

Relating LanguageCert Communicator to the CEFR

Anthony Green, University of Bedfordshire

Centre for Research in English Language Learning and
Assessment, Putteridge Bury, Hitchin Road, Luton, LU2 8LE, UK

Language
Cert



14 January 2019

Abstract

This study was undertaken to relate the *LanguageCert Communicator* Exam to the Common European Framework of Reference (CEFR; Council of Europe 2001). It includes both the Spoken (Speaking) and Written (Listening, Reading, Writing) Exams for which separate certificates are awarded. The study employed the staged approach recommended by the Council of Europe (2009) which includes Familiarisation, Specification, Standardisation, Benchmarking/ Standard setting and Validation.

Following Familiarisation, which involves building and confirming understanding of the CEFR, Specification was carried out by *LanguageCert* staff in collaboration with the researchers. This made use of a standard text template developed from the forms used in the Council of Europe (2009) Manual, but designed to better convey the outcomes to test users and other stakeholders.

Benchmarking and Standard setting combined a qualitative perspective based on the analysis of test materials and rating scales with the 'Benchmarking with FACETS' approach suggested by North and Jones (2009) which makes use of calibrated performance samples and cut scores for the CEFR level descriptors. The twin-panel approach involved two-day meetings in Greece and the UK between a total of 16 expert panellists (nine meeting in Athens and seven in Luton). The panellists reviewed test material and sample performances and related these to the CEFR. The review of material confirmed that all four papers (Listening, Reading, Writing and Speaking) reflected the B2 level of the CEFR in the targeted Communicative Activities.

Findings from the Benchmarking of performance samples and Standard setting panels broadly supported the current interpretation that passing scores on the four *Communicator* subtests (Listening, Reading, Writing and Speaking) represent B2 on the CEFR in the areas tested, but results from both panels suggested that the current passing scores for B2 should be raised across all four papers.

This report begins with a brief description of the *Communicator* Exams and of the Common European Framework of Reference (CEFR). It outlines the work undertaken to relate the City & Guilds qualifications to the CEFR and how this contributed to the development of the Council of Europe's *Manual for Relating Language Examinations to the CEFR* (Council of Europe 2009). It then sets out the methods employed in this study and presents the results of the four phases recommended by the Council of Europe (2009): Familiarisation, Specification, Benchmarking/Standard setting and Validation. It concludes with recommendations on cut scores to be adopted for *LanguageCert Communicator* to represent a minimal B2 level of performance and for further work to maintain the relationship between the Exams and the CEFR.

LanguageCert International ESOL Communicator

LanguageCert International ESOL (English for Speakers of Other Languages) is a set of English Language qualifications, each targeting a different level of the CEFR and intended for teenage, young adult and adult candidates who may be preparing for entry to higher education or professional employment. They include the *Preliminary* (A1), *Access* (A2), *Achiever* (B1), *Communicator* (B2), *Expert* (C1) and *Mastery* (C2) levels. These are based on the City & Guilds examinations of the same name, which were acquired by *LanguageCert* in 2015.

The focus of this report is on the *Communicator* (B2) tests, which consist of a Written Exam (Listening, Reading, Writing) available in both computer-based and paper-based formats and a face-to-face Spoken Exam (Speaking). Certificates are awarded separately for the Written and Spoken Exams and results fall into three categories: Fail, Pass (75 out of 150 on the Written Exam; 25 out of 50 on the Spoken Exam) or High Pass (101 out of 150 on the Written Exam; 35 out of 50 on the Spoken Exam).

Following the acquisition of the International ESOL qualifications, all the test material at all levels (a total of approximately 8,500 items) was subjected to a comprehensive quality review procedure. A team of consultants, including a former Director of Assessment at Cambridge Assessment English and two experienced test developers, was commissioned to carry out an initial quantitative review of the materials. Among other features, they considered the statistical data on test and item performance (difficulty and discrimination), the material's suitability for the target test population, its topicality (might the content be, or soon become out-of-date?) and the accuracy of the answer keys. Most of the test material from the City & Guilds International ESOL Reading and Listening tests had to be modified as a decision was taken to adopt a three-option multiple-choice format rather than the four-option format that had predominated. Based on the reviews, items were either accepted for use on the revised examinations, rejected or edited to improve their quality. Of the 8,500 test items analysed, approximately 45% were accepted with no change other than, where relevant, eliminating the fourth option. A further 37% were more substantially edited and 18% rejected.

As new material is generated, following specifications that embed the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001), this is added to the

legacy material from City & Guilds to form the *LanguageCert* item bank. As more data becomes available through piloting and official test administrations, this is calibrated to the *LanguageCert* Item Difficulty (LID) scale using Rasch analysis techniques. This scale is used to inform the compilation of parallel forms of the examinations at each level, helping to maintain consistency of standards over time.

Although the overall design of the City & Guilds tests has been retained, the tests have been substantially revised. The changes, reflecting the current test design are summarized in Table 1 below.

Table 1 Changes to the format of the Communicator tests

Time	Test	City & Guilds Communicator			LanguageCert Communicator	
		Part/ task	No. of items/ extent	Item type	No. of items/ extent	Item type
30 mins	Listening	1	8	4-option MCQ	7	3-option MCQ
		2	6	4-option MCQ	6	3-option MCQ
		3	8	Form completion	7	Form completion
		4	8	4-option MCQ	6	3-option MCQ
130 mins	Reading	1	6	4-option MCQ	6	3-option MCQ
		2	6	Choose from 8 sentences to complete 6 gaps (2 distractors)	6	Choose from 7 sentences to complete 6 gaps (1 distractor)
		3	9	Match questions to the relevant text	7	Match questions to the relevant text
		4	9	Open-ended 'wh' questions requiring short answers	7	Open-ended 'wh'-questions requiring short answers
c. 10 mins	Writing	1	100-150 words	Formal response to a written, graphic or visual input with four content points to be addressed and the intended reader specified	100-150 words	Formal response to a written input with four content points to be addressed and the intended reader specified
		2	150-200 words	Personal letter, narrative or descriptive composition	150-200 words	Personal letter, narrative or descriptive composition
c. 10 mins	Speaking	1		Give and spell name; Give nationality; Answer five questions		Give and spell name; Give country of origin; Answer five questions
		2		Two or three situations are presented by the interlocutor at each level and candidates are required		Two or three situations are presented by the interlocutor at each level and candidates are required

		to respond to and initiate interactions.		to respond to and initiate interactions.
3		Hold a short discussion to make a plan, arrange or decide something using written text as the prompt		Hold a short discussion to make a plan, arrange or decide something using written text as the prompt
4	30 seconds preparation, 2-3 minutes talking	After 30 seconds pf preparation time talk about a topic provided by the interlocutor and answer follow-up questions	30 seconds preparation, 2-3 minutes talking	After 30 seconds pf preparation time talk about a topic provided by the interlocutor and answer follow-up questions

The Common European Framework of Reference

The aim of the CEFR is to encourage and facilitate reflection, communication and networking in language education. It is intended to “make it easier for practitioners to tell each other and their clientèle what they wish to help learners to achieve, and how they attempt to do so’ (Council of Europe 2001a, p.4) and to ‘make their... objectives and methods clear and explicit for the benefit of those who use the products of their work” (p.5). In the case of examinations, the CEFR is intended to assist users in articulating both their content standards (the nature of the skills being tested) and their performance standards (levels of proficiency).

The CEFR encompasses a horizontal dimension and a vertical dimension. The horizontal aspect of content coverage is embodied in the *Descriptive Scheme*, a set of parameters for conveying content standards relating to language use. The vertical dimension is represented by a set of *Common Reference Levels* that divide objectives for organised language learning into levels of proficiency (See Figure 2). The Common Reference Levels provide a common basis for communicating and comparing performance standards. It is acknowledged that although the framework is intended to be comprehensive, it is not exhaustive and sub-categories, expanding on the published framework, may be needed for both the Descriptive Scheme and Common Reference Levels to better capture learner needs and abilities in specific language teaching or assessment contexts.

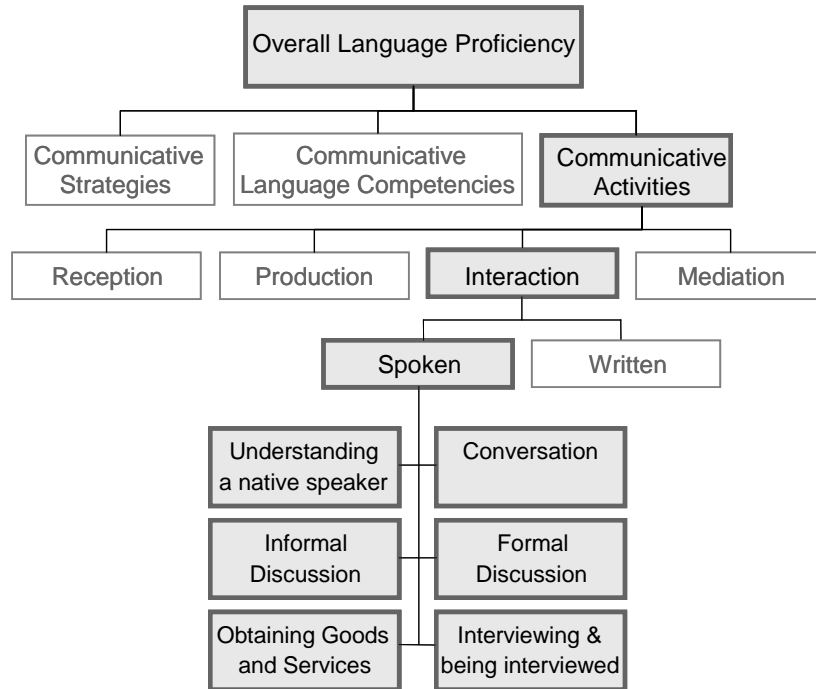


Figure 1 The location of illustrative scales for spoken interaction within the CEFR descriptive scheme (from Council of Europe 2001b)

Illustrating the flexibility of the framework, a number of illustrative scales are provided to represent a range of potential content (conceptualizing language use from the perspectives of *Communicative Activities*, *Communication Strategies* and *Communicative Language Competences*). Figure 1 shows the kinds of choice available to users by following one strand in the scheme (*Communicative Activities*) from the abstract level of overall language proficiency to illustrative scales describing levels of performance in six *Activities* that involve *Spoken Interaction*.

On the vertical dimension, following the branching principle of the CEFR (Council of Europe 2001a, p.25), the three broad reference levels (A, B and C) are each divided into two (e.g. A1 and A2), with a further distinction suggested between the ‘criterion levels’ (e.g. A2) and the ‘plus levels’ (A2+, B1+ and B2+) to give nine levels in total (see Figure 2). It is suggested that further and finer distinctions can be drawn according to local need. With the publication of the *Companion Volume with New Descriptors* (Council of Europe 2018), a further level, termed Pre-A1, has been added to reflect a pre-generative level of language proficiency that depends on a very limited repertoire of words and formulaic expressions.

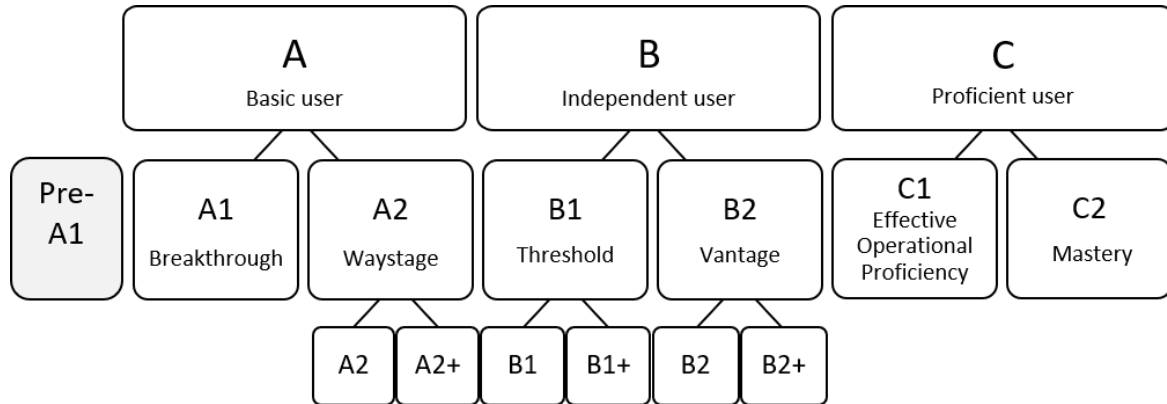


Figure 2 The branching system of levels in the CEFR (based on Council of Europe 2001, p.32, Figure 2) and Council of Europe (2018)

Following piloting by a number of European test providers including City & Guilds, a *Manual for Relating Language Examinations to the CEFR* (Council of Europe, 2009) (hereafter referred to as the Council of Europe Manual) was published by the Council of Europe to inform “linking”, or the process of relating qualifications to the framework.

Linking City & Guilds Communicator to the CEFR

Following publication of the CEFR in 2001, City & Guilds was among the first examination providers to attempt to relate their qualifications to the framework. A draft manual for *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Council of Europe 2003) (henceforward referred to as the Manual) was issued by the Council of Europe to guide such initiatives. This was eventually published in its final form, following piloting, in 2009. The City & Guilds projects, led by Professor Barry O’Sullivan from the Centre of Language Assessment Research (CLARe) at Roehampton University, were among those piloting the draft Manual and they followed its four recommended phases:

- Familiarisation: a process of building familiarity among linking project participants with the CEFR and with the assessment.
- Specification: a “self-audit of the coverage of the examination (content and tasks types) profiled in relation to the categories presented in [the CEFR]” (Council of Europe 2003, p.2).
- Standardisation: “procedures to facilitate the implementation of a common understanding of the “Common Reference Levels” of [the CEFR]” (Council of Europe 2003, p.2). These included measures to relate local test performances to the CEFR and the setting of cut scores.
- Validation: the collection of evidence to support the claims linking the assessment to the CEFR.

O’Sullivan prefaced the City & Guilds linking projects with an additional “Critical Review” phase. This involved a comprehensive evaluation and limited revision of the qualifications guided

by engagement with the CEFR. The addition of the Critical Review addressed the condition set out in the Manual that “Relating examinations to [the CEFR] only makes sense if the examinations are of good quality” and that, “It is the prior responsibility of the examination providers to demonstrate the validity of their examination by showing that it assesses the constructs intended.” (Council of Europe 2003, p.1). In effect, the Critical Review phase could be regarded as a component of an extended Specification phase, adding two new elements aimed at evaluating and enriching the examination: external scrutiny and iterative, incremental improvement. This was in harmony with the intention of the draft Manual to raise consciousness of quality issues and to provide for “a cumulative process of continuing improvement” (Council of Europe 2003, p.1).

Specification

O’Sullivan (2008) appreciated that the forms provided by the Council of Europe for Specification “forced the team to consider aspects of the tests not necessarily referred to directly in the re-written specifications” (p.20). Nonetheless, the project team found the forms, “somewhat awkward and repetitious” (p.20) and lacking in transparency for users. One advantage of the involvement of the external expertise of O’Sullivan and his colleagues was their participation in other linking projects, which allowed them to build and share in a broader consensus interpretation of the form categories. In his recommendations to the Council of Europe, O’Sullivan (2008) argued that, “the design of the manual forms should be reconsidered” and that they should be brought into line with a more current approach to test specification. However, this does not appear to have been a priority in the revision of the Manual and the final version, published in 2009, included few modifications to the forms, although it did offer additional grids for content description that would support a more detailed description of individual test tasks. In practice, the forms and grids have not always been used in linking projects and where they have been, the responses they elicit are highly variable, not providing for the ready comparability of qualifications envisaged in the Manual (Green 2018).

