Relating the Password Test

(Listening and Speaking)

to the

Common European Framework of Reference for Languages

Report of the CEFR Alignment Panel Meeting

19th August 2014

Dr Guoxing Yu

Reader in Language Assessment

The University of Bristol

Contents

Introduction and summary of conclusions

Password[1] tests are designed and academically managed for English Language Testing Ltd (ELT) by the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire, UK. This report presents the results from the independent CEFR linking panel event held on 19 August 2014. The CEFR Panel considered two Password test modules – Listening and Speaking. This document details the process and outcomes with regard to the relationships between the two Password modules and the CEFR.

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 is of the relationships is given in the Tables below.

CEFR Level	Password Listening Score
A0, A1	Pre-Password
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 Speaking Score
A0, A1	Pre-Password
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.

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 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 2009, p.4) rather than embedded in the routine operation of an assessment. The project described here broadly follows the steps advocated in

the Manual; and it is ELT's intention to conduct similar events in the future when more data are available.

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) – judgments about which aspects of the CEFR are addressed by Password – and a 'vertical' comparison with relevant level descriptions – judgments 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)
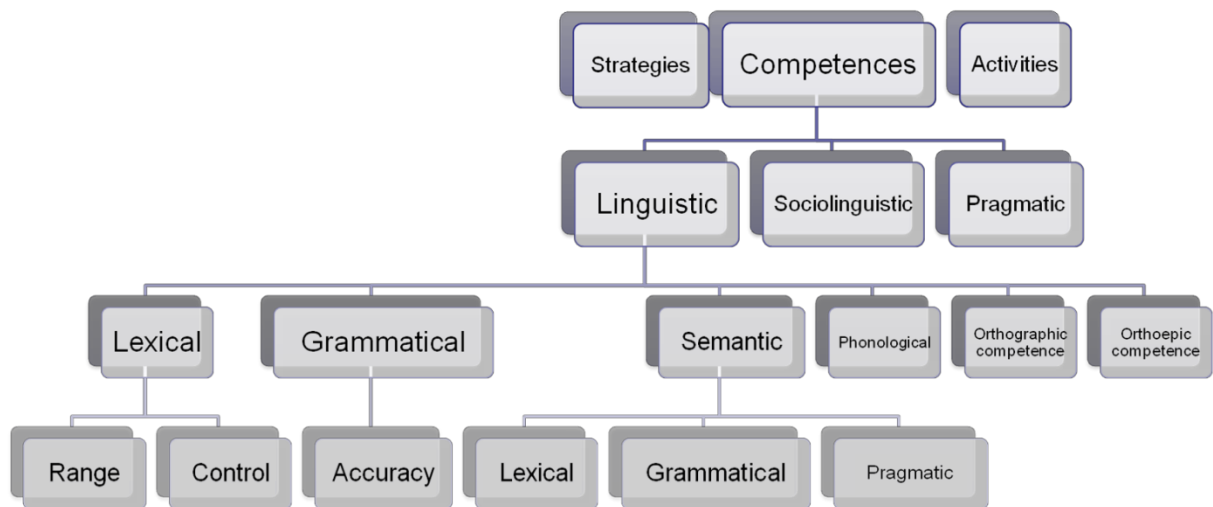
Password Listening

Password Listening was developed on the basis of Weir's (2005) socio-cognitive framework for test validation and Field's work on establishing the cognitive validity of IELTS (2012) and tests in the Cambridge English suite (2013). It aims to measure students' listening proficiency in terms of different levels of cognitive processing:

- At level A2: extraction of factual information centring on word-level cues.
- At level B1: word- and phrase-level processing; extraction of factual information.
- At level B2: contextualisation of information; determining speaker purpose and attitude; employing inference; identification of main points in an argument.
- At level C1: integrating information into a wider discourse representation; following a line of argument.

Detailed test specifications have been developed drawing upon the CEFR levels. They conform to the CEFRs outline descriptors but expand upon them to reflect the range of listening processes that students need to master in order to function effectively in an English-speaking academic context. The 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).
- Research by Field (2012) into the cognitive validity of the IELTS test. A position paper by Field (2011) on the components of expert L2 academic listening.
- Cognitive models proposed by Field for the speaking (2012) and listening (2013) constructs; criteria which emerged in the course of validating current tests against these models.

The recorded input employed in Password Listening is closely related to the subject-matter, style of delivery and discourse structure of academic presentations and lectures. Only the first section tests candidates' ability to handle everyday conversational and informational discourse across four levels (A2 to C1). The recorded material includes short clips followed by a single question (testing ability to extract local information) as well as more conventional longer passages tested by several questions. There is a final section in which candidates are required to compare the points made by two different speakers.

In relation to the vertical relationship with the CEFR levels, Password Listening test specifications cover a range from A2 to C1, with a focus on B2 and C1 levels as the targets for demonstrating ability to study in an English-speaking environment.

