

**Relating the Password Test
(Knowledge, Reading and Writing)
to the
Common European Framework of Reference for Languages**

**Report of the CEFR Alignment Panel Meeting
11th July 2012**

**Dr Tony Green
Reader in Language Assessment
CRELLA, The University of Bedfordshire**

Contents

Introduction	1
The CEFR	3
Specification.....	5
Password (language) Knowledge	7
Password Reading	9
Password Writing	11
The CEFR linking panel	14
Participants	14
Familiarisation training	15
Training in assessing performance in relation to the CEFR levels using standardised samples	16
Password Writing rating scales	16
Rating Password Writing essay performance in relation to the CEFR levels	17
Password Reading	17
Password (language) Knowledge	18
Review of the standard setting experience	18
Results: Familiarisation	19
CEFR linking – general	22
CEFR linking result - Password Writing	23
Password Writing essay rating	25
CEFR linking result - Password Reading	26

Relating Password (Knowledge, Reading & Writing) to the CEFR August 2012

CEFR linking result - Password (language) Knowledge	28
Comments on the panel event	30
References	32
Appendix. Distribution of Password essay scores	33

Introduction

English Language Testing Ltd (ELT), the creators of the Password¹ tests and the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire, its academic designers and managers are pleased to present the results from the Common European Framework of Reference for Languages (CEFR) linking panel event held on the 11th of July 2012.

The CEFR panel considered three Password test modules; Password (language) Knowledge, Password Reading and Password Writing.

This document details the process and outcomes; it is clear that the panel were satisfied that they reached well-grounded decisions on the Password - CEFR relationships.

A summary of the relationships is given in the Tables below and overleaf.

CEFR Level	Password (language) Knowledge Score
A0	Pre-Password
A1	2.0, 2.5
A2	3.0, 3.5
B1	4.0, 4.5, 5.0
B2	5.5, 6.0, 6.5
C1	Password 7.0 or higher

¹ The Password logo is a registered trademark of ELT. All other trademarks are acknowledged.

CEFR Level	Password Reading Score
A0	Pre-Password
A1	2.0, 2.5
A2	3.0, 3.5
B1	4.0, 4.5, 5.0
B2	5.5, 6.0, 6.5
C1	Password 7.0 or higher

CEFR Level	Password Writing Score
A0	Pre-Password
A1	2.0, 2.5
A2	3.0, 3.5
B1	4.0, 4.5, 5.0
B2	5.5, 6.0, 6.5
C1	Password 7.0 or higher

ELT and CRELLA undertake regular and rigorous reviews of Password test results. Over time, as new material is introduced into the test and results analysed, this may mean that minor changes have to be made to the number of points needed to obtain a given Password band and/ or CEFR level. This ensures that, as far as possible, the standard of English required to obtain each Password band score and/ or CEFR level remains constant.

The CEFR

The Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Council of Europe, 2001), known as CEF or CEFR, is designed to bring together the best of a wide range of different schemes for describing language learning levels. The CEFR summarises the scope of a global consensus on levels of functional communicative language ability lays out options for successful language learning, posing important questions to guide the development of resources for teaching, learning and assessment.

In the process of developing levels for the CEFR, *Can Do* statements were gathered from 30 different schemes, giving an initial pool of over 2,000 statements to be considered for inclusion (North, 2000). These were screened for repetition and then approved, rejected or edited by the CEFR authors (who reduced the number by half) to leave a set of meaningful, “positively worded, ‘stand-alone’ statements” (North, 2000, p.184). Users of the framework should be able to understand each statement, without having to compare it with other statements, and be able to relate the statements to learners’ practices and abilities. The framework is intended to provide a shared language for setting objectives, developing materials and making comparisons between systems. Learners should be able to look at a statement and say, ‘yes, I can do this;’ ‘no. I can’t do this;’ or ‘I’d like to learn to do this’ when using a given language.

Through having teachers rank the statements and use them to judge the performance of learners, North (2000) constructed an overarching set of common reference levels representing an expanding communicative language ability. At the end of the development process, each *Can Do* statement and each performance had been ‘calibrated:’ given a mathematical value to allow estimates of the probability that a learner judged able to perform one activity would be able to perform any of the other activities appearing on the scale. The teachers involved in the development were also asked to group the statements into categories to represent different aspects of language use.

The outcome of these processes was a network of interrelated scales (Chapters 4 and 5 of the CEFR) representing a range of features that might be addressed in language education around the world, ordered according to a consensus view of their relative difficulty and illustrative of loosely defined

underlying levels of ability. In the ‘branching approach’ adopted, distinctions are made between three general levels of learner (A: basic, B: independent and C: proficient), while the scales provided in the CEFR generally define differences at six levels: A1 and A2, B1 and B2 and C1 and C2. Finer-grained divisions are often made, as between B2.1 and B2.2. Yet more delicate distinctions can be made within each level so that the relatively small gains in language proficiency occurring within language programmes (achievement) can be captured and reported. The CEFR scales are not a closed system: additional sets of *Can Do* statements, such as the ALTE *Can Do* scales presented at as Appendix D of the CEFR (Council of Europe [COE], 2001, p. 244ff), can be integrated into the CEFR scheme using appropriate statistical techniques.

The CEFR *Descriptive Scheme* and the *Common Reference Levels* together provide a conceptual grid which users can utilize to describe their systems. Since its publication, the CEFR has been widely adopted as a convenient means of comparing language qualifications designed for different purposes, produced in different countries and developed by different assessment agencies. The adoption of CEFR levels by policy makers in setting targets for language learning achievement or requirements for the language abilities of migrants has encouraged testing agencies to link their qualifications to the CEFR and encouraged the Council of Europe to provide guidance on defensible methods for doing so. Such linking does not imply an equivalence between the results of different testing systems related to the CEFR at the same level (different tests are developed with different purposes in mind), but, when taken in conjunction with an awareness of the test content and coverage, should help users and others to ‘locate’ the test in CEFR terms.