Standardisation

The City & Guilds projects first addressed the *Communicator B2* level examinations (O’Sullivan 2008, 2010). During the Standardisation stage, panels of expert judges were assembled and, following further Familiarisation, set standards using a variation on the Examinee Paper Selection method for the Writing and Speaking papers and a variation on the Angoff method for the Reading and Listening papers. Based on these exercises, a cut score of 50% was set for all papers as representing a minimally B2 level of proficiency.

Writing and Speaking

The reports on Writing and Speaking (offered as a separate qualification – *ISESOL Communicator*) concluded that City & Guilds’ then current grade boundaries were supported by the panels. In other words, a passing grade on the Writing and Speaking tests provided a clear indication that the test taker satisfied the B2 level in these skills. The supporting evidence for this from a review of the rating scales and from the estimates made by the panels raises a number of

issues. According to O’Sullivan (2008), “the rating scale used for *Communicator* was developed based directly on the descriptors at Level B2 of the CEFR” (p.69) with band scores of 2 representing B2 on each criterion. Satisfying the criteria for B2 across the four Writing criteria of *Task Fulfilment, Range, Organisation* and *Accuracy* would thus suggest a passing score of 8 out of 12 on each task, or a total of 16 out of 24 over the two tasks rather than the 12 points used as a cut score by City & Guilds. Scoring of the *ISESOL Communicator* Speaking test involved combining a global score (out of 3) awarded for each of the four test parts (making a total of 12 possible points) with analytic scores awarded on a three-point scale for performance on the test as a whole against the four criteria of *Accuracy, Range, Pronunciation* and *Fluency* (again making a total of 12 possible points). As with Writing, it seems that a score of 2 on each scale was based on the B2 level, suggesting a passing total score of 16 rather than the operational 12 out of the maximum 24.

In the Standardisation panels, of the two samples used to represent a passing grade on *ISESOL Communicator* for Speaking (O’Sullivan 2010), one was rated by the panellists as B2, but the other as B1 (the panels were not given the option of awarding a B1+ score). The response of the project team to this unexpected result was to reclassify the sample. This was subsequently used in rater training as an example of a failing performance. Only one sample was used in the panel meeting to represent a passing grade for the Writing test, although this was, as expected, awarded a B2 rating by the panellists. O’Sullivan (2008, 2010) noted that the Council of Europe only provided one example of a B2 performance each for Writing and Speaking to support linking studies and questioned whether this could provide an adequate basis for standard setting. It was clearly important that the current study should employ more samples of performance.

Listening and Reading

For the City & Guilds Listening and Reading papers, a variation of the Angoff standard setting procedure was adopted. Panellists reviewed one representative Reading and one representative Listening paper and estimated the probability (expressed as a percentage in ten-point increments) that a test taker who minimally satisfied the criteria for B2 (see Table 2 for the definitions of these adopted by the panels) would respond correctly to each test item. The average of these estimates provided the panel’s initial recommendation for the B2 cut score. The panellists then discussed their estimates and were permitted to revise them. An adjusted average based on a multi-facet Rasch analysis of these Round 2 estimates (excluding those of any panellists whose judgements were found to be inconsistent) was taken as the panel’s recommended B2 cut score for each test. In both cases this was a score of 15 out of 30.

Table 2 Definitions of minimally competent B2 Readers and Listeners adopted by the City & Guilds linking panels

<i>Reading</i>	<i>Listening</i>
Can understand without dependence on dictionaries or glossaries articles, reports and narratives aimed at the general reader and texts in which the writers adopt particular stances	Can follow most standard spoken language, live or broadcast, such as lectures, discussion and debates, on topics normally encountered in personal, social, academic or vocational life. Has difficulty understanding implicit meaning in extended

or viewpoints. Has difficulty with specialized or unclearly structured texts and low frequency lexis.

speech and finds it difficult to understand if there is extreme background noise, inadequate discourse structure and idiomatic usage.

Alongside estimates based on test material, the *Communicator* project also involved data from test administrations. The Council of Europe provided two test tasks illustrative of the B2 level for Reading and two for Listening (from the Cambridge *First Certificate in English - FCE* - and the Finnish *Matriculation Examination* English test). These were included in pilot *Communicator* test forms used with 58 test takers judged to be mostly at B2 level in these skills. For Listening there was no significant difference in difficulty between the *Communicator* tasks and either of the two illustrative B2 tasks. For Reading, the *Communicator* task proved to be similar in difficulty to the Finnish *Matriculation Examination* task, but significantly easier than the *FCE* task. Although this exercise suggested that the City & Guilds material was broadly consistent with B2, it could not support any more definitive conclusion. O’Sullivan (2008) remarked that the illustrative tasks were not of equivalent difficulty, but that little indication was given by the Council of Europe as to whether they represented relatively easy or difficult material within the B2 range or of the scores on the tasks that might represent a minimally competent B2 performance.

One further observation made by O’Sullivan (2008) is worthy of note. He was critical of the one-off linear approach to linking implied in the Manual. He suggested that for the relationship between the tests and the CEFR to persist over time, it should be embedded in “all aspects of the development and validation process” (p.86) including training in the CEFR for item writers and maintenance of consistent levels of difficulty in the test material. Ensuring ongoing alignment with the CEFR is important to *LanguageCert* and continues to be considered at all stages of test development and delivery. This starts with test specifications and continues all the way through to the criteria employed in the Speaking and Writing tests, the standardization of examiners and *post hoc* statistical analyses.

The City & Guilds linking projects represented an important step forward for the examinations, doing much to improve their quality and that of the systems that supported them. The recent revisions following the acquisition of the qualifications by *LanguageCert* have built on the advances made a decade ago. The revisions to the design of the examinations (Table 1) have coincided with developments in the CEFR. The publication of the Companion Volume (Council of Europe 2018) and the addition of new descriptors has extended the definition of the framework. Additional material has been made available by the Council of Europe to illustrate the common reference levels. The time is therefore ripe to revisit and re-evaluate the relationship between the examinations and the CEFR.

Research methods

In common with the earlier linking studies involving the *Communicator* tests (O’Sullivan 2008), this study was guided by the approach recommended in the Council of Europe Manual (Council of Europe 2009). Discussion and consensus building between stakeholders with a range

of perspectives is central to the process and the Manual recommends convening a diverse panel of experts to judge how the content and level of the examination in question relates to the CEFR.

Because of the logistical challenges involved in setting up a meeting of researchers based in the United Kingdom with test developers and others working in Greece, where the *LanguageCert* development team is based, a decision was taken to employ an asynchronous ‘twin-panel’ approach. This was preferred to the option of using an online alternative similar to that employed by Knoch and Frost (2016) because of the importance of discussion and consensus building, which the researchers judged to be better accomplished through direct, face-to-face interaction among panellists. Twin panel approaches have been successfully used by Green (2012) (for the Pearson Versant English Placement Test) and by Brunfaut and Harding (2014) and Green & Inoue (2017) (for the Taiwanese GEPT) to facilitate collaboration between external researchers and staff at examination boards when working in distant locations.

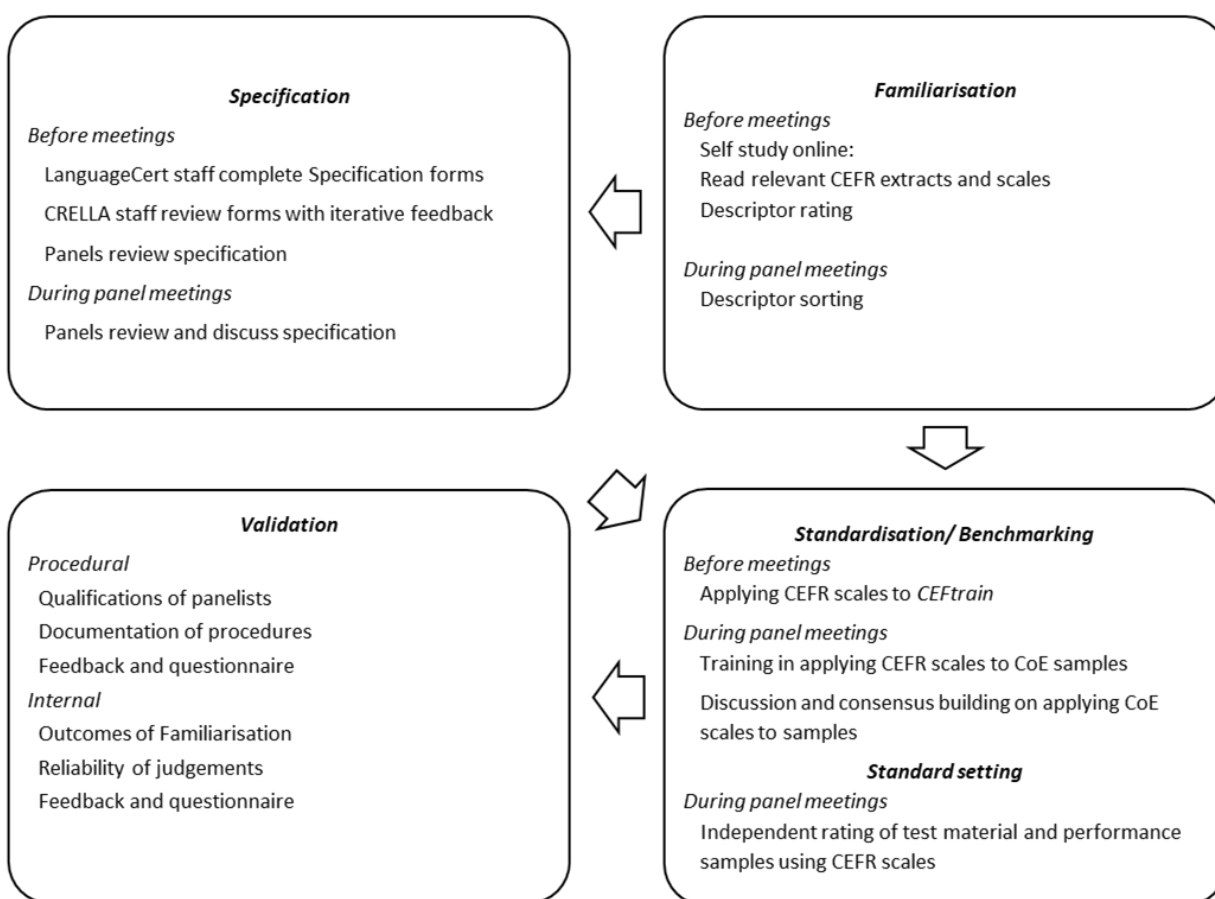


Figure 3 Summary of procedures adopted

The Council of Europe (2009) Manual divides the linking process into five conceptual stages of Familiarisation, Specification, Standardisation/Benchmarking, Standard Setting and Empirical Validation. In this project (see the summary provided in Figure 3), Familiarisation, involving building familiarity with the CEFR, was carried out both prior to and during the panel meetings.

Following the initial Familiarisation, Specification – a self-audit of the coverage of the examination content in relation to CEFR categories – was carried out by *LanguageCert* staff involved in the *Communicator* programme. The *LanguageCert* staff completed the forms for describing the test content provided by the Council of Europe Manual, and two rounds of preliminary drafts were commented on by CRELLA researchers. These were reviewed by the panellists during the panel meetings. More detailed procedures for this phase are described later. The forms are provided in the Appendix.

Since the draft Manual (Council of Europe 2003) employed in the City & Guilds projects (O’Sullivan 2008, 2010), one notable change to the recommended phases of linking projects has been the refinement of the Standardisation phase. This is now divided into three parts: *Standardisation training*, *Benchmarking* and *Standard setting* (Council of Europe 2009).

Standardisation training involves training participants in the interpretation of the levels through the use of CEFR illustrative performance samples and tasks. It may precede both Benchmarking and Standard setting. In this project, Standardisation training was carried out prior to the panel meetings using the *CEFRTrain* resource (www.helsinki.fi/project/CEFRTrain) and calibrated sample performances were used to confirm the panellists’ interpretations of the CEFR Common Reference Levels.

Benchmarking of tasks involves judging the demands of the tasks on the examination in relation to the CEFR common reference levels. Benchmarking of performance involves using CEFR scales to rate sample performances drawn from the examination in question. If the performance samples from the test are successfully identified with the intended levels, this supports the linking claim. Benchmarking of performance may be used to support claims about the correspondence between test scores and the CEFR and may be understood as an examinee work based standard setting method (Hambleton et al. 2000). This was similar to the approach for Writing and Speaking used by O’Sullivan (2008, 2010) in the City & Guilds linking panels. The current study builds on O’Sullivan (2008, 2010) by including more performance samples from *Communicator*, by using additional sample performances that have become available from the Council of Europe and English Profile and by exploiting the calibration of these on CEFR-based scales for Writing and Speaking.

Standard Setting involves “the establishment of a decision rule to allocate students to one of the CEFR levels on the basis of their performance in the examination. Usually this takes the form of deciding on cut-off scores, borderline performances.” (Council of Europe 2009, p.11). An initial round of Standardisation training was carried out prior to the panel meetings for this project (see above). Again, this study was able to build on the earlier City & Guilds studies by including additional illustrative Reading and Listening tasks since made available by the Council of Europe, although, for reasons outlined on the Council of Europe website (www.coe.int/en/web/common-european-framework-reference-languages/selection-procedures), difficulty values and cut scores are not provided. Further Familiarisation, Standardisation training, Benchmarking and Standard Setting were carried out at the meetings and are described in later sections of this report.

Empirical Validation of the claimed links between the *LanguageCert* Communicator and the CEFR may be informed by qualitative and quantitative analyses and through feedback on the linking procedures from the panellists involved. Validation is seen as an ongoing process that has no end and test providers should continue to gather evidence for the relationship between their tests and the CEFR after an initial linking project has completed.

Panels of experts

It is suggested in the Council of Europe Manual that panels of experts should be convened to make judgements concerning the relationship between a test and the CEFR. The experts, the Manual suggests, should be drawn from both within and outside the organisation producing the test in question. Internal experts should include representatives of the different key stages in language test development while panellists external to the institution should represent a range of different viewpoints. These might include teachers preparing learners for the test, users of results such as receiving institutions and experts in relevant fields such as Applied Linguistics, Language Testing and, for tests of languages for specific purposes, experts in the content domain.

Table 3. Panellist affiliation, gender and years of ELT experience*

<i>Affiliation</i>	<i>Athens</i>	<i>Luton</i>	<i>Panel</i>	<i>Athens</i>	<i>Luton</i>
<i>LanguageCert</i> staff	3	2	Female	6	5
CRELLA, University of Bedfordshire staff		2	Male	2	2
Other (academic)	1	3			
Other (language school sector)	5				
			<i>Years of ELT</i>	<i>Athens</i>	<i>Luton</i>
			Less than 5		
			6 to 10	2	2
			11 to 15		3
			16 to 20	1	
			Over 20	5	2

* Only 15 of the 16 panellists responded to this question

For this project, participants completed a short background questionnaire asking about their qualifications and experience. the Athens panel (Table 3) comprised:

- Three members of the *LanguageCert* staff with a range of responsibilities within the testing programme
- Four experienced language teachers and teacher trainers who had some knowledge of both the CEFR and the *Communicator* examination and made use of them in their work (which included preparing students to take *Communicator* and other tests). These teachers also had experience as examiners for the speaking tests of other international examinations.
- An academic consultant to the LanguageCert programme.