Table 1: CEFR Overall listening interaction

	Overall Listening Interaction
C2	*Has no difficulty in understanding any kind of spoken language, whether live or broadcast, delivered at fast native speed.*
C1	*Can understand enough to follow extended speech on abstract and complex topics beyond his/her own field, though he/she may need to confirm occasional details, especially if the accent is unfamiliar.*
	Can recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts.
	Can follow extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly.
	Can understand standard spoken language, live or broadcast, on both familiar and unfamiliar topics normally encountered in personal, social, academic or vocational life. Only extreme background noise, inadequate discourse structure and/or idiomatic usage influences the ability to understand.
B2	*Can understand the main ideas of propositionally and linguistically complex speech on both concrete and abstract topics delivered in a standard dialect, including technical discussions in his/her field of specialisation.*
	Can follow 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.
	Can understand straightforward factual information about common everyday or job related topics, identifying both general messages and specific details, provided speech is clearly articulated in a generally familiar accent.
B1	*Can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure etc., including short narratives.*
	Can understand enough to be able to meet needs of a concrete type provided speech is clearly and slowly articulated.
A2	*Can understand phrases and expressions related to areas of most immediate priority (e.g. very basic personal and family information, shopping, local geography, employment) provided speech is clearly and slowly articulated.*

A1	*Can follow speech which is very slow and carefully articulated, with long pauses for him/her to assimilate meaning.*

Password Speaking

Password Speaking was developed with the levels of CEFR (overall Spoken Interaction, see Table 2; and overall Spoken Production, see Table 3) in mind. The test employs speaking tasks designed to reflect the demands of the A2 to the C1 level based on the empirical work of Weir (1983), Brown, Iwashita and McNamara (2005), Green (2012), Nakatsuhara and Inoue (2013) and Nakatsuhara (2014).

Table 2: CEFR level descriptors for overall spoken interaction

	Overall Spoken Interaction
C2	*Has a good command of idiomatic expressions and colloquialisms with awareness of connotative levels of meaning. Can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices. Can backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it.*
C1	*Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language.*
	Can use the language fluently, accurately and effectively on a wide range of general, academic, vocational or leisure topics, marking clearly the relationships between ideas. Can communicate spontaneously with good grammatical control without much sign of having to restrict what he/she wants to say, adopting a level of formality appropriate to the circumstances.
B2	*Can interact with a degree of fluency and spontaneity that makes regular interaction, and sustained relationships with native speakers quite possible without imposing strain on either party. Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments.*
	Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field. Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. Can express thoughts on more abstract, cultural topics such as films, books, music etc.

B1	*Can exploit a wide range of simple language to deal with most situations likely to arise whilst travelling. Can enter unprepared into conversation of familiar topics, express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).*
	Can interact with reasonable ease in structured situations and short conversations, provided the other person helps if necessary. Can manage simple, routine exchanges without undue effort; can ask and answer questions and exchange ideas and information on familiar topics in predictable everyday situations.
A2	*Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time. Can handle very short social exchanges but is rarely able to understand enough to keep conversation going of his/her own accord.*
A1	*Can interact in a simple way but communication is totally dependent on repetition at a slower rate of speech, rephrasing and repair. Can ask and answer simple questions, initiate and respond to simple statements in areas of immediate need or on very familiar topics.*

Table 3: CEFR level descriptors for overall spoken production

	Overall spoken production
C2	*Can produce clear, smoothly flowing well-structured speech with an effective logical structure which helps the recipient to notice and remember significant points.*
C1	*Can give clear, detailed descriptions and presentations on complex subjects, integrating sub themes, developing particular points and rounding off with an appropriate conclusion.*
	Can give clear, systematically developed descriptions and presentations, with appropriate highlighting of significant points, and relevant supporting detail.
B2	*Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples.*
B1	*Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points.*
A2	*Can give a simple description or presentation of people, living or working conditions, daily routines, likes/dislikes etc. as a short series of simple phrases and sentences linked into a list.*
A1	*Can produce simple mainly isolated phrases about people and places.*

The Password Speaking test elicits monologic performance on generally accessible topics (see Table 4). The tasks reflect the kind of speaking commonly required in English for Academic Purposes programmes and on tests of academic language skills. There is also a clear correspondence between the Password Speaking tasks and the CEFR activities and illustrative scales of Information Exchange, Sustained Monologue and Addressing Audiences.

Table 4: Overview of Password Speaking test[2]

Section	Level	Timing	Domain	Task Type	Nature of spoken output elicited	Example topics
1	A2	3 mins	*Personal*	A series of 5 written questions presented onscreen to which the candidate is invited to respond.	Test takers are expected to talk about familiar topics	• School and work • Home country/town • Holidays • Hobbies • Food and culture • Weekend/seasonal activities
2	B1	4 mins	*Social*	A series of 3 written prompts presented onscreen, each one outlining a brief scenario or situation to elicit a short response.	Test takers are expected to demonstrate specific speech acts by giving an appropriate functional response for the given scenario.	The situations are social but in school/university contexts (e.g. requesting a librarian to purchase a book for the university's library)

---

[2] Minor modifications were made to the test presentation following the May 2014 pilot

3	B2	4 mins	*Academic*	A short written prompt presented onscreen with instructions to elicit a long turn.	Test takers are expected to engage in a long turn, comparing and contrasting the options offered and giving their opinions	• Life and work balance • Study methods
4	C1	4 mins	*Academic*	A graph or chart presented onscreen with instructions to elicit a short presentation.	Test takers are expected to give a short presentation, based upon the graph/chart, interpreting and evaluating the information given and offering their opinions	• Birth rate and social changes • Obesity, fast food and life expectancy
Total time:		15 mins				

SECTION 1 is targeted at the A2 level and lasts approximately 3 minutes. It focuses on the *Personal* domain and requires the candidate to respond to a series of simple questions on familiar topics.