To facilitate this process, the Council of Europe published *Relating examinations to the Common European Framework of Reference: A Manual* (henceforth referred to as ‘the Manual’) released in draft form in 2003 and updated in 2009.

The Manual presents three related sets of procedures that users are advised to follow in order to link assessments to the CEFR:

- Specification of examination content.
- Standardisation of judgments.
- Empirical validation through analysis of test data.

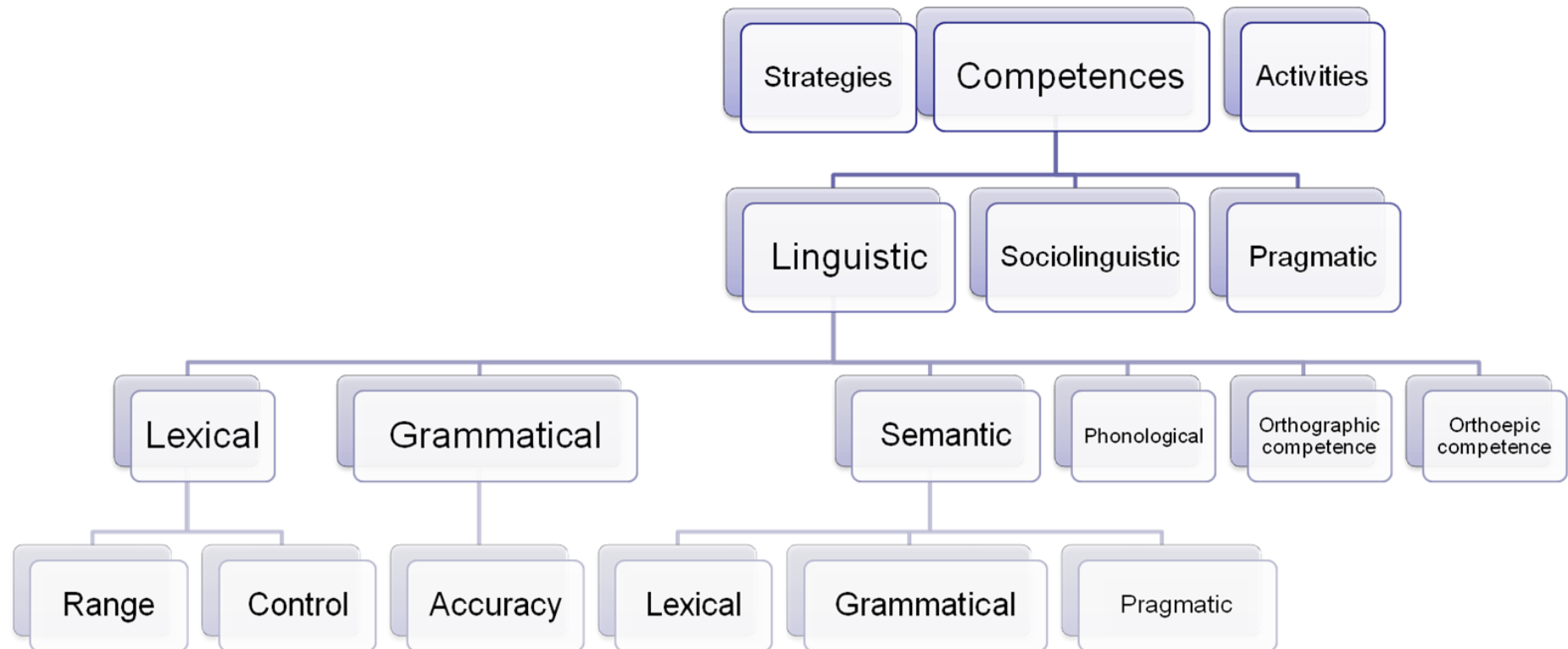
However, the Manual stresses that it does not provide the 'sole guide to linking a test to the CEFR' (p.1). Indeed, the process of linking it advocates has attracted some criticism where such exercises are conducted as 'one-off events' (Milanovic 2010, p.4) rather than embedded in the routine operation of an assessment.

The project described here broadly follows the steps advocated in the Manual, but also provides an account of how the CEFR levels are embedded in the approach to the routine production of test material.

Specification

This initial phase in relating Password to the CEFR requires identification of areas of overlap between the two. There is a 'horizontal' comparison to be made about test content in relation to the CEFR descriptive scheme (see Figure 1) – judgements about which aspects of the CEFR are addressed by Password – and a 'vertical' comparison with relevant level descriptions – judgements about which level on Password might correspond to which level on the CEFR. The objective is to arrive at a preliminary overview of the relationship which may then be explored in greater depth in later phases.

Figure 1 The Common European Framework of reference: the user/learner's competences (Chapter 5)



Password (language) Knowledge

Password (language) Knowledge was developed on the basis of Weir's (2005) socio-cognitive framework for test validation. The chief concern is with the processing of language at the word and sentence levels fundamental to both comprehension and production and with the academic social context: the test is concerned with the language used in academic textbooks, in student writing and that encountered in the daily lives of students.