The Luton panel (Table 3) comprised:

- Two members of staff at the Centre for Research in English Language Learning and Assessment at the University of Bedfordshire.
- Two members of the *LanguageCert* test development team.
- Three external experts with expertise in language assessment and Applied Linguistics.

One panellist did not respond to the remaining questions on the questionnaire. Two of the fifteen participants that responded reported that they had taken part in CEFR linking exercises before (both were members of the Luton panel). Five panellists held PhDs and a further three had Masters degrees in language education or related fields. The remaining participants held at least a Bachelor's degree, two reporting that they held a Cambridge DELTA and two a CELTA teaching certificate.

The panel meetings were both coordinated by a researcher from the University of Bedfordshire who has worked with the CEFR since before its publication and has been conducting standard setting studies since 2005 as well as participating in standard setting panels conducted by others. The coordinator ensured the comparability of the procedures across the two linking panels.

Familiarisation

Familiarisation involves training activities to ensure that participants in the linking process have a detailed knowledge of the CEFR, the levels and the illustrative scales. According to the Council of Europe Manual (Council of Europe 2009), familiarisation with the CEFR is a logical pre-requisite for any effective linking to the CEFR and is an essential preliminary to both Specification and Standardisation/Benchmarking.

This study included self-access Familiarisation and Standardisation training activities for all panellists and for the additional *LanguageCert* staff involved in Specification as well as exercises conducted during the panel meetings. The self-access activities included the following:

- Read in detail at least Section 3.6 of the CEFR
- Carry out Phase1 Standardisation training (described below)
- Access an online questionnaire and assign CEFR levels to 32 descriptors drawn from the CEFR

During the panel meetings, further Familiarization activities included rating a set of Can-Do descriptors and discussing the outcomes.

Specification

Specification involves profiling the coverage of the *LanguageCert* Communicator (content and tasks types) in relation to the categories in CEFR Chapters 4 (*Language use and the language learner*) and 5 (*The user/learner's competences*). According to the Manual (Council of Europe 2009), Specification can be used as the basis for (tentative) preliminary claims about a test's implicit level as well as its coverage and is seen as an essential preliminary to 'benchmarking' and 'standard setting' procedures. Forms A2 and A8–A22 in Chapter 4 of the Council of Europe (2009) Manual provide a basis for content analysis and for relating that content to the CEFR. Outcomes of the Specification phase include a detailed description of the test, its qualities and administration procedures.

An adaptation of the Council of Europe forms intended to provide a more descriptive report for users (Green & Inoue 2017) was used to capture relevant information about the *Communicator*

examination. These forms elicit information about features of the testing system including Test Development (A2), Marking (A3), Grading (A4), Reporting Results (A5), Data Analysis (A6) and (A7) the Rationale for Decisions taken in relation to the test. Form A1 (General Examination Description) and the CEFR content grids for Speaking (p.126-131) were used to capture information specific to the examinations at each level of the *Communicator* scheme, to the relevant paper (Listening, Reading, Writing and Speaking) and to the tasks included in each test Part. Table 4 shows the Table of Contents from this document.

Table 4 Table of contents for specification based on Council of Europe (2009), Green & Inoue (2017)

A. The purpose and use of LanguageCert International ESOL Communicator Level

1. Overview of LanguageCert International ESOL Communicator Level

B. Producing LanguageCert International ESOL

2. Development of LanguageCert International ESOL
3. Writing LanguageCert International ESOL

C. The content of LanguageCert International ESOL – Communicator

4. Summary of the content of LanguageCert International ESOL - Communicator
 5. Information on Assessment Tasks
 6. Scoring Assessment Tasks
 7. Reporting scores on LanguageCert International ESOL as a whole
 8. Assessment Results and Analysis
-

The completed Specification forms were sent to the University of Bedfordshire for review together with the (*LanguageCert*) test specifications for the four *LanguageCert* Communicator subtests (Listening, Reading, Writing and Speaking). The researchers annotated the forms with queries and comments and sent them back to the *LanguageCert* team who then revised them. The forms were then checked again by the *LanguageCert* team, who also responded to further queries raised, and a draft was prepared for use in the panel meetings. The completed forms can be found in the Appendix.

During the meetings, panellists were asked to identify which of the CEFR scales appeared most relevant to the subtest in question and to come to a conclusion about the best fit between the *Communicator* test material and the CEFR.

Standardisation training, Benchmarking and Standard setting

The Council of Europe Manual recommends that Standardisation training should be carried out before both (a) *Benchmarking local performance samples* and (b) *Standard Setting*. Like Familiarisation, Standardisation training helps panellists to develop a common understanding of the CEFR Common Reference Levels. The Council of Europe Manual states that Standardisation training typically follows three phases involving working with previously standardised samples and progressing from training in applying the scales towards rating independently:

Phase 1 involves illustration with standardised performances/ tasks (employing samples provided by the Council of Europe) to exemplify the intended levels and use of relevant scales. This was carried out by panellists before the panel meetings. The panellists viewed and rated

performance samples from the *CEFTTrain* project (www.helsinki.fi/project/CEFTTrain). After rating the performances, they compared their ratings with, and read the justifications for those published by the *CEFTTrain* project group.

Phase 2 involves Controlled Practice with standardised performances/ tasks. To practice making judgements using CEFR scales. Phase 3 is a Free Stage with standardised performances/ tasks. The objective is to establish the consistency with which panellists are able to apply the scales. In this project, the panellists' accuracy and consistency in interpreting the scales was addressed as an element of the Benchmarking of (Written and Spoken) performance using a multi-facet Rasch measurement approach and the Standard setting of receptive skills (Reading and Listening). Panellists used CEFR scales to rate selected and previously calibrated sample performances from the Council of Europe (Centre international d'études pédagogiques: CiEP 2008 and English Profile 2010) as well as previously scored *Communicator* performances. For Listening and Reading, they rated illustrative tasks provided by the Council of Europe alongside material from *Communicator* tests.

As in the earlier City & Guilds studies, an important aspect of the Benchmarking/ Standard setting phase of the project was the definition of the "minimally competent" B2 level learner of English. In making their judgements, the panellists were asked to estimate how well a minimally competent B2 level language learner might perform on the test material. This would not be a learner who could demonstrate mastery of all the Can-do descriptors calibrated at the B2 level, or one who was just capable of the easiest B2 level Can-do descriptor, but who would (marginally) be better described as B2 rather than as B1+. Following the formulation used in developing the CEFR illustrative scales, they, "Could be expected to perform [a bare majority of B2 activities] without support in normal circumstances" (North 2001).

Speaking and Writing

For Speaking, the panels listened to nine standardised speaking performance samples (from CiEP: hereafter 'CiEP samples'). The CiEP samples had previously been rated in terms of CEFR levels at the cross-language benchmarking seminar held in Sèvres, reported in Breton, Lepage and North (2008). The cut-off logit values for each CEFR level established through that seminar are published in North and Jones (2009). The Athens and Luton panellists also rated 14 local performance samples. Each of these samples represented a candidate's performance on one part of the *Communicator* speaking subtest. These had previously been scored using the *Communicator* rating scales and had each been selected by *LanguageCert* as representative of a *Communicator* score band.

For Writing, there remains only one written sample for English from the Council of Europe website representing the B2 level (the same B2 sample used in the City & Guilds project by O'Sullivan 2008). Along with this sample, the panellists were given five samples of learner writing previously employed at an English Profile seminar held in Cambridge in 2010 and identified there by over 30 participants from around the world as representative of CEFR levels A2 to C1. The

panellists were also provided with 18 local performance samples from the *Communicator* Writing subtest. For both Speaking and Writing, the previously calibrated CEFR samples and local samples were presented in randomised order during the same session so that each panellist scored all of the calibrated CEFR samples and a selection of *Communicator* samples. These performances were scored using the relevant Council of Europe (2018) assessment scale: CEFR Table 3 Qualitative features of spoken language (expanded with phonology) (pp.155-6) or Manual Table C4: Written Assessment Grid (pp.157-8). For both Speaking and Writing, panellists awarded a single (holistic) score, but noted any discrepancies between the overall rating and individual scoring criteria (such as an overall B2 level performance that featured only a B1 level of *fluency*, but a C1 *range*).

The panellists' ratings were analysed using the many-facet Rasch measurement (MFRM) software, FACETS (Linacre 2015). Through this analysis, the *Communicator* performance samples were located on a common scale in relation to the cut-off points for the CEFR levels. This provided quantitative evidence of the links between the *LanguageCert* Communicator and the CEFR. Anchoring techniques (Linacre, 2013) were used for both the cut-off points for the CEFR levels and for the ability measures of performance samples. North and Jones (2009), recommend that these values can be used to anchor the scale steps used in other CEFR-linking projects. In addition to the cut-off logit values for the CEFR levels, logit values (i.e. ability measures) for CiEP samples were supplied by North (2016, personal communication) and used as anchors. This further strengthened the connections between the Council of Europe Sèvres seminar, the standardised samples and this study.

For this project, as suggested in North and Jones (2009), in order to ensure that the panellists' application of CEFR levels to the performance samples was consistent with the scales from the Sèvres and English Profile seminars, the panellists' ratings were first analysed without any anchoring; this was to check whether any levels were too closely placed or placed in reverse order. This unanchored FACETS run on the panellists' ratings in this study demonstrated that the scale steps were generally even and the adequacy of using anchoring values from the Sèvres and English Profile seminars was verified.

The CEFR levels awarded by the panellists (both in Athens and in Luton) to the spoken performance samples from the CiEP; written samples from the Council of Europe (First Certificate Band 3) and English Profile; and samples from the *Communicator* examination were collated and analysed using the MFRM software, FACETS. The CEFR levels were converted to numerical values as follows: A1 = 1; A2 = 2; A2+ = 3; B1 = 4; B1+ = 5; B2 = 6; B2+ = 7; C1 = 8, C2 = 9. The CiEP samples had previously been rated according to the CEFR levels in the cross-language benchmarking seminar in Sèvres, reported in Breton, Lepage and North (2008). Ability measures in logit values for these performances were obtained from North (2016, personal communication) together with the cut-off logit values for all the CEFR levels identified in the same seminar, which are published in North and Jones (2009) (see Table 11).

Table 5 Cut-off Logit Values for each CEFR Level (from North & Jones, 2009)

<i>CEFR level</i>	<i>Measure</i>
A1	-4.29
A2	-3.23
A2+	-2.21
B1	-1.23
B1+	-0.26
B2	0.74
B2+	1.72
C1	2.80
C2	3.90

As no ability estimate was available for one performance, Sylvia¹, identified by CiEP as exemplifying B2. Her performance was assigned a value at the mid-point of the B2 range.

No empirically derived logit values are available for the written samples. On the assumption that these represented performances at the mid-point of each level, they were assigned a value reflecting this. A B2 writing sample was thus assigned a value of 1.23 logits. One exception was a sample from the English Profile seminar identified as being on the B1+/B2 borderline. This was assigned a value of 0.74 (the score marking the B1+/B2 boundary). These spoken and written performances were used as anchors in the analysis.

A total of 23 written and 22 spoken performance samples was used in the panel meetings. To make efficient use of the time available, the panellists did not all score every performance sample. Instead, for Writing each panellist was assigned between 12 and 16 samples to score and each Writing sample was scored by between 6 and 13 panellists. For Speaking, the Athens panel was asked to score 13 and the Luton panel 15 samples. Six Speaking samples were scored by all panellists.

Reading and Listening

For Reading and Listening, an extended Angoff method (Hambleton et al. 2000) was employed. In the *Communicator* Reading and Listening tests, each of the four test tasks or testlets on each test paper consists of a text or recording accompanied by six or seven three-option multiple choice questions. Alongside the material from the *Communicator* tests, the panellists also estimated the scores that a minimally competent B2 test taker would earn on sample tasks provided by the Council of Europe to illustrate the B1, B2 and C1 levels.

Although the extended Angoff method was developed for use with constructed-response tasks, the selected-response items on these tests are not entirely independent of each other and it may not be appropriate to judge the difficulty of each item without reference to the source text or recording and the other items in the task. The tasks were therefore treated as polytomous (scaled) items.

¹ Rather confusingly, this individual is referred to in the accompanying notes as Sylvie rather than Sylvia (as her name appears on the CiEP video). She can be recognised because she is partnered with a boy named Paul.

Panellists were asked to judge for which CEFR level the material seemed most suitable and to estimate the score (out of the maximum number of points available) that a minimally competent B2 test taker would earn on each task. A further advantage of this methodology is that judges find the extended Angoff approach more intuitive and are able to reach decisions more quickly than with the procedures employed in the City & Guilds project (O'Sullivan 2008). This made it possible to review multiple forms of the tests as well as a number of Council of Europe sample tasks within the limited time available.

After arriving at their initial estimates, panellists were given an opportunity for discussion and (in the case of the *Communicator* material) were provided with facility values from trials of the test material. Following discussion, they were allowed to revise their estimates before submitting them. The average of the panellists' estimates was then calculated for each task. The standards on the total score scale (out of 26 points) on the *Communicator* tests were calculated as an aggregate of these task averages.

Because of the lack of previously calibrated difficulty values for the Council of Europe material, it was not possible to place the *Communicator* material and illustrative tasks on a common scale and multi-facet Rasch analysis was not employed. However, panellists were encouraged to discuss their estimates and were free to revise them following the discussion. This served to confirm the panellists' interpretation of the CEFR levels applying to Spoken and Written Reception.

Validation

It is stated in the Council of Europe Manual that for any claim of linking to be made, satisfactory evidence must be provided for the validity of scores derived from the test for its intended purpose. Users are directed to sources such as the European Association of Language Testing and Assessment Guidelines for Good Practice in Language Testing and Assessment (www.ealta.eu.org) for advice on issues of test quality. Evidence for the validity of the *Communicator* tests falls outside the scope of this report, but can be obtained from the *LanguageCert* website at: www.languagecert.org.

The Council of Europe Manual recommends that an adequate linking process should provide evidence for 'procedural validity', or the effectiveness of the methods adopted, and 'internal validity' of the standard setting results: the accuracy and the consistency of the judgements made. This report provides evidence for the procedural validity of each stage of the study under the relevant headings. It includes evidence of the qualifications of the participating panellists, the documentation of rigorous and systematic procedures and evidence that participants demonstrated sufficient knowledge of the framework and were satisfied that the processes were successful. The qualifications of the panellists are outlined above and evidence of their knowledge of the CEFR is reported under the heading of Familiarisation. The triangulation of institutional, researcher and independent panellist perspectives contribute to the validity of the Specification phase. Evidence that panellists were satisfied with the procedures was sought through a post-panel questionnaire

concerning their understanding of the CEFR levels and satisfaction with the meeting and procedures followed.

In relation to internal validity, evidence for the consistency of participant judgements and the extent to which these were based on a sound understanding of the CEFR can be found in the statistics reported for the Benchmarking/ Standardisation stage of the project, from the outcomes of Familiarisation tasks and from the post-panel questionnaire.

Results

Familiarisation

Prior to the panel meetings all 16 panellists were asked to respond to an online questionnaire that included questions relating to their background and experience, a self-assessment of their level of familiarity with the CEFR and a descriptor rating exercise that involved assigning 32 descriptors from the CEFR.

