

Aligning Cambridge ESOL Examinations to the CEFR: Issues & Practice

Paper

34th Annual Conference
International Association for Educational Assessment
Conference Theme: Reinterpreting Assessment: Society, Measurement and Meaning

Hanan Khalifa
Principle Co-ordinator, Research & Validation Division
University of Cambridge ESOL Examinations
Khalifa.h@cambridgeesol.org

&

Angela ffrench
Examinations Manager, Assessment & Operations Division
University of Cambridge ESOL Examinations
ffrench.a@cambridgeesol.org

Abstract

Today, the Common European Framework of Reference (CEFR) plays a central role in language education and policy not only within Europe but worldwide. It has growing relevance for language testers and examination boards, helping to define language proficiency levels and interpret language qualifications. For many language testers, it has become imperative to make the case that their exams are aligned to the CEFR. The Council of Europe has attempted to facilitate this by providing a toolkit of resources, including a draft pilot *Manual* for relating language examinations to the CEFR and a technical reference supplement.

In this paper we discuss Cambridge ESOL's experience in piloting the *Manual* and in aligning its own examinations to the CEFR. We compare the *Manual's* approach to alignment, which involves training participants to carry out a set of one-off activities such as familiarisation, specification, standardisation and empirical validation, and Cambridge ESOL's iterative and cyclic approach to alignment. The comparison will demonstrate how we can build an alignment argument that is based on existing test development and validation systems while generating evidence in line with the aims of the *Manual*.

Introduction

The Common European Framework of Reference (CEFR) was published by the Council of Europe in 2001 as a common basis for the description and elaboration of learning, teaching and assessment. It has within a short period of time become highly influential in Europe and beyond.

If we look for examples within the educational field, we will find that in France, secondary school students are expected to achieve B1 in their first foreign language and A2 in their second (Martyniuk 2008) and in Norway, national reading tests for school children are based on and reported as CEFR levels (Moe 2008). The European Survey on Language Competence Project (SurveyLang) launched in 2008 relates to the CEFR levels and will report achievement levels of 15-16 year old school students across Europe for the two most widely-taught foreign languages in each European Union country. Beyond Europe, educational authorities are increasingly preoccupied with international benchmarking. In Taiwan, for example, the Ministry of Education set about establishing a common standard of English proficiency through the adoption of the CEFR (see Wu and Wu 2007).

If we look at domains other than education, we will also find a CEFR impact. For example, the language requirements in relation to immigration and citizenship in Europe are increasingly couched in terms of CEFR levels. In Austria and Finland an A2 level is currently required for citizenship and in Germany and Denmark a B1 level is preferred. The United Nations has adopted the CEFR levels for training purposes at UN institutions around the world.

As a result of its fast growing influence, many language testers now seek to align their examinations to the CEFR. Publishers are also concerned to lay claim to linkages for their ESOL textbooks at various CEFR levels.

The Council of Europe has attempted to facilitate this linking exercise by providing a toolkit of resources, including a draft pilot *Manual* for relating language examinations to the CEFR, a technical reference supplement to this (Council of Europe 2003, 2004), and exemplar materials illustrating the CEFR levels (see http://www.coe.int/t/dg4/linguistic/Illustrations_EN.asp). The Council of Europe has also provided forums where practitioners shared their reflections on the use of the *Manual* and their experience in using the different linking stages as suggested in the *Manual*. In these forums, bodies offering examinations across Europe (France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Finland, Japan, Slovenia, UK and Turkey) presented data concerning their progress to date in linking a whole range of language examinations to the CEFR.

Approaches to alignment

In this paper we share an examination board's experience, namely Cambridge ESOL's, with the alignment procedures suggested by the *Manual*. A discussion on alignment cannot really take place without reference to the unique relationship the CEFR has with Cambridge ESOL Main Suite Examinations in terms of their shared purposes, namely, provision of a learning ladder and proficiency framework and in terms of informing each other's evolution and development (see North 2008, Taylor & Jones 2006 for a full discussion of this relationship).

Manual's approach to alignment

The *Manual* outlines the alignment process as a set of stages: *Specification*, *Standardisation* and *Empirical Validation with Familiarisation* as an essential component of the first two procedures. Familiarisation aims at ensuring participants have detailed knowledge of the CEFR. Specification involves "mapping the coverage of the examination in relation to the categories of the CEFR" (Council of Europe 2003:6). Standardisation aims at "achieving and implementing a common understanding of the meaning of the CEFR levels" (Council of Europe 2003:7). Empirical validation of alignment claims involves the collection and analysis of data on (a) task performance and item characteristics, (b) test quality, (c) learner performance and test scores, (d) rater behaviour and (e) the adequacy of the standards set. To facilitate the implementation of the *Manual's* approach a set of activities, forms and illustrative examples are also provided. The table below summarizes the steps that are currently recommended in the *Manual* (see Council of Europe 2003: pp 6-10).