Individual questions appear on screen allowing approx. 10 seconds' preparation time, before the recording begins automatically and the candidate is expected to speak following a screen prompt. The number of seconds allowed for their spoken response is given (20 seconds) before moving on to the next question.

Section 1

Section 1 has 5 questions. The questions are about you.

You have 10 seconds to read each question and think about your answer.
A bell will ring. Start speaking. You will have 20 seconds to speak.
The bell will ring again. Stop speaking. You will see the next question.

Click on 'Continue' to start Section 1.

For example, "Where do you live?", "Describe your home town.", "What do you enjoy doing in your free time?", "What did you do last weekend?", "Where will you go for your next holiday?".

SECTION 2 is targeted at the B1 level and lasts approximately 4 minutes. It focuses on the *Social* domain and requires candidates to demonstrate specific speech acts. Section 2 presents 3 short scenarios or situations that require the candidate to give an appropriate spoken response. For each scenario, the candidate is given 3 simple bullet points to help structure the content of their response. The candidate is given 20 seconds to read the task prompt. Then the candidate receives a screen prompt to start speaking and has 45 seconds to respond before moving on to the next situation.

SECTION 2

Section 2 has 3 questions.
The questions are about what you say to people you meet.
For each question you have 20 seconds to read the situation and plan your answer.
A bell will ring. Start speaking. You will have 45 seconds to speak.
It is important say as much as you can.
> After 45 seconds, the bell will ring again. Stop speaking. You will see the next question.

Click on 'Continue' to start Section 2.

For example,

Question 1: Your English homework is late. It was due on Monday. Today it is Tuesday. You have not finished the homework. You see your English teacher. He asks 'Can I have your homework, please?'. Talk to your teacher: apologise, explain the problem, say what you will do.

Question 2: You need a book for college, but you cannot find the book in the library. You see the college librarian. Talk to the librarian and ask for help: explain the problem, ask for help, thank the librarian.

Question 3: A new student has joined your class. It's time for lunch. The new student does not know where the canteen/dining room is. Talk to the student and help him/her: introduce yourself, say how you can help, invite him/her to go with you.

Section 3 is targeted at the B2 level and lasts approximately 4 minutes. It focuses on the *Academic* domain. Candidates look at a written prompt presenting 2 options and give their opinions on a topic by comparing and contrasting the options. They have 30 seconds' preparation time followed by an on-screen prompt to start speaking. The candidate must speak for about 1 minute.

Section 3

Section 3 has only 1 question. You must give your opinions about a topic.
You will see two choices. You must compare the two choices: which is better?

You have 30 seconds to read the question and plan your answer.
A bell will ring. Start speaking. You will have 1 minute to speak.
It is important say as much as you can.
After 1 minute, the bell will ring again. Stop speaking. You will see the next section.

Click on 'Continue' to start Section 3.

For example, Studying alone OR studying with friends. Which of these two study methods do you think is more effective, and why?

Section 4 is targeted at the C1 level and lasts approximately 4 minutes. It focuses on the *Academic* domain. Candidates look at a graph or chart and give a short presentation on a topic for about two minutes, interpreting and evaluating the information given and giving their opinions. They have 1 minute preparation time and must speak for about 2 minutes.

SECTION 4

Section 4 has only 1 question. You will see a graph or diagram.

- o What is the situation shown in the graph/diagram?.
- o Why this is happening?
- o What do you think will happen in the future?

You have 1 minute to read the question and plan your answer.

A bell will ring. Start speaking. You will have 2 minutes to speak.
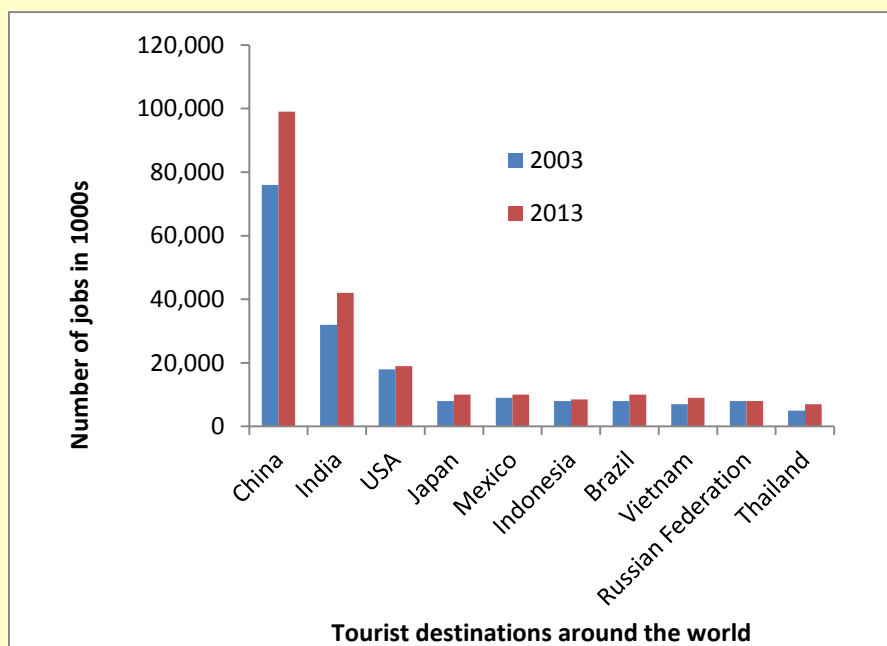
It is important say as much as you can.