Table 6 How would you describe your current knowledge of the CEFR? (N=15)

	<i>Greece</i>	<i>UK</i>	<i>Total</i>
Thorough			
Good	7	6	13
Basic	1	1	2
Poor			
Zero			

15 of the 16 panellists completed this online activity. Two reported a ‘basic’ and 13 a ‘good’ knowledge of the CEFR at this point. Some panellists explained their self-assessments with comments such as the following: “*Basic*: Although I have read through them, I think there are lots of details that need reading repeatedly to learn them thoroughly.” (Athens panellist); “*Good*: I reckon I know CEFR more through my working experience than actually studying and learning it before being practically exposed to it.” (Athens panellist); “*Good*: I feel able to examine examples of language use by learners and distinguish between the major levels of the CEFR. I am able to relate the statements below to their relevant CEFR level, although do not have comprehensive memory recall of the different levels.” (Luton panellist).

Overall, 74.2% (356) of the 480 level estimates made by the 15 panellists who responded were correct (the level assigned by the panellists matched the level assigned in the CEFR). 95.6% of the panellists’ estimates were within one band of the correct level. However, these figures were somewhat higher for the Luton panel (83.3% correct, 98.4% within one band) than for the Athens panel (68.1% correct, 93.8% within one band). This suggests that the Luton panellists were a little more conversant with the CEFR scales than their Athens counterparts, but that both groups had a very good understanding of the Common Reference Levels. At the meetings in Athens and Luton, panellists rated a set of a further 11 descriptors drawn from all levels of the CEFR (Council of

Europe 2001 p.58 and p.74). In both cases panellists worked in groups to discuss their responses and were consistently correct in their estimates, although individual results were not recorded.

The *Communicator* tests were already familiar to the Athens panellists, but not to those in Luton. Before the meetings, all panellists were provided with the Specification forms for *Communicator* (see Appendix). These set out features of the tasks included in each test part including the nature of any input and characteristics of the expected responses. The rating scales used by *Communicator* examiners were included, together with a sample form of each test.

Specification

Speaking

The panellists compared the *Communicator* rating scales for Speaking with the CEFR. They identified points of similarity and difference between the scales. These are set out in Table 6.

Table 7 Correspondences between *Communicator* rating scales for Speaking and the CEFR

<i>Communicator</i>	<i>Descriptor</i>	<i>CEFR</i>	<i>Descriptor</i>
Task fulfilment and coherence Band 2	Handles the interaction with relative ease, managing the conventions of turn-taking, using appropriate phrases, well, though not always very 'elegantly'.	Turntaking, B2	Can initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly.
	Expresses his/her message clearly, with relevant supporting detail where appropriate.	Thematic development, B2 Informal discussion (with friends), B2	Can develop a clear description or narrative, expanding and supporting his/her main points with relevant supporting detail and examples. Can account for and sustain his/her opinions in discussion by providing relevant explanations, arguments and comments.
	Content/contributions are mostly relevant to the conversation/topic, and it is rare that any misunderstanding occurs.		
	Uses a limited range of cohesive devices to link his/her utterances into clear, coherent discourse but there may be some 'jumpiness' in long contributions.	Coherence and cohesion, B2	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.
Accuracy and range of grammar Band 2	Sufficient range of the grammatical structures expected at B2, a relatively high degree of grammatical control.	Grammatical control, B2	Shows a relatively high degree of grammatical control.
	Errors occur, but they do not impede communication.	Grammatical control, B2	Does not make mistakes which lead to misunderstanding.
Accuracy and range of vocabulary Band 2	Sufficient range of vocabulary to deal with the tasks at B2 with a degree of formality appropriate to the circumstances.	Vocabulary Range, B2	Has a good range of vocabulary for matters connected to his/her field and most general topics.

<i>Communicator</i>	<i>Descriptor</i>	<i>CEFR</i>	<i>Descriptor</i>
	Some errors occur, but they do not impede communication.		Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution.
Pronunciation, intonation and fluency Band 2	Pronunciation reasonably clear and is understood without difficulty.	Phonological Control: Overall Phonological Control, B1 Phonological Control: Sound Articulation, B1	Pronunciation is generally intelligible; can approximate intonation and stress at both utterance and word levels. However, accent is usually influenced by other language(s) he/she speaks. Is generally intelligible throughout, despite regular mispronunciation of individual sounds and words he/she is less familiar with.
	Intonation is natural, has a fairly even tempo and contributes to the content. Few noticeably long pauses, although there may be some hesitation when searching for patterns and expressions.	Spoken Fluency B2 General Linguistic Range, B2	Can produce stretches of language with a fairly even tempo; although he/she can be hesitant as he/she searches for patterns and expressions, there are few noticeably long pauses. Has a sufficient range of language to be able to give clear descriptions, express viewpoints and develop arguments without much conspicuous searching for words, using some complex sentence forms to do so.
	No undue strain on the listener.		Can interact with a degree of fluency and spontaneity that makes regular interaction with speakers of the target language quite possible without imposing strain on either party. Can deliver announcements on most general topics with a degree of clarity, fluency and spontaneity which causes no strain or inconvenience to the listener.

Although some discrepancies were identified (notably in the recently updated CEFR scales for Phonological Control, which could inform a minor revision of the *Communicator* scales), the parallels between the *Communicator* rating scales and the CEFR reflected the intentions of the test developers that Band 2 should embed the B2 level. Commonalities were also noted between the Band 3 *Communicator* descriptors such as, “wide range of vocabulary and a consistently high level of vocabulary” and C1 level statements such as, “Has a good command of a broad range of language” and, “Consistently maintains a high degree of grammatical accuracy”. On the other hand, Band 1 often seemed to reflect A2 rather than B1: “Connects ideas relatively simply and may be difficult to follow in places”; “Frequent hesitation”; “Errors occur, some of which impede comprehension”; “Undue strain on the listener”; “influence of L1 is strong on pronunciation, intonation and/or stress patterns” (Band 1) seemed to reflect A2 descriptors, “Can make him/herself understood... pauses, false starts and reformulation are very evident”; “still systematically makes basic mistakes”; “Can link groups of words with simple connectors” and, “A strong influence from other language(s) he/she speaks on stress rhythm and intonation may affect intelligibility” rather than B1, “Can keep going comprehensibly even though pausing... is very evident”; “Can keep going comprehensibly...”; “Can link a series of shorter, discrete simple

elements into a connected linear sequence of points” and “Pronunciation is generally intelligible...”.

The panellists reviewed Speaking task forms and discussed at which CEFR levels test takers would be able to carry them out. The results are summarised in Table 7.

Table 8. At which CEFR level would a test taker be able to respond successfully to this task?

Test Part	A1	A2	B1	B1+	B2	B2+
VER 417CS Part 1			1	4	2	
VER 417CS Part 2			3	2	2	
VER 417CS Part 3				2	5	
VER 417CS Part 4			1	4	2	
VER 417HS Part 1			5		2	
VER 417HS Part 2			5	1	1	
VER 417HS Part 3			1	2	4	
VER 417HS Part 4				2	5	
[Specification] Part 1	2	1	1		5	
[Specification] Part 2			1		8	
[Specification] Part 3			1		8	
[Specification] Part 4			1		7	1

The Luton panel reviewed Tests 417CS and 417HS (eight speaking tasks). They concluded that all of the tasks were accessible to test takers at B2, but that much of the material would also be manageable for test takers at the B1 or B1+ levels. Parts 3 and 4 were generally agreed to be more challenging than Parts 1 and 2, although most panellists found 417CS Part 4 to be suited to B1+ test takers (similar to Part 1). Parts 1 and 2 of 417HS were judged to be easier than 417HS Parts 1 and 2 and were considered by the majority of panellists to be suitable for B1 level test takers. The Athens panel reviewed the version of the Speaking test included in the Specification phase of this project. Most identified all parts of the test with B2, but two felt that Part 1 could be accessible to test takers at A1 with one panellist placing it at A2 and another at B1. One panellist considered all parts of the test to be suitable for A1 level test takers, but another felt that Part 4 would only be accessible to test takers at B2+.

The picture that emerged from this phase of the meetings was of a test that was well attuned to the B2 level, although some of the tasks might also be successfully carried out by test takers at B1. This seems appropriate, particularly in the earlier parts of the test, intended as a relatively undemanding ‘warm-up’ phase.

Writing

The panellists compared the *Communicator* rating scales for Writing with the scales in the CEFR and Manual Table C4: Written assessment grid (Council of Europe 2018, p.173). They identified points of similarity and difference between the scales.

Table 9 Correspondences between Communicator rating scales for Writing and the CEFR.

<i>Communicator</i>	<i>Descriptor</i>	<i>CEFR</i>	<i>Descriptor</i>
Task Fulfilment, Task1 Band 2	- covers at least 2 content points with some expansion and communication mainly achieved	Creative Writing, B2+	Can write clear, detailed descriptions of real or imaginary events and experiences marking the relationship between ideas in clear connected text, and following established conventions of the genre concerned.
	- genre and tone mostly appropriate		
Task Fulfilment, Task 2 Band 2	- covers both points with some expansion and communication mainly achieved	Essays and Reports, B2	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.
	- genre and tone mostly appropriate		
Accuracy and range of Grammar Band 2	- uses simple and some complex forms with a good degree of control	Grammatical Control, B2	Shows a relatively high degree of grammatical control. Has a good command of simple language structures and some complex grammatical forms, although he/she tends to use complex structures rigidly with some inaccuracy.
	- errors do not impede meaning, <i>but may cause rereading</i>		Does not make mistakes which lead to misunderstanding.
Accuracy and range of Vocabulary Band 2	- uses a range of everyday vocabulary accurately, with occasional misuse of less common items	Vocabulary Range, B2	Has a good range of vocabulary for matters connected to his/her field and most general topics.
	- errors do not impede meaning, but may cause re-reading	Vocabulary Control, B2	Lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication.
Organisation Band 2	- text is generally well-organised and coherent using a variety of linking words and cohesive devices	Coherence & Cohesion, B2	Can produce text that is generally well-organised and coherent, using a range of linking words and cohesive devices.
	- organization mainly appropriate to text type		Can structure longer texts in clear, logical paragraphs.
	- some punctuation errors that don't impede communication		Spelling and punctuation are reasonably accurate but may show signs of mother tongue influence.

As with Speaking, the panellists found that Band 2 of the *Communicator* scales generally reflected the intended B2 level of the CEFR. As with Speaking, correspondences were noted between Band 3 and C1: “text is well organised and coherent, using a variety of cohesive devices” matched “controlled use of organisational patterns, connectors and cohesive devices” (C1) and fully addressing the content points suggested, “expand and support points of view at some length” (C1). However, some descriptors at Band 3 also seemed to match CEFR B2. For the criteria of *Accuracy and range of grammar* and *Accuracy and range of vocabulary*, “Does not make errors which cause misunderstandings” (C1) was considered to match B2 better than the Band 2 descriptor, which included the qualification, “but may cause rereading” (italicised in Table 8 above). Similarly, “genre and tone appropriate” (Band 3, *Task fulfilment*) reflected, “following

established genre conventions” at B2 for *Description*. Again, descriptors at Band 1 seemed to fall short of B1. For example, “errors impede meaning at times” was closer to A2 (“Errors may sometimes cause misunderstandings”) than B1 (“Occasionally makes errors that the reader usually can interpret correctly on the basis of the context”). Equally, “text is connected using basic linking words and a limited range of cohesive devices” at Band 1 might reflect the “simple connectors” mentioned at A2 rather than the “connected, linear text” mentioned at B1 for *Coherence*.

The limited connection between Communicator Band 1 on the Speaking and Writing papers and B1 level descriptors could be regarded as problematic if it is intended that the test should discriminate between the A2 and B1 levels. However, a test taker classified as Band 2 across the criteria would generally match the description of B2 in the CEFR illustrative scales. It may also be the case that B1 level test takers confronted with tasks that are too demanding (but suitably challenging for B2 level learners) might fail to perform in a way that would evidence their ability to carry out B1 level tasks, making A2 level descriptors more appropriate to describe their performance on the test.

The panellists reviewed two Writing tasks (the ones used to elicit the sample performances for this project) and discussed at which CEFR levels test takers would be able to carry them out. As with the Speaking tasks, they agreed that all of the tasks were suitable for test takers at B2, but half of the panellists considered that Task 1 would also be manageable for test takers at the B1 level. All agreed that Task 2 was more demanding than Task 1, but a minority (four of the 16) thought that Task 2 was also suitable for B1 test takers.

Table 10 At which CEFR level would a test taker be able to respond successfully to this task?

Test Part	B1	B2
Part 1	8	8
Part 2	4	12

Listening

Panellists were asked to identify at which level they felt the Listening tasks on the Communicator tests they reviewed would be most appropriate, although not all panellists recorded a conclusion for all Parts of the tests under review. The results are shown in Table 10.

Table 11 At which CEFR level would this task be most appropriate?

Test Part	B1	B1+	B2	B2+
Calibration Part 1	4	1	11	
Calibration Part 2		1	15	
Calibration Part 3 ¹	2	1	12	1
Calibration Part 4		2	14	
VER. 417A Part 1	3	1	7	
VER. 417A Part 2		3	8	
VER. 417A Part 3 ¹	2	1	12	1

VER. 417A Part 4			10
VER. 417H Part 1	4	2	6
VER. 417H Part 2	1	4	7
VER. 417H Part 3		2	10
VER. 417H Part 4	1		11

The majority of panellists identified all Parts of the *Communicator* Listening test with the B2 level, although in most cases, a minority felt that the tasks were suitable for test takers at B1 or B1+. One panellist identified one of the tasks (Part 3 of the Calibration test and VER 417A) with B2+.

Parts 1 (intended task focus: recognize context, meaning and function in a range of short conversations on concrete and abstract topics and select appropriate response), 2 (identify specific aspects of spoken dialogue) and 4 (follow a discussion, listen for specific information and identify attitudes and opinions), which all involve conversations, suggested *Understanding conversation between other speakers*. A B2+ descriptor mentions keeping up with, “an animated conversation” and B2 refers to the listener being able to, “catch much of what is said around him/her” and these seemed to reflect Part 1 and Part 4 in particular. Part 2 seemed to reflect the ability to, “identify the main reasons for or against an argument” or the C1 descriptor, “identify the attitude of each speaker”. On the other hand, several panellists noted that speech was “clearly articulated”, “standard speech”, both described at B1+ level. Parts 1 and 2 also suggested *Listening to the radio and audio recordings* which mentions the ability to, “identify speaker viewpoints and attitudes” at B2+, but that also locates the, “clear standard speech” felt to characterise the recordings at B1+.

Part 3 (extract key information from a monologue) suggested *Listening to announcements and instructions* including, at B2, being able to, “understand announcements and messages on concrete and abstract topics spoken in standard speech at normal speed” and, “understand detailed instructions well enough to be able to follow them successfully”, although there was some doubt as to whether the test tasks might only require the ability to, “follow detailed directions” at B1. It was again noted that the recordings seemed to be, “clearly articulated in standard speech with minimum interference from background noise”: described at B1.

Panellists also found correspondences with other scales across test Parts. It was suggested that topics were, “both abstract and concrete” as mentioned in *Overall Listening comprehension* at B2 and, “both familiar and unfamiliar” at B2+, combined with “extended speech and complex lines of argument” (B2), although there was some disagreement about whether speech was, “propositionally and linguistically complex” (B2) and whether content consisted of, “straightforward factual information about common everyday or job related topics”. Similar wording in the scale for *Listening as a member of a live audience* was also referenced: “Can follow complex lines of argument in a clearly articulated lecture” (B2) balanced by, “Can follow a lecture... provided the subject matter is familiar and the presentation straightforward and clearly structured” (B1+).