Detailed test specifications have been developed to reflect the core language knowledge that students need to acquire before they will be able to cope with understanding and producing academic texts. These specifications are based on a number of sources:

- Research carried out by the Centre for Research in English Language Learning and Assessment, University of Bedfordshire into the nature of academic language use (Weir et al 2009).
- Communicative functions (and their associated grammar and vocabulary) found in popular English text books that are widely used on pathway programmes preparing learners for academic study through the medium of English.
- The Breakthrough, Threshold and Vantage specifications for English (the T-Series) describing the A2, B1 and B2 (and above) levels of the Common European Framework of Reference for Languages (van Ek and Trim 1991a, 1991b, 2000, Council of Europe 2001).
- Common patterns of error in grammatical structures and vocabulary choice made by pathway learners in their written work.
- Research evidence on the essential grammar and vocabulary needed to support academic study (Weir 1983; Nation 1990). We used corpus based wordlists such as the academic wordlist (Coxhead 2000) and word frequency lists based on the British National Corpus to identify words that learners would need to know in order to access academic texts across disciplines.
- Grammar and vocabulary books designed for learners of English such as Murphy (2004) and McCarthy and O'Dell (2008) and reference books such as Greenbaum and Quirk (1993), Carter and McCarthy (2006) and Schmitt (2000).

In relation to the horizontal aspect of content coverage, the Password test is concerned more with the learner's competences described in Chapter 4 than with the activities described in Chapter 3 of the CEFR. For reasons given in the Password design document, the language Knowledge test does not directly elicit spoken production, reception, interaction or mediation, but is largely limited to written reception with an option for users to collect samples of written production to supplement this. In relation to competences (see the representation of aspects of the Descriptive Scheme in Figure 1), the test is focussed on lexical and grammatical competences, although aspects of semantic competence are also invoked. Of particular relevance are the illustrative scales for grammatical accuracy and for vocabulary control and range given below.

In relation to the vertical relationship, as noted above, the test specifications embed the grammatical and lexical specifications found in the Council of Europe T-Series of specifications for English, which serve to elaborate the CEFR levels.

CEFR: Grammatical accuracy (CEFR p.114)

- C2** Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).
- C1** Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot. Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.
- B2** Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding. Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.
- B1** Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.
- A2** Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.
- A1** Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.

CEFR: Vocabulary control (CEFR p.112)

- C2** Consistently correct and appropriate use of vocabulary.
- C1** Occasional minor slips, but no significant vocabulary errors.
- B2** Lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication.
- B1** Shows good control of elementary vocabulary but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.
- A2** Can control a narrow repertoire dealing with concrete everyday needs.
- A1** No descriptor available

CEFR: Vocabulary range (CEFR p.112)

- C2** Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.
- C1** Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.
- B2** Lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication.
- B1** Shows good control of elementary vocabulary but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.
- A2** Can control a narrow repertoire dealing with concrete everyday needs.
- A1** Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.

Password Reading

Password Reading was developed with the levels of the CEFR in mind. The test employs texts designed to reflect the demands of the A2 to the C1 level based on the empirical work of Green (2012) in identifying the distinctive features of texts at each level in terms of their grammatical, lexical and discursive features.

As reflected in the specifications, the content of the texts is intended to reflect a trajectory leading from relatively general texts with a social and broadly informative rhetorical function to the more argumentative text types encountered in academic study. The texts range from those of a social,

interactional nature (emails, postcards, short personal letters) to those of a type found in serious magazines and introductory text books.

The tasks at the lower levels of the test focus on local processing of words and explicitly stated propositions. The higher levels require test takers to connect ideas at the paragraph level and to form a conceptualisation of the meaning of a text as a whole.

The key activities from the CEFR that were identified as relating closely to the content of Password Reading are as follows:

Reading Correspondence (CEFR p.69)

- C2** No descriptor available
- C1** Can understand any correspondence given the occasional use of a dictionary.
- B2** Can read correspondence relating to his/her field of interest and readily grasp the essential meaning.
- B1** Can understand the description of events, feelings and wishes in personal letters well enough to correspond regularly with a pen friend.
- A2** Can understand basic types of standard routine letters and faxes (enquiries, orders, letters of confirmation etc.) on familiar topics
Can understand short simple personal letters.
- A1** Can understand short, simple messages on postcards.

Reading for Orientation (CEFR p.70)

- C2** No descriptor available
- C1** No descriptor available
- B2** Can scan quickly through long and complex texts, locating relevant details.
Can quickly identify the content and relevance of news items, articles and reports on a wide range of professional topics, deciding whether closer study is worthwhile.
- B1** Can scan longer texts in order to locate desired information, and gather information from different parts of a text, or from different texts in order to fulfil a specific task.
Can find and understand relevant information in everyday material, such as letters, brochures and short official documents.
- A2** Can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus, reference lists and timetables.

Can locate specific information in lists and isolate the information required (e.g. use the "Yellow Pages" to find a service or tradesman).

Can understand everyday signs and notices: in public places, such as streets, restaurants, railway stations; in workplaces, such as directions, instructions, hazard warnings.

- A1** Can recognise familiar names, words and very basic phrases on simple notices in the most common everyday situations.

Reading for Information & Argument (CEFR p.70)

- C2** No descriptor available

C1 Can understand in detail a wide range of lengthy, complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.

- B2** Can obtain information, ideas and opinions from highly specialised sources within his/her field.

Can understand specialised articles outside his/her field, provided he/she can use a dictionary occasionally to confirm his/her interpretation of terminology.

Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

- B1** Can identify the main conclusions in clearly signalled argumentative texts.

Can recognise the line of argument in the treatment of the issue presented, though not necessarily in detail.

Can recognise significant points in straightforward newspaper articles on familiar subjects.

- A2** Can identify specific information in simpler written material he/she encounters such as letters, brochures and short newspaper articles describing events.