After 2 minutes, the bell will ring again. Stop speaking. This is the end of the test.

Click on 'Continue' to start Section 4.

For example:

EMPLOYMENT LINKED TO TOURISM

Password Speaking responses are made available to the receiving institution which may score them using analytic rating scales, guidelines and standardised sample performances supplied by ELT.

Table 5: Password Speaking rating scale

Lower	PASSWORD SPEAKING SCALE								Higher
Password	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0 or higher
CEFR	A2		B1			B2			C1
Task summary	Describe in simple terms family and other people, living conditions, educational background, present/recent job.		Function satisfactorily in a variety of social situations likely to arise in an academic study context.			Explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.			Present clear, detailed descriptions of complex subjects integrating sub-themes and developing particular points in an appropriate way.
Task Attainment	Provides appropriate personal information and talks about familiar topics in an effective way.		Responds appropriately and effectively to the demands of specific social situations			Offers a clear personal viewpoint on a topic, including appropriate discussion of points for and against.			Presents a clear and accurate summary of graph/diagram content, together with some interpretation of the data.

Range	Uses basic vocabulary and structures, often memorised phrases, groups of a few words and formulae, in order to communicate limited (concrete, personal) information.	Has sufficient vocabulary to express him/herself with some hesitation and circumlocutions on familiar topics (such as family, hobbies and interests, work, travel, and current events).	Has a sufficient range of language to give a clear description and express a viewpoint using a mix of simple and complex language.	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style.
Accuracy	Uses some simple sentences correctly, but still systematically makes basic mistakes.	Uses reasonably accurately a repertoire of frequently used "routines" and patterns associated with more predictable situations.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstanding.	Consistently maintains a high degree of grammatical accuracy: errors are rare and cause no problem.
Fluency	Makes him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident.	Keeps going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident.	Produces stretches of language with a fairly even tempo; although may be hesitant while searching for specific words, there are few noticeably long pauses.	Expresses him/herself fluently and spontaneously without noticeable effort.

Organisation	Links groups of words with simple connectors like "and", "but" and "because".	Links series of shorter, discrete simple elements into a connected, linear sequence of points.	Uses 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.	Produces clear, smoothly flowing, well-structured speech with controlled use of organisational patterns, connectors and cohesive devices.
Pronunciation	Generally clear enough to be understood.	Clearly intelligible even if mispronunciations do occur.	Has acquired a clear, natural pronunciation.	Uses intonation and sentence stress flexibly to communicate meaning effectively.

The CEFR linking panel

As recommended in the Manual, a group of panellists was assembled for a CEFR linking event on 19th August 2014. The following sections describe the panellists and the procedures followed.

Participants

The panel was made up of 11 experienced and well-qualified English language educators[3]. It consisted of three groups: language testing researchers/lecturers working in UK higher education institutions, language test users, and a Password (language) Knowledge and Reading test question item writer. Five of the panellists hold doctoral degrees in language testing, four have Master's degree in English language teaching or related subjects, one is completing his Master's degree, and one has both Cambridge ESOL Certificate in English Language Teaching for Adults (CELTA) and Cambridge ESOL Diploma in English Language Teaching for Adults (DELTA) qualifications. The two panellists who did not hold a Master's degree both held CELTA and DELTA qualifications. All panellists had over eleven years' experience in English language teaching; seven claimed 20 or more years in the field, three between 16-20, and one between 11-15. Ten of the eleven had previous experience of CEFR linking panels. In addition to previous experience in CEFR linking panels, the members also reported a range of other experience in using CEFR, for instance, in writing textbooks, designing tests, running workshops, student assessment, and research. One used the CEFR on daily basis in work. Reporting on a five point scale (*Zero/ Poor/ Basic/ Good/ Thorough* knowledge of the CEFR), one claimed a "Thorough", eight a 'Good' and the remaining two a 'Basic' knowledge of the CEFR. Ten of the panellists were independent of English Language Testing Ltd – the owner of Password; the remaining one had experience as a question item writer for the Password (language) Knowledge and Reading tests, but not for Listening or Speaking.

---

[3] 12 panellist members were expected, unfortunately one couldn't attend due to illness.