The consensus view of the *Communicator* Listening tests across the two panels was that they reflected the demands of the B2 level, but that they also included material that might be suitable for B1+ level test takers.

Reading

Panellists were asked to identify at which level they felt the Reading tasks on the Communicator tests they reviewed would be most appropriate. As with Listening, not all panellists recorded a conclusion for all Parts of the tests under review. The results are shown in Table 11.

Table 12 At which CEFR level would this task be most appropriate?

Test Part	B1	B1+	B2
Calibration Part 1		2	12
Calibration Part 2	1	1	13
Calibration Part 3 ¹		5	11
Calibration Part 4 ²			15
VER. 417A Part 1	1		14
VER. 417A Part 2			15
VER. 417A Part 3 ¹		5	11
VER. 417A Part 4 ²			15
VER. 417H Part 1			12
VER. 417H Part 2			12
VER. 417H Part 3		1	11
VER. 417H Part 4			12

In all cases, the majority of panellists identified the test tasks with B2, although for one task (Part 3 of the Calibration test and version 417A), five panellists felt that the material would be suitable for test takers at the B1+ level.

The panellists also matched test material to illustrative CEFR scales and identified descriptors that seemed to reflect the test tasks. Compared with Listening, a narrower range of CEFR scales was invoked in the discussion. Two scales seemed to be particularly salient: *Reading for information and argument* and *Reading for orientation*. Both scales were cited by panellists in relation to all four Parts of the test, although, reflecting the intentions of the developers, the former was more closely identified with Part 1 (intended task focus: Show detailed understanding of information, ideas and opinions in a longer text) and Part 2 (Show understanding of how meaning is built up in discourse) and the latter with Part 3 (Show awareness of the purpose of different texts and the ability to locate specific information. To scan for information.) and Part 4 (Locate and obtain specific information).

From the *Reading for orientation* scale, “Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints” and, “Can recognise different structures in discursive text: contrasting arguments, problem-solution presentation and cause-effect relationships” are both found at B2 and were identified by panellists

with tasks throughout the tests. It was also noted that tasks involved the ability to, “obtain information ideas and opinions” (B2+), although there was disagreement over whether any of the texts on the test could be considered, “specialised” as described in the CEFR.

For *Reading for information and argument*, relevant descriptors included, “Can scan quickly through long and complex texts, locating relevant details” and, “Can quickly identify the content and relevance of news items, articles and reports on a wide range of professional topics” at B2, but also, “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” and, “Can scan through straightforward, factual texts in magazines, brochures or in the web, identify what they are about and decide whether they contain information that might be of practical use”, which are placed at B1+.

Overall, the panellists agreed that the test tasks were best suited to the B2 level and that they mainly involved *Reading for information and argument* and *Reading for orientation*. Across all parts of the test, the review of the material suggested that it was well-matched to the B2 level.

Benchmarking and Standard setting: Speaking and Writing

Table 12 shows the distribution of the scores awarded by the panellists to the spoken performance samples they listened to.

Table 13 CEFR scores awarded by panellists to samples of spoken performance. Figures in parentheses are ‘official’ results awarded at the Council of Europe Sèvres seminar or by LanguageCert examiners as relevant.

Candidate	Observations	Mean*	St. dev.	Mode	CEFR
candidate 12 (2: Fail)	7	1.00	0.00	1	A1
candidate 14 (1: Fail)	7	1.00	0.00	1	A1
candidate 11 (4: Fail)	8	1.88	0.35	2	A2
candidate 13 (4: Fail)	9	2.22	0.44	2	A2
candidate 03 (6: Pass)	6	4.50	0.84	4	B1+
candidate 01 (7: Pass)	8	4.63	1.41	4	B1+
candidate 05 (6: Pass)	16	4.88	1.15	6	B1+
<i>Theo (prtnr Amelie) (B1)</i>	8	5.00	1.31	5	B1+
<i>Amelie (B1+)</i>	7	5.00	1.00	6	B1+
candidate 02 (8: Pass)	8	5.38	0.52	5	B1+
candidate 09 (8: Pass)	16	5.40	0.74	6	B1+
candidate 06 (6: Pass)	7	5.43	0.53	5	B1+
candidate 08 (10: High Pass)	8	5.75	0.89	6	B2
candidate 10 (8: Pass)	16	6.00	0.63	6	B2
candidate 07 (9: High Pass)	7	6.00	0.00	6	B2
<i>Theo (prtnr Blandine) (B2)</i>	16	6.25	0.58	6	B2
<i>Sylvia (B2)</i>	9	6.33	0.50	6	B2
<i>Blandine (B2)</i>	16	6.44	0.73	6	B2
<i>Paul (B2)</i>	16	6.56	0.73	6	B2
<i>Charlotte (C1)</i>	7	7.71	0.49	8	C1

<i>Annabelle (C1)</i>	9	8.44	0.53	8	C1
<i>Xavier (C2)</i>	8	8.63	0.52	9	C2

*Note: A1 = 1; A2 = 2; A2+ = 3; B1 = 4; B1+ = 5; B2 = 6; B2+ = 7; C1 = 8, C2 = 9

From Table 12, it can be seen that judgements of the Council of Europe sample performances were in line with the levels assigned at the Sèvres seminar, but that many of the *Communicator* sample performances that had been awarded a Pass were judged to be B1+ rather than B2.

Table 14 Fit Statistics for Speaking for rater measurement (in order of rater severity)

<i>Rater ID*</i>	<i>Measure</i>	<i>S.E.</i>	<i>Infit Mnsq</i>
L4	0.47	0.29	0.99
A3	0.31	0.31	0.32
A9	0.28	0.31	0.38
L3	0.23	0.28	0.30
L5	0.15	0.28	0.55
A5	0.06	0.3	0.46
A4	0.03	0.3	0.57
A1	0.03	0.3	0.48
A2	0.03	0.3	0.40
A8	0.03	0.3	0.27
L7	0.00	0.28	0.42
A6	-0.15	0.31	1.15
L1	-0.16	0.28	0.66
L6	-0.33	0.28	0.98
A7	-0.48	0.36	1.59
L2	-0.49	0.28	0.45

*Note. A indicates Athens; L indicates Luton.

In the second column of Table 13 above, panellists are ordered according to their severity; the higher they are on the logit scale, the harsher they are. The FACETS rater measurement report (summarised in Table 13) shows that the panellist located at the top of the *Rater ID* column, L4 (i.e. panellist 4 from the Luton panel), has the highest logit value of 0.47, and is therefore the harshest. The separation reliability² was 0.0, which indicates that the panellists did not differ significantly in severity. It can also be seen that there is no substantial difference between the two panels: the Athens panellists were, on average marginally harsher than the Luton panel (by 0.034 logits): not enough to affect the classification of any of the performances. The infit mean squares indicate that most of the panellists fit well with the model, falling within the effective measurement range for judgement data of 0.4 to 1.2 (Wright & Linacre, 1994). The exceptions included A7 (with an infit mean square value of 1.59). It was decided to retain this rater in the dataset because this infit mean square value was smaller than 2, not so large as to pose a threat to effective measurement

² The separation reliability calculated by FACETS indicates how reliably the analysis is separating the raters into different levels of severity, and thus is different from traditional inter-rater reliability.

(Linacre, 2013). Raters A3, A8, A9 and L3 ‘overfit’ the model. This indicates that their ratings were more consistent with the consensus than predicted by the Rasch model. In this context, overfit is not problematic.

Matching the Communicator Exam Performances to CEFR Levels

Once it was established that the rater measurements were acceptable, all the performance samples were ordered according to their logit measures and were compared against the threshold logit measures for CEFR levels derived from the Sèvres cross-language benchmarking seminar (Table 14).

Table 15 Person Measurement Report for Communicator spoken performance samples and corresponding CEFR Levels

CEFR	Sample ID ³	Fair Average	Measure	S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd
	Cand. 12 (2: Fail)	1.04	-6.47	1.91	1	0	1	0
	Cand. 14 (1: Fail)	1.04	-6.47	1.91	1	0	1	0
	Cand. 11 (4: Fail)	1.81	-3.18	0.45	0.22	-2.24	0.22	-2.21
A2			-3.23					
	Cand. 13 (4: Fail)	2.12	-2.73	0.38	0.25	-2.35	0.24	-2.41
A2+			-2.21					
B1			-1.23					
	<i>Theo (Amelie) (B1)</i>	3.58	-1.16	0.6	2.86	2.74	2.84	2.72
	Cand. 1 (7: Pass)	3.89	-0.85	0.34	1.05	0.26	1.05	0.26
	Cand. 3 (6: Pass)	4.29	-0.46	0.41	0.49	-0.92	0.49	-0.92
B1+			-0.26					
	<i>Amelie (B1+)</i>	4.71	-0.04	0.38	0.51	-0.94	0.51	-0.94
	Cand. 5 (6: Pass)	4.72	-0.03	0.25	1.05	0.26	1.05	0.25
	Cand. 2 (8: Pass)	5.19	0.44	0.35	0.23	-2.18	0.23	-2.18
	Cand. 9 (8: Pass)	5.23	0.47	0.26	0.55	-1.36	0.55	-1.37
	Cand. 6 (6: Pass)	5.3	0.54	0.38	0.15	-2.49	0.15	-2.49
B2			0.75					
	Cand. 8 (10: High Pass)	5.62	0.87	0.35	0.37	-1.53	0.37	-1.53
	Cand. 10 (8: Pass)	5.85	1.1	0.25	0.41	-2.08	0.41	-2.07
	Cand. 7 (9: High Pass)	5.88	1.13	0.38	0.12	-2.69	0.12	-2.7
	<i>Paul (B2)</i>	6.17	1.42	0.25	0.69	-0.88	0.69	-0.89
	<i>Theo (Blandine) (B2)</i>	6.2	1.45	0.25	0.24	-3.12	0.24	-3.1
	<i>Blandine (B2)</i>	6.22	1.48	0.25	0.57	-1.36	0.57	-1.36
	<i>Sylvie (B2)</i>	6.37	1.63	0.34	0.25	-2.24	0.25	-2.23
B2+			1.72					
C1			2.80					
	<i>Annabelle (C1)</i>	7.84	3.28	0.39	0.86	-0.17	0.85	-0.2
	<i>Charlotte (C1)</i>	8.15	3.73	0.48	0.73	-0.39	0.81	-0.22
C2			3.90					
	<i>Xavier (C2)</i>	8.7	5.01	0.67	0.88	0.03	0.91	0.09

Table 14 shows the FACETS person measurement report for all the *Communicator* performance samples, with the threshold values for CEFR levels indicated by grey shading. The

³ The numbers and letters in parentheses after each Sample ID represent the CEFR level, or LanguageCert score awarded at the Sèvres seminar or by LanguageCert examiners as relevant.

LanguageCert performance samples represented three score levels on the *Communicator* speaking tests:

- Fail: These were all classified by the panellists as A2 or below.
- Pass: These were Pass scores on *Communicator* of between 6 and 8 and were classified by the panellists as B1+, although Candidate 1 was marked more harshly and Candidate 6 more leniently than their official scores suggested. Anchoring these ratings with FACETS suggested that the weaker performances at this level might be better classified as B1.
- High Pass: Both test takers with High Pass scores (candidates 7 and 8) were scored as B2, along with one Pass candidate with a score of 8 (candidate 10). This suggests that scores of 9 or 10 could most clearly be classed as B2. This leaves open the question of whether scores of 11 or 12 might correspond to B2+.

Figure 4 Variable map of spoken performance samples rated by Athens and Luton panels

Measr	+Samples	-Raters	-Scale	CEFR
6				(C2)
5	Xavier (C2)			
4	Charlotte (C1)			C1
3	Annabelle (C1)			
2				B2+
1	Sylvie (B2) Blandine (B2) Candidate 10 (8:P) Paul (B2) Theo (Blandine) (B2) Candidate 7 (9:HP)			B2
0	Candidate 8 (10:HP) Candidate 6 (6:P) Candidate 2 (8:P) Amelie (B1) Candidate 5 (6:P)	L4 A3 A9 L3 L5 A1 A2 A4 A5 A8 L7 A6 L1 L6 L2 A7	CEFR	B1+
-1	Candidate 3 (6:P) Candidate 1 (7:P) Theo (Amelie) (B1)			B1
-2				A2+
-3	Candidate 13 (4:F) Candidate 11 (4:F)			A2
-4	Candidate 12 (2:F) Candidate 14 (1:F)			(A1)

The results of the MFRM analysis can be summarised in a variable map or ruler (Figure 4). In the first column, *Measr* represents the logit scale on which all the variables are placed. It is centred around zero and larger values represent higher ability on the part of test takers, greater difficulty of test tasks or harsher judgement (a tendency to award lower scores) on the part of panellists (referred to as raters).

The second column shows where the performance samples fall on the logit scale. The higher they are placed, the higher the performance was rated. In Figure 3, Xavier and Charlotte (Council of Europe exemplars of C2 and C1) were the highest performing test takers, and the lowest performing were Candidate's 12 and 14: *Communicator* test takers who were scored by the panellists as A1. The figures in parentheses show the scores originally awarded to these performances. In the case of the CiEP samples, these are CEFR levels awarded at the Sèvres seminar and in the case of *Communicator*, overall scores on the *Communicator* Speaking scale awarded by *LanguageCert* examiners: the sum of the scores for the four criteria of *Task Fulfilment & coherence*, *Accuracy and Range of Grammar*, *Accuracy and Range of Vocabulary* and *Pronunciation, Intonation and Fluency*.

The third column shows the relative severity of the raters and shows how far they diverge in terms of severity. The fourth and fifth columns display the rating scale (i.e. the CEFR) and its levels ranging from 1 (A1) to 9 (C2).

Table 16 summarises the correspondences found by the panellists between *Communicator* (Speaking) and the CEFR. Although the benchmarking of performance samples was ambiguous as to whether a score of 8 or 9 might better represent a minimally B2 performance. However, a score of 8 would reflect a Band 2 performance averaged across the rating criteria: the level identified as representing B2 in the review of the scales. Conceptually, this would seem to represent the most appropriate cut score for B2.

Table 16 Recommended passing Scores of *Communicator* Speaking and Corresponding CEFR Levels based on benchmarking of materials and performance samples

<i>Communicator</i>	<i>Score</i>	<i>CEFR Level</i>
Fail	1-5	Pre-A1 to A2+
Pass	6	B1
Pass	7	B1+
High Pass	8-12	B2

Writing

Table 17 CEFR scores awarded by panellists to samples of written performance. Figures in parentheses are 'official' results awarded by the Council of Europe, English Profile or by *LanguageCert* examiners as relevant.