Table 1: Summary of linking steps

Familiarisation
Consideration of a selection of CEF question boxes
Discussion of the CEF levels as a whole
Self-assessment of own language level in a foreign language
Sorting individual CEF descriptors from a CEF scale into levels
The outcome of the familiarisation process is: <ul style="list-style-type: none"> • narrative description reporting on the activities used and the results obtained
Specification
Describing the examination in question: objectives, learner needs, exam papers/sections, weighting/rationale for weighting, text and task/item types, marking criteria, scoring scheme and reporting results
Content analysis of the examination profiled in relation to the categories and levels of the CEF
The outcomes of the specification process are: <ul style="list-style-type: none"> • a chart profiling coverage graphically in terms of levels and categories of the CEF • narrative description of the process

<ul style="list-style-type: none"> completed versions of relevant forms 	
Standardisation	
Training with standardised samples of performance for productive skills	
Training with standardised test tasks for receptive skills and linguistic competence	
Benchmarking local samples of performance	
Standard setting for receptive skills and linguistic competence	
<p>The outcomes of the standardisation process are:</p> <ul style="list-style-type: none"> examples of tasks and task-templates samples of spoken and written performance procedures for marking and rating narrative explaining how CEF standardised performance samples were exploited to benchmark local samples to the CEF and set up moderating systems. 	
Empirical Validation	
<p>Internal Validation</p> <ul style="list-style-type: none"> pretesting items & piloting tests confirmation of good item characteristics & appropriate difficulty values confirmation that the exam tests what it claims to test calibrating items/tasks to provide parallel test forms 	<p>External Validation</p> <ul style="list-style-type: none"> correlating test results to tests reliably and validly calibrated to the CEF correlating teacher ratings to Can-Do descriptors calibrated to CEF
<p>The outcomes of the empirical validation process are:</p> <ul style="list-style-type: none"> graphic representation of the nature and strength of the empirical relationship of the examination to the CEF narrative providing information on: form of instruments, data collection design, number of subjects, analysis methods, problems encountered & solutions adopted, results obtained, interpretation of results 	

When considering the recommended linking scheme in the *Manual* and its toolkit of resources, one cannot help but admire its rigour and thoroughness. However, it seems that the *Manual* envisages the alignment process as a specific project being organised, possibly on a one-off basis, where participants are trained to carry out a set of activities, and reports are generated which constitute the evidential outcomes. This may very well be the case if no change tends to occur in the context within which an examination has been developed and administered or if the exam does not undergo changes. The criticality of the alignment process lies in its sustainability. The section below provides an account of how an examination board interprets the *Manual's* approach to alignment while at the same time ensures the sustainability of the process.

Cambridge ESOL's interpretation of the approach

Cambridge ESOL views the alignment of its tests to the CEFR as a key aspect of their validity; it views alignment as an embedded and integrated feature of its test development and validation model. The model has an ongoing iterative cycle from perceived need, through test design, trialling and administration to post exam review (for full discussion of the model see Saville 2003). The cycle allows for changes in learning, pedagogy and assessment trends, as well in the targeted candidature, to be incorporated into an examination. Therefore, by default, the linking process is perceived as an on-going activity rather than as a single activity at a given time. In this section, we show how we can build an argument based on the high quality of existing processes, while generating evidence in line with the aims of the *Manual*.

Familiarisation

The *Manual* perceives the familiarization procedure as an "indispensable starting point" before a linking exercise to the CEFR can be carried out effectively. It goes on to say that "participants in

the linking process [should] have detailed knowledge of the CEF” (Council of Europe 2003:1). This advice leads to the emergence of several issues, e.g., who are these participants, how many, how familiar are they with the CEFR on a familiarity-unfamiliarity continuum, should we take their assessment experience into account or should we start with an assumed zero baseline of knowledge, what mode of delivery should the familiarisation activity take: face to face or distant?