- A1** Can get an idea of the content of simpler informational material and short simple descriptions, especially if there is visual support.

Password Writing

The Password Writing tasks elicit a short (200 words in 30 minutes) expository essay on a generally accessible topic. The tasks reflect the kind of writing commonly required in English for Academic Purposes programmes and on tests of academic language skills. The responses are scored on four rating criteria: *Arguments and evidence; Organization; Grammar/ vocabulary range; Grammar/ vocabulary control.*

The latter two criteria clearly reflect the same areas of competence described in the CEFR as the Password language Knowledge test, but in relation to less controlled, productive language use. The

first two criteria involve elements of other competencies described in the CEFR such as *Coherence and Cohesion* and *Propositional Precision*. There is also a clear correspondence between the Password Writing task and the CEFR activity and illustrative scale for *Reports and Essays*.

Coherence and Cohesion (CEFR, p.125)

- C2** Can create coherent and cohesive text making full and appropriate use of a variety of organisational patterns and a wide range of cohesive devices.
- C1** Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.
- B2** Can use a variety of linking words efficiently to mark clearly the relationships between ideas.
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.
- B1** Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
- A2** Can use the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points.
Can link groups of words with simple connectors like "and", "but" and "because".
- A1** Can link words or groups of words with very basic linear connectors like 'and' or 'then'.

Propositional Precision ('the ability to formulate thoughts and propositions so as to make one's meaning clear' CEFR, p.129)

- C2** Can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of qualifying devices (e.g. adverbs expressing degree, clauses expressing limitations).
Can give emphasis, differentiate and eliminate ambiguity.
- C1** Can qualify opinions and statements precisely in relation to degrees of, for example, certainty/uncertainty, belief/doubt, likelihood etc.
- B2** Can pass on detailed information reliably
- B1** Can explain the main points in an idea or problem with reasonable precision.
Can convey simple, straightforward information of immediate relevance, getting across which point he/she feels is most important.
Can express the main point he/she wants to make comprehensibly.

- A2** Can communicate what he/she wants to say in a simple and direct exchange of limited information on familiar and routine matters, but in other situations he/she generally has to compromise the message.
- A1** No descriptor available

Reports and Essays (CEFR, p.62)

- C2** Can produce clear, smoothly flowing, complex reports, articles or essays which present a case, or give critical appreciation of proposals or literary works.
- Can provide an appropriate and effective logical structure which helps the reader to find significant points.
- C1** Can write clear, well-structured expositions of complex subjects, underlining the relevant salient issues.
- Can expand and support points of view at some length with subsidiary points, reasons and relevant examples.
- B2** Can write an essay or report that develops an argument systematically with appropriate highlighting of significant points and relevant supporting detail.
- Can evaluate different ideas or solutions to a problem.
- Can write an essay or report which develops an argument, giving reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options.
- Can synthesise information and arguments from a number of sources.
- B1** Can write short, simple essays on topics of interest.
- Can summarise, report and give his/her opinion about accumulated factual information on familiar routine and non-routine matters within his field with some confidence.
- Can write very brief, reports to a standard conventionalised format, which pass on routine factual information and state reasons for actions.
- A2** No descriptor available
- A1** No descriptor available

Overall Written Production (CEFR, p.61)

- C2** Can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points.

- C1** Can write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion.
- B2** Can write clear, detailed texts on a variety of subjects related to his field of interest, synthesising and evaluating information and arguments from a number of sources.
- B1** Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence.
- A2** Can write a series of simple phrases and sentences linked with simple connectors like “and”, “but” and “because”.
- A1** Can write simple isolated phrases and sentences.

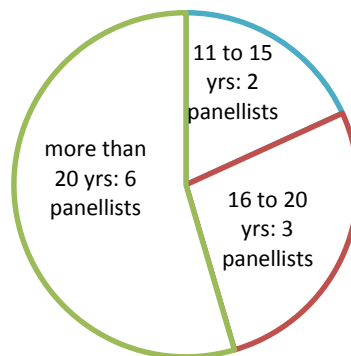
The CEFR linking panel

As recommended in the Manual, a group of panellists was assembled for a CEFR linking panel event. The following sections describe the participants and procedures followed.

Participants

The panel was made up of eleven experienced and well-qualified English language educators. All worked in the Higher Education sector in the UK and had over ten years’ experience in ELT. Three worked as lecturers in language assessment. Six claimed 20 or more years in the field (Figure 2). Two of the panellists held PhDs in applied linguistics. Of the remaining nine, seven had Masters level qualifications and three of these were studying towards PhDs. The two panellists who did not hold a Masters degree both held Cambridge ESOL Diplomas in English Language Teaching for Adults (DELTA). Nine of the eleven had previous experience of CEFR linking panels. Reporting on a five point scale (*Zero/ Poor/ Basic/ Good/ Thorough* knowledge of the CEFR), eight claimed a ‘Good’ and the remaining three a ‘Basic’ knowledge of the CEFR. Panel members’ experiences of the Framework included test design, course design and student assessment. Nine of the panellists were independent of ELT/ Password, the remaining two had experience as question item writers for the test.

Figure 2 Panellists’ responses to the question, ‘How many years have you worked in English language education?’



Familiarisation training

Prior to the panel event, the panellists were asked to familiarise (or re-familiarise) themselves with the Common European Framework levels by reviewing the illustrative scales. Specifically, they were directed to the general reference levels presented on pages 28-29 of the English version of the Framework, Section 3.6, which provides an overview of salient features of each level, and relevant illustrative scales from Chapters 4 and 5 of the CEFR. Those who were less familiar with the framework or who felt that they needed further guidance were also asked to undertake self-access training using the *CEFRtrain* website (www.helsinki.fi/project/ceftrain). As a check on their interpretation, all panellists carried out an online descriptor sorting activity in which they identified the CEFR levels of thirty-two descriptors taken from areas of the Framework that were of greatest relevance.