Figure 2 Panellists' responses to the question, 'How many years have you worked in English language education?'
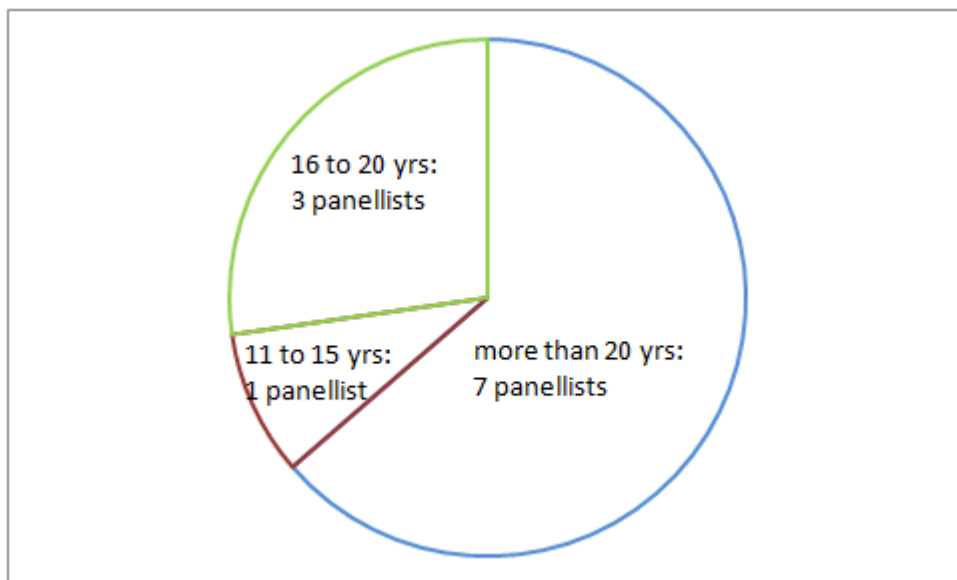
Familiarisation training

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 *CEFTrain* 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-seven 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 or four. Each group had representative(s) of test user, item writer, and language testing researcher. First, working individually, each panellist listened to five listening comprehension tasks from the CD-ROM entitled *Relating language examinations to the Common European Framework of References for Languages: Learning, Teaching, Assessment (Reading and Listening Items and Tasks: Pilot Samples)*. Each of the five listening comprehension items of Cambridge English tests: Key English Test, Preliminary English Test, First Certificate in English, Certificate in Advanced English, and Certificate of Proficiency in English were played twice. The test questions were displayed on whiteboard at the same time, but the information on which test the listening comprehension task came from was not disclosed. The panellists assigned a CEFR level to each of listening task independently according to the CEFR level descriptors for listening comprehension; a hard copy of the CEFR level descriptors (overall listening comprehension, overall spoken interaction and production) was made available for each group. The panellists then discussed their decisions within their group, justifying or reconsidering their choices. This was followed by a plenary discussion.

Linking Password Listening to CEFR

Next the panellists listened to a number of Password Listening comprehension tasks, and assigned independently a CEFR level to each Password Listening comprehension task and decided the percentage (or number) of items that a test taker at a certain CEFR level can answer them correctly. The panellists then discussed their decisions within their group, justifying or reconsidering their choices. This was followed by a plenary discussion. The panellists recorded their decisions and discussions in the form (Appendix 1) recommended by the Manual.

Linking Password Speaking to CEFR

Candidates' responses receive an overall Password Speaking level, which is decided after considering six rating criteria: Task *achievement, Range, Accuracy, Fluency, Organisation, and Pronunciation* (see Table 5). The rating criteria are closely based on the CEFR level descriptors. Scores are awarded in relation to scales that range from 'Pre-Password' to 'Password 7.0 or higher'. There is the option to award half bands (i.e. 3.5 or 5.5 etc.) to candidates that satisfy the requirements of a higher level on some of the scoring criteria.

The Speaking rating scales were presented on a whiteboard, and the panellists were asked to compare these with the illustrative scales of the CEFR and suggest which CEFR levels best corresponded to each overall band level on the Password Speaking rating scales.

To link the Password Speaking tasks to CEFR levels the panellists were presented with the Password Speaking tasks, and were asked to assign CEFR levels to the tasks. They discussed and justified their decisions in the group and to the panel. The panellists then listened to three illustrative samples of learners' performance responding to the Speaking tasks.

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. After the panel meeting, the panellists responded to an online feedback questionnaire based on a template provided by the Council of Europe (2009).

Analysis of test takers' performance in Password Listening and Speaking

At the Panel meeting, there was insufficient time to discuss test takers' performance in Password Listening and Speaking tests. After the meeting, the Password academic management team at CRELLA shared with the Panel their findings of the preliminary analysis of the data from the pilot studies.

Results

Familiarisation

The online familiarisation survey data reflected the panellists' awareness of the CEFR. On average, each statement was assigned a CEFR level correctly by 70% of the panellists. One statement (descriptor 1 in Table 6) was correctly placed by all eleven panellists; eight statements by ten of the eleven panellists.