Within Cambridge ESOL, the participants who ensure that an examination is related to the CEFR span across the organisation and beyond. There is currently a core of 52 assessment, operations and research staff who work with a network of external stakeholders on test specification, item writing, test construction, performance scale construction and on the application and use of performance testing scales. In December 2007, the network included:

- 145 personnel involved in writing materials
- almost 16,000 oral examiners (over 700 of whom are team leaders, and a large proportion of whom do not reside in the UK)
- 500 writing examiners

Roles and responsibilities are at times interchangeable, e.g., a chair of one exam could be an item writer for another.

Experienced members of this large community certainly have a close understanding of Cambridge ESOL levels. For example, item writers providing materials for a particular examination are familiar with how to interpret the level, e.g., in terms of text difficulty, linguistic features, genre choice, etc. Item writer guidelines and examination handbooks provide detailed information on text selection and item writing at a certain level. Through their work with a particular proficiency level, many will have come in contact with the CEFR or at least Waystage 1990, Threshold 1990 and Vantage specifications on which the CEFR is based. Many will also be familiar with the ALTE Can-Do descriptors which have been calibrated to the CEFR scale and which are reported in examination handbooks. For Cambridge ESOL, then, familiarization with the CEFR is seen as a part of consolidating and building on existing knowledge. At the same time the prominence of the CEFR raises the need for a more general awareness-raising, of particular importance for staff/network just entering the organisation.

One way the organisation has incorporated *Manual*-advocated familiarisation activities into its practice is through its ongoing induction and training programme; annual seminar programme, and annual team meetings. The familiarisation event may take the form of a face-to-face workshop or self-accessed materials using an electronic platform for queries and feedback. Through the familiarisation activities, Cambridge ESOL aims to foster a common understanding of the objectives and aspirations of the CEFR and its descriptive scheme and a broad awareness of the nature of the relationship between Cambridge ESOL examinations and the CEFR; to ensure a shared knowledge of differentiating features across certain level thresholds (e.g. B1/B2 and B2/C1) to enable the rating of tasks and performances across these levels.

Specification

The *Manual* has provided several forms for use in describing the examination and mapping its content onto the CEFR. These forms elicit general information on the examination, on test development, the construct being measured, marking, analysis, grading and results reporting. Cambridge ESOL documents such information in internal documents, e.g., item writer guidelines,

routine test production process, standard operating procedures for exam production, and grading manuals as well as external documents such as examination handbooks or annual reports on examination performance which are available on its website. These documents specify the exam constructs and levels and explain how they are implemented very effectively and already make use of many things now incorporated into the CEFR but familiar for many years – the Waystage and Threshold specifications, for example.

Mapping of the examination content onto the CEFR does not happen as a single activity nor is it carried out by a single member of staff; it occurs at different stages of Cambridge ESOL’s test development cycle and a number of internals and externals participate in the process bringing together a variety of expertise. It happens at the planning phase, the design phase and the development phase. At these stages specifications are produced linking candidate needs to requirements of test usefulness including frameworks of reference such as the CEFR. Decisions are made with regards to exam proficiency level, text and task parameters as well as features of performance scales which illustrate the proficiency level, marking methodology and procedures for setting grade boundaries. Pretesting and trialling take place to confirm decisions made and/or allow for modifications. The use of a socio-cognitive approach to test development has also helped Cambridge ESOL define and operationalize the measured construct in terms of the CEFR levels. Further explicit reference to the CEFR is being introduced into Cambridge ESOL’s processes over time where this serves to complement or clarify, for example when examinations are revised and updated. Task design and scale construction for performance tests is a case in point as we see in tables 2 and 3 below.

Exemplification from the CEFR scales and the FCE speaking scale

FCE is a Cambridge ESOL examination at B2 level measuring all four skills. Performance on the speaking paper is assessed on the following criteria: Grammar and Vocabulary (G&V), Pronunciation (P), Discourse Management (DM), Interactive Communication (IC). Table 2 shows how the analytical scales in the Cambridge ESOL assessment criteria for speaking satisfy the requirements of the CEFR. Elements of Range and Accuracy (CEFR) are evidenced in the Cambridge ESOL criteria of Grammar & Vocabulary and Pronunciation. Fluency and Coherence (CEFR) are captured under the Discourse Management criterion, while Interaction (CEFR) and Interactive Communication (Cambridge ESOL) focus on very similar aspects of the performance.

