, whereas in Switzerland and Austria it is *Samstag* exclusively. Differences are not only found in the lexicon, but also in the grammar, e.g. in the use of prepositions (Germany/Switzerland: *auf dem Land;* Austria: *am Land*) or in verb conjugation (southern German language area: *er/sie ist gestanden*; northern German language area: *er/sie hat gestanden*). Similar phenomena can be seen in the sociocultural sphere, e.g. in the use of greeting and address forms. Such differences between the various language regions are specially marked in our lists.

4.7 Linkage

Language activity is a complex process which can never be depicted comprehensively with all its conditions. Nevertheless, we have attempted to link elements from the *Kontaktschwelle* model with new components of the CEF in order to give the user quick and easy access to the various lists. These links should enable the user to develop his or her own tasks and to combine the necessary competences, texts, strategies, techniques and linguistic means.

5. Practical use

5.1 Possible applications

After completion of the ENDaF Project, four descriptions of levels of attainment will be available for German as a foreign language corresponding to the Council of Europe scales A1 - B2 whose "Can do"-statements have been empirically tested. As a result, national and international test developers will have a basis on which to develop examinations and diplomas which are in line with the Council of Europe levels. The materials put together in the ENDaF Project should be of practical value. They will therefore appeal to course providers, materials and curriculum developers, examiners, testing bodies and to teachers. This means, for instance, that

- the ENDaF lists and examples can easily be expanded by curriculum planners and teachers to meet the specific needs of particular learners or whole classes. Examples of "Can do" statements are a treasury for test developers who want to collate action-oriented materials for test and exams at a particular level
- tasks, lists and examples from ENDaF can be useful for teachers who would like to develop and create their own materials for their classes. The same materials can be handed out to the learners as checklists to assess their learning progress.
- learning objectives become transparent. Teachers can explain to the learners which subject matter is vital in reaching a certain level and what a particular examination demands
- profiles of tests and examinations are generally well-known. This can help curriculum developers and directors of studies to define levels for diplomas and tests. Employers and higher education institutions thus obtain clear information about the competence of graduates.

The various uses to which the materials may be put can be summarised as follows:

- *Use existing materials as an initial basis*
  The lists and examples worked out in ENDaF are a basis for the linguistic requirements at a certain level. It is possible to use the lists to compile essential vocabulary for a theme or important structures. The list of "Can do" statements can serve as a checklist for exam preparation while the same list can support an exam developer in writing test items.
- *Use existing materials as a stimulus and expand them*
  The materials may encourage users to extend them or to transfer examples from their own context. For instance, teachers can use the list of learning techniques to put together sensible techniques for a target group of learners; then they can add examples from their own practical experience or get the learners to do so.

  In the same way, language activities can be compiled which are relevant to a given target group. Examiners and teachers at specialised schools (e.g. tourism training institutions) can complement the lists with additional job-specific structures and examples. It should also be possible to develop special profiles for different levels within the relevant skills. Students, for instance, require receptive skills more urgently at least at the beginning of their studies than productive or interactive ones. With the help of existing or self-made "Can do"s a differentiated entry and exit profile can be created for a language course for academic purposes.

5.2 Publication

The lists and accompanying texts will be published on CD-ROM. Only the most important information such as a set of instructions for the use of the CD-ROM, background texts on the ENDaF model as well as key lists of "Can do" statements for the description of attainment levels will appear in print. In print, the user will find a brief outline and on the CD-ROM the actual language elements will be presented which he/she needs for the creation of a test or a curriculum. The CD-ROM will offer various search options. The user will be able to compile tailor-made lists and either print them out or export them into a word processing or spreadsheet program.

The CD-ROM will also offer the opportunity to retrieve sets of data via the various lists or through an index of terms and expressions.

The CEF has proved to be a source of inspiration to ENDaF for the identification and development of teaching and learning activities, and we look forward to seeing further applications and extensions of the Framework.

References

Baldegger, M., M. Müller, G. Schneider (1980) Kontaktschwelle Deutsch als Fremdsprache. München: Langenscheidt.

Bimmel, P. and U. Rampillon (2000). *Lernautonomie und Lernstrategien*. München: Langenscheidt.

Rampillon, U. (1995). *Lernen leichter machen. Deutsch als Fremdsprache.* Ismaning: Hueber