Table 6 Distribution of panellists' judgments of CEFR levels in the online familiarisation survey prior to the panel meeting

	Descriptor (with CEFR level in brackets)	A1	A2	B1	B2	C1	C2	%
1	Has no difficulty in understanding any kind of spoken language, whether live or broadcast, delivered at fast native speed (C2)						11	100
2	Can understand enough to follow extended speech on abstract and complex topics beyond his/her own field, though he/she may need to confirm occasional details, especially if the accent is unfamiliar. (C1)				1	10		91
3	Can recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts. (C1)		1			8	2	73
4	Can follow extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. (C1)					9	2	82
5	Can understand standard spoken language, live or broadcast, on both familiar and unfamiliar topics normally encountered in personal, social, academic or vocational life. (B2)				8	2	1	73
6	Only extreme background noise, inadequate discourse structure and/or idiomatic usage influence the ability to understand. (B2)				5	4	2	45

#	Descriptor							%
7	Can understand the main ideas of propositionally and linguistically complex speech on both concrete and abstract topics delivered in a standard dialect, including technical discussions in his/her field of specialisation. (B2)				7	4		64
8	Can follow 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. (B2)				10	1		91
9	Can understand straightforward factual information about common everyday or job related topics, identifying both general messages and specific details, provided speech is clearly articulated in a generally familiar accent. (B1)		1	8	2			73
10	Can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure etc., including short narratives. (B1)		2	8	1			73
11	Can understand enough to be able to meet needs of a concrete type provided speech is clearly and slowly articulated. (A2)		9	2				82
12	Can understand phrases and expressions related to areas of most immediate priority (e.g. very basic personal and family information, shopping, local geography, employment) provided speech is clearly and slowly articulated. (A2)	5	5	1				45
13	Can follow speech that is very slow and carefully articulated, with long pauses for him/her to assimilate meaning. (A1)	10	1					91
14	Has a good command of idiomatic expressions and colloquialisms with awareness of connotative levels of meaning. (C2)				2	2	7	64
15	Can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices. (C2)				2	6	3	27

#	Statement						
16	Can backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it. (C2)				3	8	73
17	Can express him/herself fluently and spontaneously, almost effortlessly. (C1)				9	2	82
18	Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. (C1)			4	7		64
19	There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language. (C1)			1	10		91
20	Can use the language fluently, accurately and effectively on a wide range of general, academic, vocational or leisure topics, marking clearly the relationships between ideas. (B2)			7	4		64
21	Can communicate spontaneously with good grammatical control without much sign of having to restrict what he/she wants to say, adopting a level of formality appropriate to the circumstances. (B2)		1	6	3	1	55
22	Can interact with a degree of fluency and spontaneity that makes regular interaction, and sustained relationships with native speakers quite possible without imposing strain on either party. (B2)		1	7	1	2	64
23	Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments. (B2)			10	1		91
24	Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field. (B1)		6	5			55
25	Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. (B1)	3	5	2	1		45
26	Can express thoughts on more abstract, cultural topics such as films, books, music etc. (B1)		4	4	3		36

27	Can exploit a wide range of simple language to deal with most situations likely to arise whilst travelling. (B1)		2	9				82
28	Can enter unprepared into conversation of familiar topics, express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events). (B1)			8	2	1		73
29	Can interact with reasonable ease in structured situations and short conversations, provided the other person helps if necessary. (A2)		8	3				73
30	Can manage simple, routine exchanges without undue effort; can ask and answer questions and exchange ideas and information on familiar topics in predictable everyday situations. (A2)	1	7	3				64
31	Can produce clear, smoothly flowing well-structured speech with an effective logical structure which helps the recipient to notice and remember significant points.(C2)				1	5	5	45
32	Can give clear, detailed descriptions and presentations on complex subjects, integrating sub themes, developing particular points and rounding off with an appropriate conclusion. (C1)				1	7	3	64
33	Can give clear, systematically developed descriptions and presentations, with appropriate highlighting of significant points, and relevant supporting detail. (B2)				8	3		73
34	Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples. (B2)				7	4		64
35	Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points. (B1)			10	1			91

| 36 | Can give a simple description or presentation of people, living or working conditions, daily routines, likes/dislikes etc. as a short series of simple phrases and sentences linked into a list. (A2) | | 10 | 1 | | | | 91 |
| 37 | Can produce simple mainly isolated phrases about people and places. (A1) | 10 | 1 | | | | 91 |

Table 7 Percent of panellists correctly assigning a CEFR level for each statement

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	27.00	1	2.7	2.7	2.7
	36.00	1	2.7	2.7	5.4
	45.00	4	10.8	10.8	16.2
	55.00	2	5.4	5.4	21.6
	64.00	8	21.6	21.6	43.2
	73.00	8	21.6	21.6	64.9
	82.00	4	10.8	10.8	75.7
	91.00	8	21.6	21.6	97.3
	100.00	1	2.7	2.7	100.0
	Total	37	100.0	100.0	

The most problematic descriptor proved to be number 15 (Can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices), less than 1/3rd of the panellists assigned the correct CEFR level. Overall, however, this preliminary familiarisation exercise confirmed the panellists' self-assessment 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.

There were further familiarisation activities during the panel event on 19th August 2014, using CEFR listening samples. The panel agreed that the Cambridge Key English Test (KET) listening sample was at A2(+) level (or possibly at B1), the PET listening sample at B1 (one group of the panel thought it was at A2), the FCE listening sample at B2, the CAE at C1 (some thought it was at B2), and the CPE listening sample at C2 (some thought it might be C1). The panel's judgments on the listening samples of the Cambridge English tests were in line with the alignment made by Cambridge English (www.cambridgeenglish.org/cefr).