Table 2: Analytical scales

<i>Common reference levels: qualitative aspects of spoken language use (2001 Table 3: pp 28-29)</i>	<i>Cambridge ESOL Assessment Criteria for Speaking (2008: Main Suite and BEC)</i>
Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, without much conspicuous searching for words, using some complex sentence forms to do so. (R)	Shows a good degree of control of simple grammatical forms, and attempts some complex grammatical forms. (G&V) Uses a range of appropriate vocabulary to give and exchange views on a range of familiar topics. (G&V)
Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstanding, and can correct most of his/her mistakes. (A)	Shows a good degree of control of simple grammatical forms, ... (G&V) Intonation is generally appropriate. (Pron) Word stress is generally accurately placed. (Pron)

	Individual sounds are generally articulated clearly. (Pron)
Can produce stretches of language with a fairly even tempo: although he/she can be hesitant as he/she searches for patterns and expressions. There are a few noticeably long pauses. (F)	Produces extended stretches of language despite some hesitation. (DM)
Can initiate discussion, take his/her turn when appropriate and end conversations when he/she needs to, though he/she may not always do this elegantly. Can help the discussion along on familiar ground, confirming comprehension, inviting others in, etc. (I)	Initiates and responds appropriately. (IC) Maintains and develops the interaction and negotiates towards an outcome with very little support. (IC)
Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some 'jumpiness in a long contribution. (C)	Uses a range of cohesive devices. (DM) Produces extended stretches of language despite some hesitation. (DM)

Table 3 shows how the format, task types, test focus and assessment criteria of the FCE speaking test are used in combination to satisfy the recommendations of the CEFR. The FCE Handbook specifies the format, task types, test focus of each part of the test. For example, Part 2:

- Task type and format: An individual long turn with a brief response from the second candidate. In turn, the candidates are given a pair of photographs to talk about.
- Test focus: Organising a larger unit of discourse; comparing, describing, expressing opinions.

Table 3: Language use for Level B2

<i>CEFR Language use and the language user/learner (a selection taken from pp58-82)</i>	<i>FCE format, tasks, test focus and assessment criteria</i>
Can give clear, systematically developed descriptions and presentations, with appropriate highlighting of significant points, and relevant supporting detail.	In Part 2, each candidate is given the opportunity to engage in an individual one minute uninterrupted 'long turn'. Part of the task requires candidates to compare two photographs, which inevitably involves description, and the language produced is assessed through the assessment criteria of <i>grammar and vocabulary, discourse management, and pronunciation</i> .
Can develop an argument systematically with appropriate highlighting of significant points, and relevant supporting detail.	In Part 3, the candidates engage in a collaborative task where they work towards reaching a decision through negotiation. This requires them to make use of the stimulus provided to express and justify their opinions and through evaluation of the possibilities to come to a conclusion. The language produced is assessed through the assessment criteria of <i>grammar and vocabulary, discourse management, and interactive communication</i> .
Can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	In Parts 1 and 4, candidates are asked their opinions on a range of topics, e.g.: <ul style="list-style-type: none"> • 'Do you think computers will replace newspapers and TV in the future?' • 'What are the disadvantages of living in a big city?'
Can give a clear, prepared presentation, giving reasons in support of or against a particular point of view and giving the advantages and disadvantages of various options.	In Part 2, each candidate is given the opportunity to engage in an individual one minute uninterrupted 'long turn'. In addition to comparing two photographs, candidates are required to focus on an aspect of the topic of the photographs, e.g. '.... and say why you think the music is important to the different groups of people.'
Can take a series of follow up questions with a degree of	The FCE Speaking test assesses the candidate's ability to

fluency and spontaneity which poses no strain for either him/herself or the audience.

respond to questions from both the examiner and the paired candidate spontaneously, without preparation. Assessment of fluency and relative strain is addressed through the criteria of *discourse management* and *pronunciation*.

Standardisation

Standardisation involves “achieving and implementing a common understanding of the meaning of the CEF levels” (2003:7). The *Manual* states that this involves: (a) *Training* professionals in a common interpretation of the CEFR levels using productive skills samples and receptive skills calibrated items which are already standardized to the CEF; (b) *Benchmarking* where the agreement reached at *Training* is applied to the assessment of local performance samples; (c) *Standard-setting* where the cut-off scores for the test CEF level(s) are set.