<i>Candidate</i>	<i>Observations</i>	<i>Mean*</i>	<i>St. dev.</i>	<i>Mode</i>	<i>CEFR</i>
Script B (B1)	13	1.62	0.87	1	A2
Script 9 (3: Fail)	10	2.20	1.03	2	A2
Script 16 (3.5: Fail)	7	3.14	1.07	4	A2+
Script 11 (3: Fail)	9	3.22	1.56	2	A2+

Script F (B1)	12	3.58	1.68	2	B1
Script 6 (4.5: Fail)	12	4.08	1.24	4	B1
Script 3 (5: Fail)	10	4.40	1.17	4	B1
Script 5 (8: High Pass)	12	4.42	1.16	4	B1
Script 15 (5: Fail)	7	4.71	1.38	5	B1+
Script 1 (6.5: Pass)	8	4.75	1.04	4	B1+
Script 12 (7: Pass)	8	4.75	1.04	4	B1+
Script 2 (6.5: Pass)	8	4.75	1.16	6	B1+
Script 14 (6.5: Pass)	7	4.86	1.46	6	B1+
Script E (B2)	13	5.08	1.04	6	B1+
Script 4 (6: Pass)	10	5.30	1.06	6	B1+
Script 7 (8: High Pass)	10	5.30	0.95	6	B1+
Script 17 (7: Pass)	6	5.33	0.82	6	B1+
Script 13 (7.5: Pass)	8	5.38	0.92	6	B1+
Script A (B2)	13	5.46	0.88	6	B1+
Script 8 (9: High Pass)	12	5.58	0.90	6	B2
Script C (B2)	12	5.67	0.78	6	B2
Script 18 (9: High Pass)	6	5.83	0.41	6	B2
Script 10 (10.5: High Pass)	7	5.86	0.38	6	B2
Script D (C1)	13	7.85	0.55	8	C1

*Note: A1 = 1; A2 = 2; A2+ = 3; B1 = 4; B1+ = 5; B2 = 6; B2+ = 7; C1 = 8, C2 = 9

Table 17 shows the distribution of the scores awarded by the panellists to the scripts they read. In this case, the average scores awarded by the panel correctly identified the CEFR level of three of the six calibrated samples and the rank order of the scores was also consistent with those from the Council of Europe and English Profile. However, the panels seem to have been relatively harsh, scoring two B2 samples as B1+ and one B1 sample as A2. The scores the panellists awarded to the *Communicator* samples generally placed them in the same order as the official scores, although there were a few exceptions. The script with the most discrepant result was Script 5, scored 8 by *LanguageCert*. This was scored as B1 by a majority of panellists, scoring between two scripts with official *Communicator* scores of 5 (Scripts 3 and 15) and below all of the other scripts with official *Communicator* scores greater than 5.

Table 18 Fit Statistics for Writing for rater measurement (in order of rater severity)

Rater ID	Measure	S.E.	Infit Mnsq
L4	1.20	0.28	0.91
L6	0.84	0.27	0.73
L5	0.61	0.25	0.84
L3	0.46	0.27	0.62
L1	0.31	0.27	0.79
A2	0.22	0.26	0.60
A1	0.15	0.26	0.44
A9	0.03	0.29	0.96

A5	-0.13	0.27	0.93
A4	-0.14	0.29	1.03
L2	-0.32	0.30	1.05
A6	-0.33	0.27	0.32
A8	-0.41	0.27	0.35
L7	-0.50	0.30	1.21
A3	-0.95	0.28	1.05
A7	-1.04	0.29	0.96

The results for panellists/ raters are displayed in Table 18. The separation reliability⁴ was 0.79, which indicates that for Writing the panellists did differ significantly ($p < .05$) in their severity. This is reflected in the greater spread of Measure values in Table 18 than in Table 14 with L4, the most severe rater, over two logits harsher than panellist A7, the most lenient. In the case of Writing, the Luton panel was more severe than the Athens panel by an average of .66 logits. In other words, the Athens panel classified more performances as B2 than did the Luton panel. The infit mean squares indicate that all the panellists fit well with the model, with A6 and A8 exhibiting ‘ovefit’.

Matching the Communicator Exam Performances to CEFR Levels

Table 19 Person Measurement Report for Communicator written performance samples and corresponding CEFR Levels

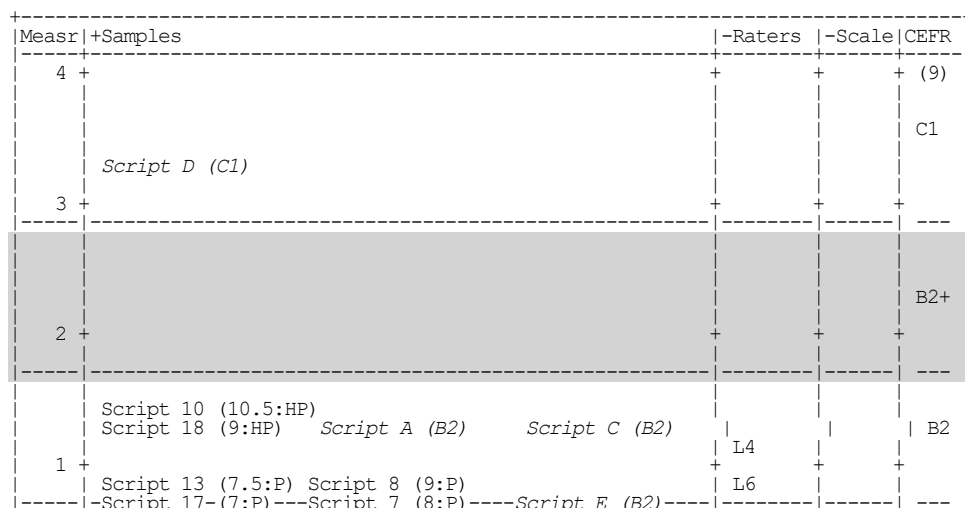
CEFR	Sample ID	Fair Average	Measure	S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd
	Script 9 (3: Fail)	2.22	-2.61	0.36	0.6	-0.92	0.56	-1.07
A2+			-2.21					
	Script 16 (3.5: Fail)	3.32	-1.41	0.38	0.86	-0.12	0.85	-0.14
	Script 11 (3: Fail)	3.36	-1.38	0.37	1.19	0.54	1.16	0.49
B1			-1.23					
	Script F (B1)	4	-0.75	0.44	2.32	2.56	2.31	2.55
	Script 6 (4.5: Fail)	4.2	-0.55	0.3	1.11	0.4	1.11	0.39
	Script 3 (5: Fail)	4.42	-0.33	0.31	0.69	-0.63	0.69	-0.63
B1+			-0.26					
	Script 5 (8: High Pass)	4.54	-0.21	0.29	0.94	-0.02	0.93	-0.03
	Script 1 (6.5: Pass)	4.7	-0.05	0.35	0.34	-1.64	0.34	-1.64
	Script 2 (6.5: Pass)	4.7	-0.05	0.46	1.68	1.3	1.69	1.32
	Script 15 (5: Fail)	4.94	0.18	0.38	0.76	-0.31	0.76	-0.31
	Script 12 (7: Pass)	4.97	0.22	0.35	0.34	-1.66	0.34	-1.66
	Script 14 (6.5: Pass)	5.08	0.33	0.38	0.83	-0.17	0.82	-0.18
	Script 4 (6: Pass)	5.33	0.57	0.32	0.69	-0.64	0.69	-0.64

Script 7 (8: High Pass)	5.43	0.68	0.32	0.4	-1.64	0.39	-1.65
Script 17 (7: Pass)	5.47	0.72	0.41	0.18	-2.07	0.18	-2.07
Script E (B2)	5.5	0.74	0.28	1.01	0.15	1.02	0.18
B2	0.75						
Script 13 (7.5: Pass)	5.63	0.87	0.35	0.19	-2.38	0.19	-2.4
Script 8 (9: High Pass)	5.65	0.90	0.29	0.25	-2.59	0.25	-2.61
Script 18 (9: High Pass)	5.97	1.22	0.41	0.06	-2.99	0.06	-2.99
Script A (B2)	5.99	1.24	0.28	0.74	-0.62	0.74	-0.6
Script C (B2)	5.99	1.24	0.29	0.48	-1.47	0.48	-1.49
Script 10 (10.5: High Pass)	6.15	1.40	0.38	0.51	-0.96	0.52	-0.94
B2+	1.72						
C1	2.80						
Script D (C1)	7.89	3.35	0.32	0.96	0.03	0.93	-0.07

Table 19 shows the FACETS person measurement report for all the *Communicator* written performance samples, with the threshold values for CEFR levels indicated by grey shading. The samples represented three score levels on the *Communicator* speaking tests:

- Fail: These were mostly classified by the panellists as A2 or B1. Script 15, with an official score of 5 was the exception, being classified by the panels as B1+.
- Pass: Pass scores between 6 and 7.5 on the *Communicator* rating scale were mostly classified by the panellists as B1+, although one (Script 13) was considered to be B2.
- High Pass: High Pass grades are not awarded for the Writing section of the Written Exam, but on the Exam as a whole, scores over 66% qualify. This would equate to 8 or more out of 12 on the Writing scale. There were five scripts with *Communicator* Writing scores of 8 or more. The two with scores of 8 were both identified as B1+, but those scoring 9 or more were classified as B2.

Figure 5 Variable map of written performance samples rated by Athens and Luton panels



	Script 4 (6:P)	L5			
	Script 12 (7:P) Script 14 (6.5:P)	L3			
	Script 15 (5:F)	A2	L1		B1+
* 0 *	Script 1 (6.5:P) Script 2 (6.5:P)	A1			
		A9		* CEFR *	
	-----Script 3 (5:F)-----Script 5 (8:F)-----	A4	A5		
		A6	L2		
		A8	L7		
	Script 6 (4.5:F)				B1
-1 +	Script F (B1)				
		A3	A7		
	Script 11 (3:F) Script 16 (3.5:F)				
-2 +					A2+
	Script 9 (3:F)				
-3 +					A2 (1)
Measr	+Samples			-Raters	-Scale CEFR

The results of the MFRM analysis are summarised in Figure 5. Again, the first column, headed *Measr*, represents the logit scale on which all the variables are placed. Higher values represent higher ability on the part of test takers, greater difficulty of test tasks or more severe judgement on the part of the raters.

The second column shows where the performance samples fall on the logit scale. The higher they are placed, the higher the performance was rated. In the case of the samples from the Council of Europe or English Profile samples (Scripts A to E, italicised), these are CEFR levels and in the case of *LanguageCert* (Scripts 1 to 18), overall scores on the *LanguageCert* Writing scale awarded by *LanguageCert* examiners: the sum of the scores for the four criteria of a) *Task Fulfilment*; b) *Accuracy and Range of Grammar*; c) *Accuracy and Range of Vocabulary*, and d) *Text Organisation*. The figures in parentheses show the scores originally awarded to these performances. In Figure 5, Script D (Band 5 on Cambridge FCE and C1 according to the English Profile seminar) was judged the highest-level performance. The lowest was Script 9 (which had been awarded 3 on the *Communicator* scale: 0 for *Accuracy and Range of Grammar*, 1 for each of the other criteria).

The third column shows the relative severity of the raters and shows how far they diverge in terms of severity. The fourth and fifth columns display the rating scale (i.e. the CEFR) and its levels ranging from 1 (A1) to 9 (C2).

Table 20 summarises the correspondences found by the panellists between the *Communicator* Writing test and the CEFR. The identification in the review of the scales of an aggregate score of 8 with B2, reflecting a Band 2 performance averaged across the rating criteria, was again broadly borne out by the benchmarking of performance samples.

Table 20 Recommended passing Scores of Communicator Speaking and Corresponding CEFR Levels based on benchmarking of materials and performance samples

<i>Communicator</i>	<i>Score</i>	<i>CEFR Level</i>
Fail	1-6	Pre-A1 to B1
Pass	6-7	B1+
Pass	8-10	B2
High Pass	11-12	B2+

Bechmarking and Standard setting: Reading

The panellists were asked to judge at which CEFR level test material from three sample *Communicator* tests would be most appropriate (the Calibration test and Versions 417A and 417H). They also judged the difficulty of sample tasks provided by the Council of Europe to exemplify levels B1 and B2 for Reading in English from the British Council Aptis tests (Aptis), the Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE), the Euroexam Centre (Euroexam) and Saint Petersburg State University (SPU). This addressed both Familiarisation with the target level and Standardisation of judgement. Although no official ‘cut scores’ for B2 are provided for these sample tasks, the panellists’ estimates appeared to reflect a good understanding of the target level with scores of around 50% expected of a minimally competent B2 test taker on the Aptis and SPSU (Table 21) and scores of over 90% on the Euroexam material, which exemplifies B1. The BIFIE material was judged to be rather more challenging: B2 level test takers were only expected to score 38% on this task, suggesting that this was interpreted as a B2+ level task.

All the material on the *Communicator* Reading tests was judged by the majority of the sixteen panellists to be most suitable for test takers at the B2 level. However, four panellists considered that Part 1 of the Calibration test was more Suitable for B1 test takers and one for B1+. One panellist identified Part 2 with B1+. Parts 3 and 4 appeared both on the Calibration test and on Test 417A. Three panellists considered that Part 3 was most suitable for B1 and two for B1+, although one considered it to be best suited to B2+. Two believed that Part 4 was best suited to B1+. For Test 417A, three panellists considered Part 1 to be suitable for B1 and two for B1+; three panellists selected B1+ for Part 2. On test 417H, four panellists identified Part 1 with B1 and two with B1+. For Part 2, one chose B1 and four B1+. On Part 3, two selected B1+ and on Part 4, one chose B1. In short, the Reading material was judged to be clearly suitable for B2 level test takers.

Table 21 In responding to this task, how many items would you expect a minimally competent B2 level reader to answer correctly?

<i>Paper</i>	<i>Obs. mean</i>	<i>Est. B2 min.</i>	<i>St. Dev.</i>	<i>K</i>	<i>B2 min %</i>
Aptis (B2)		3.86	0.38	7	55.1%
BIFIE (B2)		3.00	0.00	8	37.5%
Euroexam (B1)		6.71	0.49	7	95.9%
SPSU (B2)		3.71	0.49	8	46.4%
Calibration Part 1	5.06	4.20	0.56	6	70.0%

Calibration Part 2	5.06	3.80	0.68	6	63.3%
Calibration Part 3 ¹	5.12	4.56	0.89	7	65.2%
Calibration Part 4 ²	4.24	4.31	0.95	7	61.6%
VER. 417A Part 1	3.38	4.06	0.68	6	67.7%
VER. 417A Part 2	3.71	3.81	0.89	6	63.5%
VER. 417A Part 3 ¹	4.31	4.56	0.89	7	65.2%
VER. 417A Part 4 ²	3.57	4.31	0.95	7	61.6%
VER. 417H Part 1	4.11	3.81	0.75	6	63.5%
VER. 417H Part 2	3.95	3.63	0.62	6	60.4%
VER. 417H Part 3	5.55	4.81	0.75	7	68.8%
VER. 417H Part 4	4.45	4.28	0.45	7	61.2%
Calibration	19.47	16.88	1.86	26	64.9%
417A	14.97	16.75	2.91	26	64.4%
417H	18.06	16.53	1.96	26	63.6%

^{1,2}Note that Calibration Parts 3 and 4 are the same tasks as VER 417A Parts 3 and 4.

The panellists were asked:

In responding to this task, how many items would you expect a minimally competent B2 level reader to answer correctly?

The results are displayed in Table 21. The column headed *Obs. mean* shows the mean scores on each test section and for the Reading test as a whole when the material was piloted. As each test was piloted on a different population, this does not reflect the relative difficulty of the three test forms, but does indicate the relative difficulty of the parts within each test. As two tasks from the Calibration test (Parts 3 and 4) were used in VER. 417A, it is also possible to observe that the group taking the latter test appears to have been less proficient at Reading than the group taking the Calibration test.