The panel event itself opened with a series of familiarisation activities, as recommended in the Council of Europe Manual for relating Language Examinations to the Common European Framework of Reference for Languages (Council of Europe 2009).

For the purpose of group discussion, the panellists were arranged around tables in three groups of three and one pair. First, working individually, each panellist sorted a jumbled set of descriptors based on Table A1 of the Manual (a summary of Section of 3.6 of the CEFR) from which references to the levels had been removed. The panellists then discussed their decisions within their group,

justifying or reconsidering their choices. This was followed by a plenary discussion. Table A1 was then distributed.

Next, the panellists were given copies of the Self-Assessment Grid (Table 2 of the CEFR) and asked to rate themselves in one or two languages other than their first language. This was followed by a general discussion on issues arising from the process including consideration of variable skills profiles and the implications of different domains of language use.

Training in assessing performance in relation to the CEFR levels using standardised samples

The panellists viewed illustrative samples of learners speaking English from the training DVD supplied by the Council of Europe and awarded scores expressed in terms of CEFR levels to each. They then discussed and justified their scores to their group and to the panel as a whole. Finally, the panellists compared their scores and the justifications they had given with those provided by the Council of Europe.

Password Writing rating scales

Password Writing test takers are administered an essay. Responses are made available to the receiving institution which may score them using analytic rating scales, guidelines and standardised sample performances supplied by ELT. Scores are awarded in relation to scales that range from 'Pre-Password' to 'Password 7.0 or higher' on four rating criteria: *Arguments and evidence; Organization; Grammar/ vocabulary range; Grammar/ vocabulary control*. There are six defined levels and the option to award half bands (i.e. 3.5 or 5.5 etc.) to essays that satisfy the requirements of a higher level on some of the scoring criteria.

The panellists were given copies of the rating scales and compared these with the illustrative scales of the CEFR. On this basis, they suggested which CEFR levels best corresponded to each overall band level on the Password rating scales. The panellists justified and discussed their judgements and were given an opportunity to revise these before making their individual recommendations.

Rating Password Writing essay performance in relation to the CEFR levels

The panellists were given eighteen learner essays to rate, assigning an overall CEFR level while referring to the CEFR scales for 'Reports and Essays' and 'Overall Written Production', but awarding a single, holistic score. Of the eighteen essays, sixteen came from the Password administration (including one that was repeated from the initial set of five as a check on rater consistency – one rater awarded a higher score on the second occasion) and two were previously standardised samples from the Council of Europe illustrative samples. These provided a further check that the panellists' judgements remained in line with the CEFR.

The panellists first rated a set of six essays and results were recorded. They then compared and discussed their ratings, exploring areas of disagreement. After this round of discussion, they worked individually to rate the remaining twelve essays. These were presented in two batches of six. Two groups of panellists began by rating the first batch; the other two groups began with the second batch. Following independent rating and recording of results, there was a brief discussion within each group, then the batches were exchanged.

Password Reading

The panellists were given sample reading material from the Council of Europe sample DVD. These included three texts (each accompanied by a single item) from DIALANG (Alderson and Huhta 2005) and one text with eight items from the Cambridge ESOL KET examination. As with the samples of spoken performance, the panel rated these individually, judging the input texts and items in relation to the CEFR levels. In making their judgements, they responded to the questions, 'At what CEFR level can a test taker already reach a reasonably complete understanding of this text?' and 'At what CEFR level can a test taker already answer the following item correctly?' After making each judgement, the group first discussed their decisions then consulted the Council of Europe ratings and the supporting evidence.

The panel were then divided into two groups. Each group was presented with a form of the Password Reading test: one targeting lower and the other targeting higher level learners. Consulting the CEFR scales for 'Overall Reading Comprehension', 'Reading for Orientation' and 'Reading for Information and Argument' and working individually, each panellist judged the difficulty of each text and decided how many items they believed that a learner minimally satisfying the criteria for each CEFR level would score on each part of the test. The lower level test had five parts and the higher had three. Each test part consists of an input text and set of items. Following discussion within each group, the panellists then exchanged test forms and repeated the exercise.

Password (language) Knowledge

For the penultimate session of the day, the panel carried out a similar activity in relation to Password (language) Knowledge (Grammar and Vocabulary). This time, each of the two groups was given a full test and the panellists judged the number of items that a learner minimally satisfying the criteria for each CEFR level would score on each part of the test. The test has five parts and each test part consists of items of a certain format. After discussion within the groups, the test forms were exchanged and the exercise repeated with the second test form.

Review of the standard setting experience

Finally, the panellists were given the opportunity to comment on the experience of the panel event and whether they felt that they had been able to arrive at reasonable decisions regarding the relationship between Password and the CEFR. They discussed the experience and responded to a feedback form based on a template provided by the Council of Europe (2009).