Cambridge ESOL addresses the *Manual Standardisation* activities through its existing rigorous system of recruitment, induction, training, coordination/standardization), monitoring and evaluation (RITCME) for its item writers and examiners (see Shaw and Weir 2006 for a full account of the system and Ingham 2008 for a detailed discussion on how RITCME is implemented). RITCME is a continuous process and remedial actions are carried out effectively and efficiently. This system ensures that these participants have adequate professional background and receive appropriate training in the skills required. It also allows for ongoing professional development through standardization, monitoring and evaluation of the performance of item writers and examiners of productive skills. Obligatory standardisation of writing examiners takes place prior to every marking session, and the writing samples used are evaluated by the most senior examiners for the paper. Co-ordination of oral examiners takes place once a year prior to the main administration session and the video samples of performances which are used are rated by Cambridge ESOL’s most experienced Senior Team Leaders and Team Leaders, representing a wide range of countries and familiarity with level. The marks provided are then subject to quantitative and qualitative analysis before being approved for standardization purposes. Materials used in the training and coordination events have already been standardized to the CEFR and some of these materials have been provided to the Council of Europe as illustrative receptive and productive samples of the CEFR levels (see http://www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/illustrationse.html). Once again we see here how the *Manual* recommended *Standardisation* activities are embedded within Cambridge ESOL practice.

Standard setting as described in the *Manual* is perhaps most applicable when designing a new examination or in a context when item banking methodology is not in use. Cambridge ESOL examinations are mapped onto a measurement scale common to all examinations. For receptive skills, for example, the stability of the measurement scale is achieved by an item banking methodology that is employed in the development phase where new items are pretested and calibrated using anchor items to monitor exam difficulty. The calibrated items are then stored in the Cambridge ESOL Local Item Banking System (LIBS) where each item has a known difficulty and accordingly examination papers are constructed to a target difficulty on the CEFR A2-C2 continuum and can be graded accordingly to a high degree of precision. This is better described as *standard-maintaining* rather than *standard setting*, given that the standard is a stable one which is carried forward. The current rationale for the standard of the objective papers owes something to an essentially normative view of skill profiles in a European context (as, probably, does the CEFR),

and something to the progression depicted by the common measurement scale, which can be represented as a rational ‘ladder’ of learning objectives.

Empirical Validation

We will focus here on *Internal Validation* since *External Validation* of the linking process is still in its infancy worldwide with no published studies at present demonstrating external validation. It is worth mentioning though that Cambridge ESOL as a member of ALTE carried out in 1998-2000 a major project using the ALTE Can Do scales (Jones 2000, 2001, 2002) and Cambridge ESOL examinations, providing a strong empirical link between test performance and perceived real-world language skills.

Internal Validation as described by the *Manual* relates to a very wide range of activities at the different phases of Cambridge ESOL test development and administration cycle. It is related to:

- the planning stage where data is gathered on the targeted candidature through reliable and valid data collection instruments
- the design and development stages where data is collected to ensure that the examination is reliable, valid, and practical
- the test administration stage where quality assurance procedures are in place to ensure fair administration procedures
- the post exam and review stage where data is collected on the washback effect of the examination.

Cambridge ESOL follows a socio-cognitive approach towards test validation. This approach considers and provides evidence for the following components of validity: cognitive, contextual, scoring, criterion-related and consequential. Linking to the CEFR is embedded in all of these. Questions like “what makes a B2 reader differ from a B1 reader in terms of cognitive ability?” or “what type of task is most suited to assess the ability to integrate information from a variety of reading texts which is a mark of a C1/C2 reader?” are posed, answered and documented (See for example Shaw & Weir 2006 on examining writing).

Internal validation is carried out through a variety of activities:

- Statistical analyses of objective items before (pretest) and after live sessions. This includes the use of anchor tests, and information about candidates gathered each session via candidate information sheets;
- Qualitative analysis of Writing and Speaking tasks before (trailing) and after live sessions which is documented in examiner, senior team leader and annual validation reports.
- Statistical analysis of Writing examiners marking tendencies and monitoring via the Team Leader system through the entire marking period, in addition to a systematic ‘marks collection’ exercise in Speaking co-ordination and in the monitoring process in live sessions.

Conclusion

In its current draft form, the *Manual* appears to envisage the provision of different types of evidence demonstrating examination alignment to the CEFR and showing the quality of the methodology. The authors of the *Manual* appear to envisage that users will apply *Manual* procedures rationally

and selectively and through a reflection on this application, users contribute to a body of knowledge and experience and add to the suggested techniques. However, in order to sustain alignment claims it is imperative to bring explicit CEFR reference in test providers' practices, processes and documentation on an ongoing basis coordinated for practical purposes with revisions and updates. It is worth noting here that the CEFR as a concept appeared to function quite well in the past without extensive underpinning from measurement theory and statistics. However, these are becoming more and more important as attempts are being made to validate aspects of the CEFR empirically (North and Schneider 1998) and to align examinations to it (Kaftandjieva, 2004).