CEFR linking result - Password Listening

The panellists assigned independently a CEFR level to each Password listening task, and then decided within their group the CEFR level and the percentage of items in each section that a test taker at a certain CEFR level can answer them correctly. It should be noted, however, that given the nature of the group discussions it was not always the case that an agreement was reached within the group within the time allowed. It was at the plenary discussion time that we aimed to reach an agreement.

Table 8: Listening tasks in Section 1

	Group 1	Group 2	Group 3
1 (clock/time)	A2	A2	A2
2 (Mandela)	A2	A2	A2
3 (Amazon River)	A2	A2	A2
4 (forest fire)	B1	B1	B1
5 (doctor)	B1	B1	B1
6 (accidents)	B1	B1	B1

A1	30%	30%	30%
A2	60%	40%	40%
B1	80%	60%	70%
B2	90%	80%	90%
C1	100%	100%	

As shown in Table 8, the Section 1 tasks required test takers to have A2 to B1level listening proficiency to be able to answer the majority of the items correctly.

Table 9: Listening tasks in Section 2

	Group 1	Group 2	Group 3
1 (air pollution)	B1	B1+	B1

As shown in Table 9, the panel agreed well that the tasks required test takers to have B1 or B1+ level listening proficiency.

Table 10: Listening tasks in Section 3

	Group 1	Group 2	Group 3
1 (rain forests)	B1	B1+	B2
2 (Artic/Antarctic)	B1+	B2	B2
3 (Studying)	B1+	B2	B2+
4 (theory of mind)	B2	B2	B2+
A2	20%	30%	--
B1	60%	50%	--

| B2 | 100% | 75% | 65% |
| C1 | 100% | 90% | 100% |

As shown in Table 10, the panel agreed that Section 3 tasks were at B1+ or B2 level listening proficiency.

Table 11: Listening tasks in Section 4

(animal communication)	Group 1	Group 2	Group 3
Question 1	C1	B2+	B2+
Question 2	C1	B2+	B2+
Question 3	C1	B2+	B2+
Question 4	C1	B2+	B2+
Question 5	C1	C1	B2+/C1
A2		25%	--
B1		35%	--
B2		50%	60%
C1		75%	100%

As shown in Table 11, the panel agreed well that the task was at B2+ or C1 level.

Table 12: Listening tasks in Section 5

(modern tourism)	Group 1	Group 2	Group 3
1	B2	B1+	B2+
2	B2	B1+	B2+
3	B2	B1+	B2+
4	B2	B1+	B2+
5	B2	B1+	B2+
6	B2	B1+	B2+
A2		35%	
B1		65%	
B2		80%	
C1		100%	

As shown in Table 12, the panel suggested that it was at B1+, B2 or B2+ level.

In summary, as shown in Tables 8-12, the panel generally agreed well among themselves, as well as with the test designers (according to the test specifications), on the level of listening ability required to successfully complete the tasks. Section 1 tasks were considered at A2 – B1, Section 2 tasks at B1, Section 3 tasks at B1+ or B2, Section 4 at B2+ or C1, and Section 5 at a relatively broader range from B1+, B2, to B2+.

Listening test performance data was provided to the panel after the meeting. As shown in the data, most of the Listening test items are working well; the progression of Listening task difficulty was as intended, with Section 1 as the easiest (A2) and other sections becoming progressively more difficult, ranging from B1 to C1. There was an effective spread of test scores, which indicates that the Listening test seems appropriate to the level of pilot students.

In summary, according to the panel's judgment on the day and the pilot test performance data, Password Listening tasks ranged from A2 to C1 CEFR levels. Specifically, the Section 1 tasks were considered at A2 – B1, Section 2 tasks at B1, Section 3 tasks at B1+ or B2, Section 4 at B2+ or C1,

and Section 5 at B2 level. The test takers performance data in the pilot studies further confirmed the panel's judgment on the increasing difficulty of the listening tasks.

Each panellist, working individually, judged the difficulty of each Password Listening task. They consulted the CEFR scales for 'Overall Listening Comprehension' and 'Listening As A Member Of A Live Audience' 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 Listening score.

Table 13 Average rater judgements of CEFR level thresholds on the Password Listening test

Section	Section 1	Section 2	Section 3	Section 4	Section 5	
Score out of	*10 points*	*8 points*	*6 points*	*7 points*	*8 points*	*Total points*
A1	0	0	0	0	0	0
A2	5	3	1-2	1-2	2-3	12-15
B1	7	5	3	3	4-5	22-23
B2	8	6-7	4-5	4-5	6-7	28-33
C1	9-10	8	6	6-7	8	37-39

Taking into consideration the Panel's judgement and the actual test performance data from the pilot studies, the recommended cut scores for the Password Listening test are given in Table 14.

Table 14 Password Listening score linking

	Password Listening
A2	12
B1	19
B2	26
C1	35

Based on the outcomes of this exercise, the recommendations given in Table 15 can be made on the relationship between the CEFR and the Password Listening scales.