Results: Familiarisation

Table 1 Distribution of panellists' judgements of CEFR levels

Descriptor (with CEFR level in brackets)	A1	A2	B1	B2	C1	C2
1. Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire (A1)	10	1	0	0	0	0
2. Can use basic sentence patterns and communicate with memorised phrases, groups of a few words and formulae about themselves and other people (A1)	8	2	1	0	0	0
3. Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required (A1)	11	0	0	0	0	0
4. Can get an idea of the content of simpler informational material and short simple descriptions, especially if there is visual support (A1)	8	3	0	0	0	0
5. Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forgets to mark agreement (A2)	1	8	2	0	0	0
6. Can link groups of words with simple connectors like 'and', 'but' and 'because' (A2)	1	7	3	0	0	0
7. Can identify specific information in simpler written material he/she encounters such as letters, brochures and short newspaper articles describing events (A2)	3	7	1	0	0	0
8. Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items (A2)	2	9	0	0	0	0
9. Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations (B1)	0	5	6	0	0	0

10. Can link a series of shorter, discrete simple elements into a connected, linear sequence of points (B1)	1	5	5	0	0	0
11. Has sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events (B1)	0	2	9	0	0	0
12. Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension (B1)	0	0	7	4	0	0
13. Can identify the main conclusions in clearly signalled argumentative texts (B1)	0	1	5	5	0	0
14. Can recognise significant points in straightforward newspaper articles on familiar subjects (B1)	1	3	7	0	0	0
15. Can write short, simple essays on topics of interest (B1)	1	1	8	1	0	0
16. Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence (B1)	0	2	7	1	1	0
17. Can use some complex sentence forms to express viewpoints and develop arguments (B2)	0	0	2	6	3	0
18. Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding (B2)	0	0	0	9	2	0
19. Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse (B2)	0	1	7	3	0	0
20. Can scan quickly through long and complex texts, locating relevant details (B2)	0	0	0	4	6	1
21. Can obtain information, ideas and opinions from highly specialised sources within his/her field (B2)	0	0	1	2	5	3
22. Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints (B2)	0	0	0	10	1	0

23. Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively (B2)	0	0	0	4	4	3
24. Can synthesise information and arguments from a number of sources (B2)	0	0	0	7	4	0
25. Can write an essay or report that develops an argument systematically with appropriate highlighting of significant points and relevant supporting detail (B2)	0	0	0	7	3	1
26. Occasional minor slips, but no significant vocabulary errors (C1)	0	0	0	1	10	0
27. Can expand and support points of view at some length with subsidiary points, reasons and relevant examples (C1)	0	0	0	5	6	0
28. Can write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion (C1)	0	0	0	0	7	4
29. Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot (C2)	0	0	0	0	5	6
30. Can understand a wide range of long and complex texts, appreciating subtle distinctions of style and implicit as well as explicit meaning (C2)	0	0	0	0	4	7
31. Can provide an appropriate and effective logical structure which helps the reader to find significant points (C2)	0	1	1	4	3	2
32. Can produce clear, smoothly flowing, complex reports, articles or essays which present a case, or give critical appreciation of proposals or literary works (C2)	0	0	0	0	0	11

The online familiarisation activity reflected the panellists' awareness of the CEFR. An average of 7.14 of the eleven panellists (min. 2, max 11, standard deviation 2.07) correctly identified each descriptor. Only two statements (descriptors 3 and 32 in Table 1) were correctly placed by all panellists,

although a majority of panellists correctly placed 25 (78%) of the descriptors. In two cases (descriptors 20 and 21 in Table 1), a majority of the panellists placed the statement one level above its CEFR position (both were B2 level descriptors placed by a majority of panellists at C1). In one case, a statement was placed one level below its CEFR position by the majority: descriptor 19 in Table 1 was incorrectly placed at B1 rather than at B2. Panellists were equally divided between those placing descriptors 10 (B1) and 14 (B2) at their correct CEFR level and those placing them at an adjacent level. In both cases, the majority of the incorrect panellists underestimated the level. The most problematic descriptor proved to be number 31 in Table 1. This produced the widest spread of judgements with at least one panellist placing it at every level between A2 and C2: four identifying it as B2, three as C1 and just two as C2. However, this was the only descriptor that gave rise to such a wide level of variation.

Of the 352 judgements made by the panellists, 223 (63.4%) were in line with the CEFR, 109 (31.0%) descriptors were placed one level above or below the CEFR (46 higher and 68 lower) and 20 (5.7%) were two or more levels away from the CEFR position (ten higher and ten lower). Of these 20, six came on the problematic descriptor 30 in Table 1.

The panellist who was most closely in line with the CEFR correctly identified the level of 26 (81%) of the descriptors. The panellist correctly placing the fewest descriptors (15 – 47%) was consistently harsh in his judgements. He placed 13 descriptors one band below their CEFR level and two descriptors two bands below their CEFR level compared with four descriptors one band above their CEFR level. This panellist had reported his CEFR knowledge as ‘Basic’ although he had worked quite closely with the CEFR some years before.

Overall, this preliminary familiarisation exercise confirmed the panellists’ view that they had a good understanding of the Common European Framework, but suggested that there was some scope for improvement through further familiarisation activities during the panel event.

CEFR linking – general

Where there is no clear match between a Password band score and the CEFR boundaries – where a CEFR boundary falls between two Password bands, a decision must be made on whether to be more

cautious and to link the Password Band with the lower CEFR level or with the higher. In general, we take a cautious approach. This means that learners scoring at a high level within a Password band might, according to the panel’s recommendations, fall within a higher CEFR level than indicated in the Tables. In a few cases, where the CEFR boundary is only marginally higher than a Password level boundary, we have linked the two even though the lowest scoring Password test takers within the band might be placed by the panel at a lower CEFR level.

ELT and CRELLA undertake regular and rigorous reviews of Password test results. Over time, as new material is introduced into the tests, this may mean that minor changes have to be made to the number of points needed to obtain a given Password band or CEFR level. This ensures that, as far as possible, the standard of English required to obtain each Password band score remains constant.