As the regulatory function of the CEFR gathers pace, we need to caution here that while frameworks carry certain benefits to a variety of stakeholders, e.g., facilitating selection from a range of examinations, they also have limitations. As Taylor (2004) states "they risk masking significant differentiating features, they tend to encourage oversimplification and misinterpretation, and there is always a danger that they are adopted as prescriptive rather than informative tools". The purpose of any linking exercise is to provide a framework of how tests and levels relate to each other in broad terms within a common frame of reference. This is of particular value to end users especially in the globalized world in which we now live. The major challenge, therefore, for language testers, at least in Europe, is to begin to look explicitly at direct cross-language comparison. This will need new methodologies and kinds of evidence, but provides the best hope of a better answer to the question: "How does my B1 compare with your B1?, In what way do they vary?"

References

Council of Europe (2008) *Explanatory Notes to Recommendation CM/Rec (2008)7 of the Committee of Ministers to member states concerning the use of the Common European Framework of Reference for Languages (CEFR) and the promotion of plurilingualism*. Retrieval from <https://wcd.coe.int/>

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

Council of Europe (2003) *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEF)*. Manual:Preliminary Pilot Version. DGIV/EDU/LANG 2003, 5. Strasbourg: Language Policy Division.

Council of Europe (2004) *Reference supplement to the Preliminary Pilot Version of the Manual for Relating language examinations to the CEFR*.

Jones, N (2000) Background to the validation of the ALTE Can Do Project and the revised Common European Framework, *Research Notes* 2, 11-13.

Jones, N (2001) The ALTE Can Do Project and the role of measurement in constructing a proficiency framework, *Research Notes* 5, 5-8.

Jones, N (2002) *Relating the ALTE Framework to the Common European Framework of Reference*. In Alderson, J.C (ed.) *Case Studies in applying the Common European Framework*, Strasbourg, Council of Europe. pp167-183.

Ingham, K (2008) The Cambridge ESOL approach to item writer training: the case of ICFE Listening. *Research Notes* 32, pp 5-9.

Kaftandjieva, F (2004). Standard setting. *Section B, Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment*. Strasbourg: Council of Europe. Retrievable from http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp

Martyniuk, W (2008) Relating language examinations to the Council of Europe's Common European Framework of Reference, in Taylor, L and Weir C J (Eds) *Multilingualism and Assessment*, Studies in Language Testing 27, Cambridge: UCLES/Cambridge University Press.

Moe, E (2008) Juggling numbers and opinions: an attempt to set CEFR standards in Norway for a test of reading in English, in Taylor, L and Weir C J (Eds) *Multilingualism and Assessment*, Studies in Language Testing 27, Cambridge: UCLES/Cambridge University Press.

North, B (2008) The CEFR Levels and Descriptor Scales in Taylor, L and Weir C J (Eds) *Multilingualism and Assessment*, Studies in Language Testing 27, Cambridge: UCLES/Cambridge University Press.

North, B. & Schneider, G (1998) Scaling descriptors for language proficiency scales. *Language Testing* 15, 217-62.

Saville, N (2003) The process of test development and revision within UCLES EFL, in Weir, C and Milanovic, M (Eds.) *Continuity and innovation: revising the Cambridge Proficiency in English Examination 1913-2002*, Cambridge: UCLES/Cambridge University Press, 57-120.

Shaw, S & Weir, C.J (2007). *Examining Second Language Writing: Research and Practice*. Studies in Language Testing. Vol 26. Cambridge: Cambridge ESOL CUP.

Taylor, L (2004) IELTS, Cambridge ESOL examinations and the Common European Framework, *Research Notes*, 18, pp2-3.

Taylor, L & Jones, N (2006) Cambridge ESOL exams and the Common European Framework of Reference (CEFR). *Research Notes*, 24, 2-5. Cambridge: Cambridge ESOL.

Van Ek, J & Trim, J (1998) *Threshold 1990*. Cambridge: Cambridge University Press.

Van Ek, J & Trim, J (1998) *Waystage 1990*. Cambridge: Cambridge University Press.

Van Ek, J & Trim, J (2001) *Vantage*. Cambridge: Cambridge University Press.

Wu J R W & R Y F Wu (2007) Using the CEFR in Taiwan: the perspective of a local examination board. The Language Training and Testing Center, Taipei: Taiwan. Mimeo, paper presented at EALTA conference, Sitges, Spain July 2007.