Table15 Password Listening scales: CEFR link

A0, A1	Pre-Password
A2	3.0
B1	4.0
B1+	5.0
B2	5.5
C1	Password 7 or higher

CEFR linking result - Password Speaking

The panel listened to three test takers' (TT, SB, CI) recorded speaking and judged their performance according to the CEFR descriptors on spoken production. As shown in Table 16, there was an even stronger agreement between the three groups in their judgment about the test takers' speaking proficiency levels. The grades awarded to the three test takers (4.0, 5.5, and 7.0 or higher) by Password raters according to the scoring scale (see Table 5) were in line with the panel's overall judgment of the test performance according to the CEFR levels. The only discrepancy between the panellists judgement and the Password scores was for test taker TT. This test taker scored Password 4.0 on the test, but was judged by the majority of panellists to be only A1+ or A2 on the CEFR, which would be consistent with a Password score of 3.0. It is therefore recommended that this performance should not be used in rater training to exemplify a 4.0 performance, but might be reclassified as 3.0.

Table 16: Password Speaking

	Group 1	Group 2	Group 3	Password score awarded
Test taker: TT	A1+	A2	A1+	4.0
Test taker: SB	B1+	B2	B2	5.5
Test taker: CI	C1	C1+/C2	C1/C2	7.0 or higher

According to the panel's judgment of the three test takers' performance in the four sections of the Password Speaking test, Password Speaking tasks provide a sufficient range of opportunities for test takers to demonstrate their different level of speaking proficiency.

Based on the outcomes of this exercise, the recommendations given in Table 17 can be made on the relationship between the CEFR and Password Speaking.

Table 17 Password Speaking: CEFR link

A0, A1	Pre-Password
A2	3.0
B1	4.0
B2	5.5
C1	Password 7 or higher

Comments on the panel event

In a final session, panellists were given an opportunity to comment on the process and outcome of the CEFR alignment event. The panellists also filled in a brief online questionnaire, based on an example provided in the Council of Europe Manual. A summary of their responses to the online questionnaire is presented in Table 18.

Table 18: Feedback questionnaire responses (N=10)

Questions (4 strongly agree, 3 agree, 2 disagree, 1 strongly disagree)	Mini.	Max.	mean	Std. deviation
I understood how to answer the pre-panel online questionnaire	4	4	4.00	0.000
The online familiarisation activity was helpful for me	3	4	3.70	.483
I understood the purpose of the panel meeting	3	4	3.90	.316
I understood the instructions for the activities at the panel meeting	3	4	3.50	.527
The introductory familiarisation activity at the panel meeting was helpful for me	3	4	3.40	.516
There was adequate time for discussions at the panel meeting	3	4	3.60	.516
All participants were allowed to state their views freely	3	4	3.90	.316
The small group discussions were helpful for me	3	4	3.80	.422
The panel discussions were helpful for me	3	4	3.80	.422
The facilities at University of Arts London helped to create a positive working environment	2	4	3.20	.632
I am confident that we reached a reasonable decision on the relationship between CEFR levels and Password Test (Listening comprehension)	2	4	3.30	.675

I am confident that we reached a reasonable decision on the relationship between CEFR levels and Password Test (Speaking)	2	4	3.40	.699
I feel I now have a clear understanding of the CEFR levels	3	4	3.60	.516

It is clear from their feedback that the panellists were generally satisfied with the event and with their decisions on the relationship between Password and the CEFR. 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), 301-320

Brown, A., Iwashita, N., & McNamara, T. (2005). An examination of rater orientations and test taker performance on English for academic purposes speaking tasks. TOEFL Monograph Series. Princeton, NJ: Educational Testing.

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment, 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

Field, J. (2011) Into the mind of the academic listener. *Journal of English for Academic Purposes*, 10(2), 102–112

Field, J. (2012) 'The cognitive validity of the lecture based question in the IELTS listening paper'. In C. Weir et al. (eds.) *IELTS Collected Papers 2: Research in reading and listening assessment* Cambridge: Cambridge University Press

Field, J. (2013) Cognitive validity. In A. Geranpayeh & L. Taylor (eds.), *Examining Listening.* Cambridge: Cambridge University Press

Green, A.B. (2012) *Language Functions Revisited: Theoretical and empirical bases for language construct definition across the ability range*. Cambridge: Cambridge University Press.

Milanovic, M. 2009. Cambridge ESOL and the CEFR. Cambridge ESOL *Research Notes*. Issue 37.

Nakatsuhara, F. (2014) *A research report on the development of the Test of English for Academic Purposes (TEAP) speaking paper for Japanese University entrants*. EIKEN Foundation of Japan (formerly STEP), Japan.

Nakatsuhara, F. & Inoue, C. (2013) *Trinity College London Integrated Skills in English (ISE) 'The Interview': Function analysis report for ISE 0, 1 & 3*. Project report submitted to Trinity College London.

Weir, C. J. (1983). *Identifying the language problems of the overseas students in tertiary education in the United Kingdom*. Unpublished PhD dissertation: University of London.

Appendix 1: Blank form for recording listening comprehension alignment

Listening Comprehension					
Target level in the curriculum:					
Item types					
Source					
Length (mins)					
Authenticity					
Discourse type					
Domain					
Topic					
Curriculum linkage (an optional new category)					
Number of speakers					
Pronunciation					
Content					
Grammar					
Vocabulary					
Nr of listening					

Input text comprehensible at level					
Items comprehensible at level (enter item codes)					
A1					
A1/A2					
A2					
A2/B1					
B1					
B1/B2					
B2					
B2/C1					
C1					
C1/C2					
C2					