CEFR linking result - Password Writing

On the basis of their content analysis of the Password rating scales for Writing, the panellists assigned a CEFR level to each Password band. The results of this analysis are shown in Table 2.

Table 2 Comparisons between the Password Writing scales and the CEFR by eleven panellists

	Pre-Password	2.0	3.0	4.0	5.0	6.0	Password 7 or higher
Panellist 1	A0	A1	A2	A2	B1	B2	B2+
Panellist 2	A0	A1	A1	A2	B1+	B2+	C1
Panellist 3	A0	A1	A2	B1	B1+	B2+	C1
Panellist 4	A0	A1	A2	B1	B1+	B2+	C1
Panellist 5	A0	A1	A2	B1	B1+	B2	B2+
Panellist 6	A0	A1	A2	A2	B1	B2	B2+
Panellist 7	A0	A1	A1	A2	B1	B1+	B2+

Panellist 8	A0	A1	A2	B1	B1+	B2	C1
Panellist 9	A0	A0	A1	A2	B1	B2	C1
Panellist 10	A0	A1	A2	B1	B1	B1+	B2+
Panellist 11	A0	A1	A2	B1	B1+	B2	B2+
Mode	A0 (11)	A1 (10)	A2	B1	B1+	B2	B2+

Discussions pointed to the issue of reconciling the uniformly positive ‘Can Do’ wording of the CEFR scales with the negative formulations often used in the Password scales (e.g. *‘ideas are baldly stated and may not be entirely clear to the reader’*). However, the panellists were generally able to find sufficient common ground between the Password scales and the CEFR to justify their decisions.

The panellists unanimously agreed that an A0 learner would achieve Pre-Password and all but one thought that an A1 level would be needed for Password 2. Opinions were more divided over the higher levels with eight of the eleven identifying A2 with Password 3. A narrow majority (six) panellists identified B1 with Password 4, B1+ with Password 5 and B2 with Password 6. At the Password 4 and 5 levels, the five panellists who were in the minority all placed the Password level one band lower on the CEFR (Password 4 A2, 5 B1), but for Password 6, they were divided between B2+ (three panellists) and B1+ (two panellists). For Password 7.0 or higher, the minority of five panellists all selected C1 rather than B2.

Based on the outcomes of this exercise, the following recommendations can be made on the relationship between the CEFR and the Password Writing scales:

Table 3 Password Writing scales: CEFR link

A0	Pre-Password
A1	2.0
A2	3.0

B1	4.0
B1+	5.0
B2	6.0
C1	Password 7 or higher

Password Writing essay rating

Following analysis and discussion of the Password Writing rating scales and the CEFR, the panel then used CEFR scales to give a holistic score expressed in CEFR terms to a series of eighteen essays that had previously been scored using the Password scales. The original (Password) scores were not revealed to the panellists.

First all panellists rated five essays independently. They then discussed their ratings. The remaining essays were rated in two batches. Half of the panellists rated each batch. They then discussed their ratings before the batches were then exchanged. Two sample essays were included from sources other than Password: the Council of Europe and *CEFRain*. Both of these essays were taken from recognised training material and exemplified the crucial B1 level. However, one of the samples (a Cambridge PET essay from the Council of Europe illustrative materials) was rated as a A2+ by the majority (seven) of the panellists. This suggests that the panellists as a group may have been interpreting the CEFR scales marginally more harshly than intended. However, it should also be noted that the Council of Europe and *CEFRain* samples were not academic essays of the kind used with Password (no such samples are available) and this may have affected the ratings.

As a check on whether these scores would reflect operational rating of essays using the Password scales, the modes of the CEFR scores awarded by the panellists (see Appendix) were compared with the average of the Password scores originally awarded to these essays by individual raters using the Password scales. This excludes the two samples from the Council of Europe and *CEFRain*, which did not have Password scores. Although the data is limited, the picture that emerges is consistent with the panellists' analysis of the scales and supports their interpretation. Three of the four essays

awarded B2 had Password scores of 6.0. Of the six essays awarded A2, three had received Password scores of 3.0 and two 3.5.

Table 4 CEFR essay scores compared with original Password scores

Mode of CEFR rating	Number of essays	Average of Password bands
A1	2	2.50
A2	6	2.83
A2+	3	3.50
B1	1	4.00
B2	4	5.63

CEFR linking result - Password Reading

Each panellist, working individually, judged the difficulty of each Password Reading text. They consulted the CEFR scales for ‘Overall Reading Comprehension’, ‘Reading for Orientation’ and ‘Reading for Information and Argument’, and decided how many items they believed that a learner minimally satisfying the criteria for each CEFR level would score on each part of the test. The average of their recommendations was taken as the panel recommendation for the Password Reading score.

Table 5 Average rater judgements of CEFR level thresholds on the Password Reading test

Part	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7	Part 8
<i>Score out of</i>	<i>7 points</i>	<i>6 points</i>	<i>6 points</i>	<i>7 points</i>	<i>5 points</i>	<i>7 points</i>	<i>8 points</i>	<i>6 points</i>
A1	2.18	0.45	0.55	0.18	0.18	0.09	0.00	0.00
A2	4.91	2.73	1.55	0.91	0.82	0.91	0.36	0.00
B1	6.55	4.64	3.64	2.82	2.27	2.82	1.27	0.09
B2	7.00	5.82	4.45	5.27	3.45	4.45	3.64	0.82
C1	7.00	6.00	4.64	6.45	4.55	6.45	6.00	3.36
C2	7.00	6.00	5.00	6.73	5.00	6.91	7.36	4.91

Applying these results to the three levels of Password Reading test suggests the following relationship (see Table 6).