The panel as a whole thus considered that a minimally competent B2 level performance on any of these three *Communicator* Reading test forms would be 17 points out of a possible 26 points. The Athens panel tended to be marginally more lenient than the Luton panel (Calibration: 16.89, 417A: 16.18, 417H: 16.17 compared with Calibration: 16.67, 417A: 17.57, 417H: 17.00), although the differences between them would only have affected the recommendation for Test 417A, moving it up to 18. This compares with the recommended cut score of 15.30 out of a possible 30 points, rounded down to 15 in O'Sullivan (2008).

Bechmarking and Standard setting: Listening

The panellists were asked to judge at which CEFR level test material from the Calibration, 417A and 417H *Communicator* test forms would be most appropriate. They also judged the difficulty of sample tasks provided by the Council of Europe to exemplify levels B1 and B2 for Listening in English from the Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE), Cambridge Assessment English (Cambridge), the Matriculation Examination, Finland (Finnish Matr.) and Saint Petersburg State University (SPSU). This addressed both Familiarisation with the target level and Standardisation of judgement. Although no official 'cut scores' for B2 are provided for these sample tasks, the panellists' estimates seemed to reflect a good understanding of the target level with scores of

around 50% expected of a minimally competent B2 test taker on the BIFIE and SPSU material exemplifying B2 and scores of over 90% expected on the Cambridge material exemplifying B1. Although the Finnish Matr. material was intended to exemplify B1/B2, the panel estimated that it was equally challenging as the BIFIE and SPSU tasks: B2 level test takers were expected to score 51% on this task (Table 22).

All the material on the *Communicator* Reading tests was estimated by the majority of panellists to be most suitable for test takers at the B2 level. The only disagreement expressed was for the Calibration test and Version 417A. Three panellists considered that Part 3 (the same task on both tests) was most suitable for B1+. One panellist considered that Part 1 on both tests and two panellists that Part 2 of 417A were more suitable for B1+ test takers. The Listening material, even more clearly than the Reading, was judged to be suitable for B2 level test takers.

Table 22 In responding to this task, how many items would you expect a minimally competent B2 level reader to answer correctly?

Paper	Obs. mean	Est. B2 min.	St. Dev.	K	B2 min %
BIFIE (B2)		4.08	1.50	8	51.0%
Cambridge (B1)		5.43	0.81	6	90.5%
Finnish Matr. (B1/B2)		2.57	0.66	5	51.3%
SPSU (B2)		4.03	0.97	8	50.3%
Calibration Part 1	5.66	5.13	0.57	6	85.4%
Calibration Part 2	4.60	3.81	0.65	7	54.5%
Calibration Part 3 ¹	5.32	4.50	0.50	6	75.0%
Calibration Part 4	4.68	3.94	1.11	7	56.3%
VER. 417A Part 1	4.90	5.00	0.63	7	71.4%
VER. 417A Part 2	4.10	3.91	0.70	6	65.2%
VER. 417A Part 3 ¹	4.16	4.50	0.50	7	64.3%
VER. 417A Part 4	3.68	3.40	0.70	6	56.7%
VER. 417H Part 1	5.42	4.92	0.67	7	70.2%
VER. 417H Part 2	3.85	3.83	0.58	6	63.9%
VER. 417H Part 3	3.38	3.75	0.75	7	53.6%
VER. 417H Part 4	4.78	3.58	0.67	6	59.7%
Calibration	20.26	17.38	3.73	26	66.8%
417A	16.84	16.81	1.05	26	64.7%
417H	17.43	16.08	0.97	26	61.9%

^{1,2}Note that Calibration Part 3 is the same task as VER 417A Part 3.

Again, the panellists were asked:

In responding to this task, how many items would you expect a minimally competent B2 level reader to answer correctly?

The results are displayed in Table 22. The column headed *Obs. mean* shows the mean scores on each test section and for the Reading test as a whole when the material was piloted. As with Reading, each test was piloted on a different population, so the mean scores do not reflect the relative difficulty of the three test forms, but do indicate the relative difficulty of the parts within each test. As one of the tasks from the Calibration test (Part 3) was also

used in VER. 417A, it is also possible to observe that the group taking the latter test appears to have been less proficient at Listening than the group taking the Calibration test.

The panel as a whole thus considered that a minimally competent B2 level performance on any of these three *Communicator* Listening test forms would be 17 or 18 points out of 26. Again, the Luton panel was a little harsher than the Athens panel, placing the minimum B2 standard 1 or 1.5 points higher: their recommendation would be between 17 and 19 out of 26 (Calibration: 18.29, VER 417A: 17.31, VER 417H: 16.71) against 16 or 17 (Calibration: 16.67, VER 417A:16.36, VER 417H: 15.20) for the Athens panel.

This compares with the recommended cut score of 14.70 out of 30, rounded up to 15 in O’Sullivan (2008). However, the O’Sullivan (2008) recommendation was based on the judgements of just 9 of the 11 judges on the City & Guilds panel: the two harshest judges (the only two on the panel that were external to the City & Guilds programme) were dropped from the analysis on the grounds that they were distant from consensus of the rest of the group. The average across all 11 City & Guilds panellists was 17.36, which could have been interpreted as a recommended minimum of 18 points out of 30.

Procedural Validity: the post-panel questionnaire

The Council of Europe (2009) Manual recommends asking panellists for feedback on whether the training provided had been helpful, whether they felt they had a clear understanding of the purpose of the meeting and on how they had found the arrangements. Immediately following the panel meetings in Athens and Luton, the panellists were asked to complete a brief questionnaire on the procedures and facilities. All eighteen panellists responded to all of the questions. The results of this questionnaire are shown in Table 11 below.

Table 23 Post-panel questionnaire for panellists

		4 - Strongly Agree	3 - Agree	2 - Disagree	1 - Strongly Disagree
1	I understood how to answer the pre-seminar online questionnaire (assigning descriptors to levels)	13	2		
2	I understood the purpose of the seminar	14			
3	I feel I now have a clear understanding of the CEFR descriptive scheme	10	5		
4	I feel I now have a clear understanding of the CEFR common reference levels	10	4		
5	I understood the instructions for the activities	15			
6	There was adequate time for discussion	10	6		
7	All participants were allowed to state their views freely	15			
8	The introductory discussions were helpful to me	15			

9	The discussions during the seminar were helpful to me	15			
10	The facilities at the conference centre helped to create a positive working environment	14	1		

Encouragingly, all fifteen of the panellists who responded agreed with all of the statements, suggesting that they were satisfied that the seminar and associated activities had been successful. Only five items did not receive unanimous “strongly agree” responses. These included understanding of the CEFR descriptive scheme and common reference levels (items 4 and 5) as well as the familiarisation questionnaire (“I understood how to answer the pre-seminar online questionnaire (assigning descriptors to levels”). In relation to the seminar itself, six panellists selected “agree” rather than “strongly agree” in response to the item, “There was adequate time for discussion”. This perhaps reflects the tight timescale for the meeting and the restrictions this imposed.

Finally, this report was circulated to the panellists, who confirmed that it accurately reflects the standard setting procedures that they went through.

Conclusions

The panels agreed that material throughout the Spoken and Written Exams reflected the B2 level. They particularly identified the Listening paper with the CEFR illustrative scales for *Understanding conversation between other speakers*, *Listening to the radio and audio recordings*, *Listening to announcements and instructions* and *Listening as a member of a live audience* as well as *Overall listening comprehension*. The Reading paper was associated with *Reading for information and argument* and *Reading for orientation*. Band 2 on both the Speaking and Writing papers reflected the B2 level across the rating criteria. The exception was *Pronunciation, intonation and fluency* on the *Communicator* Speaking scale which corresponds to the B1 level, suggesting that this scale could be re-evaluated in the light of the updated CEFR scales for *Phonological control*. The panels also noted unexpectedly little evidence of B1 level descriptors in the Band 1 level of *Communicator*.

In relation to the Common Reference Levels of the CEFR, the evidence from Specification, Benchmarking and Standard setting suggests that scores of 17 out of 26 for the Reading and Listening papers and 8 out of 12 for Writing and Speaking would better reflect the B2 level than the current Pass scores. Confidence in this conclusion is limited by the size of our sample and a few anomalous results. More stable estimates of the CEFR levels for Writing and Speaking will be obtained if further estimates of the level of *Communicator* performance samples are collected through routine rater training. The methodology used here has potential applications in the cross-validation of this and other linking projects as it allows for the inclusion of samples from other testing programmes where these become available.

O’Sullivan (2008) observed that linking a test to an external criterion, such as the CEFR, should not be regarded as a one-off activity and steps including enhanced test

specifications and training for test writers and examiners were introduced by City & Guilds to ensure that the relationship was maintained over time. Given the high stakes of the *Communicator* test, it is essential that the nature of the relationship should continue to be a focus. The links to the CEFR set out in this report should be confirmed and extended through procedures embedded in routine test production. It is also suggested that the results should, where possible, be validated through comparisons with other measures such as other tests or teacher judgements and that further dedicated linking exercises should be conducted at regular intervals.

References

- Breton, G., Lepage, S. & North, B. (2008). Cross-language benchmarking seminar to calibrate examples of spoken production in English, French, German, Italian and Spanish with regard to the six levels of the Common European Framework of Reference for Languages (CEFR): CIEP, Sèvres, 23-25 June 2008. Retrieved 15, August 2016 from rm.coe.int
- Brunfaut, T. & Harding, L. (2014) Linking the GEPT listening test to the Common European Framework of Reference. LTTC-GEPT Research Reports RG-05. Taipei: Language Training and Testing Centre.
- CiEP: Centre international d'études pédagogiques. (2008). Spoken Performances Illustrating the 6 levels of the CEFR, Paris: Centre international d'études pédagogiques.
- Council of Europe. (2001). A Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press.
- Council of Europe. (2003). Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) – Preliminary pilot version. DGIV/EDU/LANG (2003) 5, Strasbourg: Council of Europe.
- Council of Europe. (2009). Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. A manual. Strasbourg: Council of Europe.
- Council of Europe (2018). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors. Strasbourg: Council of Europe.
- Ffrench, A. (n.d.). DVD "Spoken performances illustrating the 6 levels of the Common European Framework of Reference for Languages": Comments on the assigned levels in English. Retrieved November 3, 2016, from www.ciep.fr
- Green, A.B. (2012). Conducting CEFR Benchmarking for the Versant English Placement Test: Final Project Report. NCS Pearson. Downloaded 6 June 2013 from www.pearson.com.au/educator/english-language/versant-englishplacement-test.
- Green, A.B. (2018). Linking tests of english for academic purposes to the CEFR: The score user's perspective. *Language Assessment Quarterly*, 15(1), 59-74.
- Green, A.B. & Inoue, C. (2017). GEPT Speaking - CEFR Benchmarking, LTTC-GEPT Research Report No. RG-09. Taipei: LTTC.
- Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24(4), 355-366.

- Knoch, U. & Frost, K. (2016). Linking the GEPT Writing Sub-test to the Common European Framework of Reference (CEFR). LTTC-GEPT Research Reports RG-08. Taipei: Language Training and Testing Centre.
- Linacre, J. M. (2013). A user's guide to FACETS: Rasch-model computer programs. Retrieved August 15, 2016, from <http://www.winsteps.com/manuals.htm>
- Linacre, J. M. (2015) Facets computer program for many-facet Rasch measurement, version 3.71.4. Beaverton, Oregon: Winsteps.com
- North, B. & Jones, N. (2009). Further Material on Maintaining Standards across Languages, Contexts and Administrations by exploiting Teacher Judgment and IRT Scaling. Retrieved 30 August, 2016 from http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions, 8(3), 370.

Appendix

LanguageCert IESOL Communicator B2 Specification

Table of Contents

A. The purpose and use of LanguageCert International ESOL Communicator Level	47
1. Overview of LanguageCert International ESOL Communicator Level	47
B. Producing LanguageCert International ESOL	50
2. Development of LanguageCert International ESOL	50
3. Writing LanguageCert International ESOL	51
C. The content of LanguageCert International ESOL – Communicator B2.....	53
4. Summary of the content of LanguageCert International ESOL – Communicator B2..	53
5. Information on Assessment Tasks	59
6. Scoring Assessment Tasks	97
7. Reporting scores on LanguageCert International ESOL as a whole.....	104
8. Assessment Results and Analysis.....	106

A. The purpose and use of LanguageCert International ESOL Communicator Level

Overview of LanguageCert International ESOL Communicator Level

LanguageCert International ESOL Communicator represents one of six levels of the LanguageCert International ESOL Qualifications: a range of qualifications that offer a communicative approach to the testing of Listening, Reading, Writing and Speaking.

Separate certificates are issued for

- LanguageCert International ESOL Communicator (Listening, Reading, Writing)
- LanguageCert International ESOL (Speaking). LanguageCert International ESOL (Speaking)

Both reflect the B2 Vantage level of the Common European Framework of Reference (CEFR). In this document, for the sake of brevity, the name LanguageCert International ESOL Communicator is used to refer to both LanguageCert International ESOL (Speaking) Communicator and LanguageCert International ESOL (Listening, Reading, Writing) Communicator.

LanguageCert International ESOL Communicator is an international assessment taken by teenagers aged 14 or over and adults and young adults, who are not native speakers of English and who wish to achieve a high quality, internationally recognised qualification in English that is available and recognised worldwide.

The people expected to take **LanguageCert International ESOL Communicator** are:

- non-native speakers of English worldwide;
- young people or adults attending an English language course either in the UK or overseas;
- students learning English as part of their school or college curriculum;
- people needing English for their everyday or working life;
- learners who require externally recognised certification of their command of the English language;
- those who are attending English language courses over a period and require a series of graded examinations which provide steps in the ladder of English language proficiency;
- learners attending short English language courses.

LanguageCert International ESOL Communicator provides evidence of a person taking the test's ability to speak, write and understand verbal and written English at CEFR level B2.

LanguageCert International ESOL Communicator is primarily concerned with the personal and public domains.

It involves the following CEFR activities: Spoken Reception (Listening), Written Reception (Reading), and Written Production (Writing) in English. LanguageCert International ESOL Communicator (Speaking) involves Spoken Production and Spoken Interaction (Speaking) in English.

LanguageCert International ESOL Communicator relates to a wide range of contexts for language use judged to be relevant to the test takers.

LanguageCert International ESOL Communicator aims to encourage people preparing for the test to learn to communicate in the real world. This is supported by the communicative approach adopted in which all four language skills are tested. All test materials are based on the concept of communication and interactional authenticity. This approach is coherent across the whole test development process.

LanguageCert International ESOL is owned and administered by LanguageCert. LanguageCert is an Awarding Organisation dedicated to language skills assessment and certification. It is a member of PeopleCert, a global leader in the certification industry that has delivered millions of exams in 200 countries.

LanguageCert International ESOL Communicator results should be used by individuals in a variety of circumstances, such as;

- by language learners and teachers to set individual language learning goals;
- by universities or other higher or further education institutions to choose students with adequate language abilities for entry to degree courses taught in English;
- by teachers who need B2 level to work in public schools (in Spain, for example) or other professionals who need to demonstrate proof of English language competence;
- by prospective employees as proof of being able to communicate at B2 level;
- by employers to select workers with suitable language skills for customer-facing roles which require knowledge of English.

The examination aims to demonstrate that individuals certified with a LanguageCert International ESOL qualification will be able to communicate in English effectively – at B2 level – regardless of geographical borders and local standards. The examination does not directly relate to specific contexts of use, but to a variety of communicative circumstances.