Table 6 Outcome 2: Password Reading score linking

	PR Low	PR Mid	PR High
A1	15%		
A2	39%	33%	
B1	68%	46%	33%
B2		71%	42%
C1			75%

Based on the outcomes of this exercise, the following recommendations can be made on the relationship between the CEFR and Password Reading:

Table 7 Password Reading: CEFR link

A0	Pre-Password
A1	2.0
A2	3.0
B1	4.0
B1+	5.0
B2	6.0
C1	Password 7 or higher

CEFR linking result - Password (language) Knowledge

The panellists judged how many points learners minimally satisfying the criteria for each level of the CEFR would be able to score on each section of the Password (language) Knowledge test. The results of this process are displayed in Table 8.

Table 8 Password (language) Knowledge scores associated with each CEFR level

Level	Part 1	Part 2	Part 3	Part 4	Part 5	Total
A1	1.8	3.6	0.7	0.3	1.9	8.4
A2	4.5	7.2	1.9	0.9	4.4	18.8
B1	9.5	13.8	5.2	2.5	11.6	42.7
B2	12.5	22.4	9.7	4.8	20.0	69.5

C1	14.8	27.5	13.2	6.8	25.8	88.2
C2	15.0	29.6	14.8	8.5	28.7	96.7

As all of the items on these tests have a selected response format (ranging from two to five options), test takers at lower levels might score a certain number of points through random guessing. Although the panel suggested that scores of 19 on Password for A2, the threshold is adjusted upwards to 35 to take account of the possible effects of guessing. If users intend to interpret the A1 level cut score, they must therefore take account of the number of points scored on the easiest items on the test (those at the beginning of each test part). If a score of 15 is obtained from the first few items of each part of the test it is more likely to reflect a genuine level of ability than a score of 15 obtained from items scattered throughout each test part. Although it might be argued that blind guessing could also affect scores above this level, in practice scores on Password at the A2 level and above are consistent with scores obtained from other sources that do not rely on selected response formats (such as the associated writing samples and the IELTS test), suggesting that it is more appropriate not to adjust for possible guessing at these levels.

Recommended cut scores for the Password (language) Knowledge test are given in Table 9.

Table 9 Password (language) Knowledge score linking

	Password (language) Knowledge
A1	15
A2	35
B1	43
B2	69
C1	88

Based on the outcomes of this exercise, the following recommendations can be made on the relationship between the CEFR and Password (language) Knowledge:

Table 10 Password (language) Knowledge: CEFR link

A0	Pre-Password
A1	2.0
A2	3.0
B1	4.0
B1+	5.0
B2	6.0
C1	Password 7 or higher

Comments on the panel event

In a final session, panellists were given a brief questionnaire, based on an example provided in the Council of Europe Manual, and were invited to comment on the event. A summary of the responses is presented in Table 11.

Table 11 Feedback questionnaire responses

Questions (4 Strongly Agree, 3 Agree, 2 Disagree, 1 Strongly Disagree)	Average	Standard deviation
I understood how to answer the pre-seminar online questionnaire	3.82	0.40
I understood the purpose of the seminar	3.73	0.47
I feel I now have a clear understanding of the CEFR levels	3.45	0.52
I understood the instructions for the activities	3.45	0.52
There was adequate time for discussion	3.27	0.65
All participants were allowed to state their views freely	3.64	0.50
The introductory discussions were helpful for me	3.64	0.50
The discussions following the first round of judgements were helpful for me	3.64	0.50
The discussions following the second round of judgements were helpful for me	3.55	0.52
The facilities at the conference centre helped to create a positive working environment	3.64	0.50

I am confident that we reached a reasonable decision on the relationship between CEFR levels and test scores	3.45	0.69
--	------	------

It is clear from their anonymous feedback (in Table 11) that the panellists were satisfied with the event and with their decisions on the relationship between Password and the CEFR. Open-ended responses characterised the day as ‘useful’, ‘productive’ and ‘positive’. One panellist commented that *‘it shows how difficult it is to apply CEFR, but I was encouraged that there was more agreement than disagreement’*. Overall the panel were satisfied that they had reached well-grounded decisions on the Password - CEFR relationships.

References

Alderson, J. Charles., Huhta, Ari (2005). The development of a suite of computer-based diagnosis tests based on the Common European Framework. *Language Testing*, 22 (3), 301

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment 2001, Council of Europe, Cambridge University Press.
www.coe.int/t/dg4/linguistic/source/Framework_en.pdf

Council of Europe (2009). Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A Manual, Strasburg: Language Policy Division. www.coe.int/t/dg4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf

Appendix. Distribution of Password essay scores

<i>Essay</i>	<i>Password</i>	<i>A1</i>	<i>A2</i>	<i>A2+</i>	<i>B1</i>	<i>B1+</i>	<i>B2</i>	<i>B2+</i>	<i>C1</i>	<i>Mode</i>
1	3.5	3	5	3						A2
2	3		5	2	4					A2
3	3	1		6	3	1				A2+
4	6					3	7	1		B2
5	3	4		7						A2+
6	2.5	9	2							A1
7	4.5			6	4	1				A2+
8	B1 (PET)			7	1	3				A2+
9	3	4	6	1						A2
10	6					2	6		2	B2
11	4				6	3	2			B1
12	Pre	4	7							A2
13	3.5	2	7	1	1					A2
14	5	4	4	2						A2
15	B1 (CEFTTrain)		1	2	5	3				B1
16	2.5	10	1							A1
17	4.5			1	3	2	4	1		B2
18	6					3	3	3	2	B2