LanguageCert International ESOL Communicator results are recognised by the following organisations and authorities:

- **England** – Ofqual as a valid proof of English language ability at B2 level for employers in UK; a pre-requisite for UK qualifications that require evidence of suitable English language ability.
- **Wales** – Qualifications Wales.
- **Czech Republic** – Ministry of Education In the STANDARDISED LANGUAGE EXAMINATIONS list and MŠMT “MATURITA” list, as valid proof of English language ability for state/governmental employment, depending on the position applied for.
- **Greece** – ASEP (Supreme Council for Civil Personnel Selection) As a prerequisite for access to opportunities for employment in the Public Sector for posts requiring English language ability at B2 level. They may also be accepted for admission to programmes at the Public Universities and Institutes of Higher Education across the country, as well as for employment at private organisations.
- **Hungary** – Office of the Language Testing Accreditation Center.
*Please note that in Hungary the exam format is slightly different; in the two separate exams, the skills examined are: Listening together with Speaking, and Reading together with Writing.
- **Italy** – Ministry of Education, Universities and Research B2 level “Communicator” qualification is accepted for PON scholarships.
- **New Zealand** – recognised by the New Zealand Qualifications Authority (NZQA).accepted for admission for Certificates and Diplomas of lower levels of New Zealand’s education system.
- **Romania** – recognised by the Romanian Ministry of National Education and Scientific Research.
- **Spain** – recognised by ACLES (Association of Language Centres in Higher Education) as proof of B2 level ability, as proof of linguistic competence by teachers and for admission to universities.
- **Andalusia** – by Junta de Andalucía, the official authority responsible for the accreditation of CEFR mapped foreign language qualifications for teachers of bilingual education.
Aragon – recognised by the Ministry of Education, Culture and Sports of Aragon.
Navarra – recognised by the Department of Education of the Government of Navarra.

To find out more about organisations that recognize the results please visit

<http://www.languagecert.org/en/Recognitions/languagecert-international-esol/Pages/Recognitions-LanguageCert-International-ESOL-Qualifications.aspx>

B. Producing LanguageCert International ESOL

Development of LanguageCert International ESOL

The LanguageCert International ESOL Communicator examination was developed by LanguageCert. The examination was heavily based on the City & Guilds IESOL qualification, which was acquired by LanguageCert in May 2015 (see Section 1 for information about LanguageCert).

The content of LanguageCert International ESOL Communicator is also based on a theory of communicative language ability adapted from models of communicative language ability (CLA) such as Bachman & Palmer (1990) and the foreign language specifications provided by the Council of Europe in the B2 Vantage specification .

The key sources for the content of LanguageCert International ESOL are the following:

- The acquired City & Guilds exam specifications, on which the LanguageCert International ESOL was originally based;
- Models of communicative language ability (CLA) such as Bachman & Palmer (1990);
- The findings of ongoing Stakeholder engagement projects, including feedback from people taking the test, test centres, and other relevant users;
- The Common European Framework of Reference;
- Vantage (Ek, J. and Trim, J. (2007). *Vantage*. Cambridge: Cambridge University Press.) for tasks, grammar, lexicon, functions and registers.

The above were considered in depth during a major revision of the examination undertaken by City & Guilds in 2014 and also during an exhaustive *Vetting and Editing Project* conducted by LanguageCert upon acquisition of the qualification, relevant materials and intellectual property in 2015-16.

To find out more about how LanguageCert International ESOL Communicator was developed and to find a detailed list of topics included in the syllabus, you can read the **Qualification Handbooks** also available online ([here](#)).

Writing LanguageCert International ESOL

Test writers for LanguageCert International ESOL are selected based on their qualifications and extensive experience. Invitations for training are sent to professionals with a minimum of the following required skills and qualifications:

- Bachelor's degree in English Language or similar field;
- A prestigious Teaching Qualification is desirable;
- Minimum 3 years' experience in the ELT sector;
- Demonstrable materials production experience;
- Working knowledge of Common European Framework for Languages;
- Exceptional command of the Assessment Language. Excellent command of the English Language (C2);
- Being well-organized, self-driven, flexible and able to work in a high-volume, fast-paced, and deadline-driven environment.

Test writers for LanguageCert International ESOL are given training by expert senior consultants who are highly experienced in English language assessment and more specifically in the development, production and delivery of English examinations and teaching qualifications. Initial training usually lasts three days and writers are required to attend training twice a year.

To ensure that the content follows the developers' intentions and to ensure that it is parallel across different versions of LanguageCert International ESOL Communicator exams, when preparing material, the writers follow Item Writer Guides. These contain detailed instructions and exhaustive task specifications (information on task and text type, task focus, question requirements, number of items per task, assets, distractors, etc.).

Assessment material is reviewed by subject matter experts: exceptionally experienced assessment professionals (externally commissioned professional item writers). The LanguageCert associate in charge of the coordination of commissioning performs a second check, before delivering the items to an editor who makes sure that items and tasks perform and render in accordance with specifications. Editing meetings are also arranged twice a year, in which item writers are invited to participate.

In designing and developing tasks for use in each assessment, LanguageCert puts fairness to people taking the test at the very heart of all its work by producing materials which will not favour or discriminate against any people taking the test. This includes ensuring test materials are as free from specific regional or national cultures as possible, and that topics are universal. A list of potentially sensitive (e.g. death, natural disasters), culturally sensitive (e.g. religion, alcohol) and upsetting (e.g. crime, accident) topics is provided to item writers to aid in this. LanguageCert's Language Assessment Development

Manager directs the item development process and is responsible for accepting submitted items and signing them off for pretesting and/or use in live exams.

Before being used in live exam sessions, all **materials are** tested on groups of language learners. Pretesting is carried out at LanguageCert examination centres and educational institutions. The aim is to pre-test each test with a representative sample of people taking the test. This naturally depends on availability, but the suitability of the sample is always checked by the Language Assessment Development Manager. LanguageCert strives to ensure adequate sample sizes for pre-testing. Typically, each test is pre-tested on a sample of at least 100 people taking the test and LanguageCert ensures that the demographics of the sample are carefully considered, in order to have both a sufficiently random sample and one that fits the characteristics of the test takers. We make sure that relatively equal numbers of males and females participate in the pre-testing and that the age range matches that of prospective test takers. If 100 pre-test people taking the test cannot be found, final calibration is performed during live testing.

To find out more about how LanguageCert International ESOL Communicator is prepared, you are welcome to attend one of the training webinars regularly offered by LanguageCert as part of the Academic Support Services. (<http://www.languagecert.org/en/Qualifications/Academic-Support-Services/Pages/Academic-Support-Services.aspx>)

The content of LanguageCert International ESOL – Communicator B2

Summary of the content of LanguageCert International ESOL – Communicator B2

i. LanguageCert International ESOL Communicator (Listening, Reading, Writing)

(examination duration: listening: approximately **30 minutes**; reading and writing: **2 hours and 10 minutes**)

Skill and Focus	Task	Format	Marks
Listening 1: understand context, meaning and function in short conversations on concrete and abstract topics	The person taking the test listens twice to seven unfinished conversations between two speakers, one being a formal conversation.	One three-option multiple choice for each conversation to choose the appropriate completion or continuation of the conversation	7
Listening 2: identify a specific aspect of a conversation	The person taking the test listens twice to three conversations with two speakers to identify: topic, purpose, context, speakers, gist, relationships between speakers, roles, functions, attitudes, feeling and opinions.	Two three-option multiple choice questions for each of the three conversations	6
Listening 3: extract key information from a monologue to complete a task	The person taking the test listens twice to a radio broadcast, narrative, presentation, etc to identify specific information.	A note or message pad with seven headings and space to write the correct information for each heading	7
Listening 4: follow a discussion between two speakers	The person taking the test listens to a discussion twice and identifies gist, examples, fact, opinion, contrast, purpose, key ideas, attitude, cause and effect.	Six three-option multiple choice questions	6

As per the table above the maximum number of marks available for the Listening section is 26 (one mark per question). Marks are converted into a scaled score out of 50 (total scaled score: 150) to ensure equal weighing of all skills.

Skill and Focus	Task	Format	Marks
Reading 1: understand in detail information, ideas and opinions	The person taking the test reads one long text of 450-500 words: a news story, article, review or proposal.	Six three-option multiple choice questions	6
Reading 2: understand how meaning is built up in a text	The person taking the test reads a text of 340-400 words with six sentences removed, e.g. a topic sentence, summarising sentence, developing idea, emphasising a point, opinion, contrast, sequence, forward and back reference, transition to a new idea.	Choice of seven sentences to choose from in order to complete the six gaps, one being a distractor	6
Reading 3: understand the purpose of text and locate specific information and awareness of writers' stance and attitude	The person taking the test reads four texts (380-420 words in total), e.g. an email, article, advert, brochure, etc. with a linked theme, but with a different purpose.	Seven multiple matching questions to identify information from the text	7
Reading 4: understand specific information through detailed reading	The person taking the test reads a continuous text of 450-500 words: a narrative, descriptive, explanatory, expository, biographical, instructive text.	Seven open-ended 'wh'-questions requiring short answers of up to five words.	7

As per the table above the maximum number of marks available for the Reading section is 26 (one mark per question). Marks are converted into a scaled score out of 50 (total scaled score: 150) to ensure equal weighing of all skills.

Skill and Focus	Task	Format	Marks
Writing 1: respond appropriately to a given text to produce a formal response for an intended public audience	Instructions are given to the person taking the test to respond formally using a written input with three content points to be addressed, while the intended reader is specified.	100 to 150-word written composition	12
Writing 2: produce a personal letter, a narrative or descriptive composition	Instructions are given to write an informal piece of writing for a specified reader on a general subject not requiring specialist knowledge with two content points to be addressed.	150 to 200-word written composition	12

As per the table above the maximum number of marks available for the Writing section is 24, a total of 12 awarded for each task. Marks are converted into a scaled score out of 50 (total scaled score: 150) to ensure equal weighing of all skills.

The maximum number of marks available (76 marks) for the Listening, Reading and Writing sections are converted into a scaled score out of 50 per skill (total scaled score: 150) to ensure equal weighing of all skills. The pass mark is 50% (75 out of 150). For a more detailed description of assessment arrangements, please refer to section 6 of this document.

ii. LanguageCert International ESOL (Speaking). LanguageCert International ESOL (Speaking)

Skill & Focus	Task	Format	Marks
<p>Speaking Part 1: To communicate personal information; to give personal information and to express opinions and ideas</p>	<p>Give and spell name Give country of origin Answer five questions</p>	<p>One-to-one interview between an interlocutor and person taking the test.</p>	<p>LanguageCert International ESOL (Speaking) is marked holistically. Marks are awarded for performance on all 4 Parts of the Speaking test – a marker awards marks on four assessment criteria on a scale from 0 to 3 to give a total out of 12.</p>
<p>Speaking Part 2: to communicate in real-life situations using a range of functional language to elicit or respond as appropriate</p>	<p>Two or three situations are presented by the interlocutor at each level and people taking the test are required to respond to and initiate interactions.</p>		
<p>Speaking Part 3: to co-operate to reach agreement/decision</p>	<p>Hold a short discussion to make a plan, arrange or decide something using written text as the prompt</p>		
<p>Speaking Part 4: to narrate, describe communicate ideas and express opinions</p>	<p>After 30 seconds of preparation time, talk about a topic provided by the interlocutor and answer follow-up questions for 2 minutes</p>		

The LanguageCert International ESOL (Speaking) Communicator B2 exam offers a comprehensive test of spoken English which can either be a stand-alone examination or a complement to the International ESOL (Reading, Writing and Listening) examination. The tasks in the examinations are designed to test the use of English in real-life situations.

The exam consists of a one-to-one interview between one candidate and an interlocutor. The interlocutor is responsible for delivering, conducting and recording the spoken interview, which is assessed by a trained marking examiner at a later stage.

The maximum marks awarded for the Speaking test is 12 (three per assessment criterion). Marks are then converted into a scaled score out of 50 (please see Section 6 of this document for a more detailed description) to ensure equal weighing of all skills. The pass mark is 50% (25 out of 50).

Scores for LanguageCert International ESOL Communicator (LRW) and LanguageCert International ESOL (Speaking) are not combined, as they are taken as separate exams. A separate certificate is issued for LanguageCert International ESOL Communicator (LRW) and for LanguageCert International ESOL (Speaking).

Information on Assessment Tasks

Relating language assessments to the Common European Framework of References for languages: Describing your assessment.

LanguageCert International ESOL (LRW) Communicator B2

B2 Listening

Part 1

I. Rationale

This Listening part 1 of LanguageCert ESOL Communicator B2 involves seven recorded dialogues, each consisting of three turns, played twice. After hearing each dialogue for the second time, the person taking the test is asked to choose one out of a three-option multiple choice, in order to complete the last part (turn) of the dialogue. Each dialogue is in spoken English, with elided forms, idioms and designed in a way that reflects an authentic conversation in real-life situations.

The instructions are given in English in written mode.

The person taking the test taking the Listening Part 1 has ten seconds after hearing each item twice, to check and complete their answers before they move on to the next part.

At the end of the exam they are given ten minutes to transfer their answers to the answer sheets.

This task is similar to real-life exchange at work or in personal life.

II. What the person being assessed reads/listens to/ sees (the input)

The input may include informal contexts such as conversations between friends; neutral contexts such as conversations between strangers, or a variety of others but also formal, all accessible to a varied age-range. Item writers adapt authentic dialogues or write authentic-style exchanges that a person taking the test could be engaged in. Dialogues are written in spoken English, incorporating natural intonation features as well as colloquial English, natural responses and interjections.

The input mainly relates to the personal and public domain.

Communication themes may include personal identification, daily life, free time, relations with other people, food and drink, shopping, places and weather, entertainment. The speakers' relationships could be teacher-student, schoolmates, work colleagues or a variety of others.

The input is designed specifically for Part 1, taking all specifications into account, ensuring themes and content is on level and that level appropriacy is maintained throughout.

Spoken input is read from a script, recorded on the premises under the presence and supervision of the Academic Associate responsible for the material development and production, in case of last-minute corrections and to ensure accurate execution of the script. The actors recording the materials are highly experienced in working on educational materials and they are all native speakers of English.

To make material suitable for the LanguageCert ESOL Communicator B2, changes might be made, such as rewriting items that might become outdated, re-recording might also take place, phrases might get simplified in case they are too difficult for the level and certain parts might be rephrased if they do not seem to work while recording (e.g. do not sound natural).

The input is mostly concrete and familiar; issues one might encounter at work, at a doctor's appointment, with friends, with family etc.

The input is about six minutes long for part 1 including repetition and pauses. Each dialogue is up to 25 words long.

The vocabulary of the input involves a range of everyday life transactions, some professional exchanges and some more formal, understanding gist, context, purpose and key ideas and distinguishing between fact and opinion.

The grammar of the input involves a wide range of structures and sentence patterns according to B2 level of the CEFR.

The delivery is at a moderate, natural rate.

With standard UK English and US English accents (currently, UK English predominates).

There is no distortion or background noise on the recording. The recording involves two speakers, a male and a female.

People being assessed hear the recording twice.

The input is likely to be comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do (the expected response).

The response involves choosing one out of three written multiple-choice options. Ten seconds are allowed before the second hearing of the same dialogue.

Responses are expected to be of 30 words maximum in total.

Choosing the correct response involves understanding narratives, sequences, instructions, descriptions and explanations on abstract and concrete topics.

The main purpose of the response, in most cases, is either emotive (it involves recognising the emotional state of the speaker) or metalingual (verifies understanding) and can either complete or continue the dialogue.

In Listening part 1 of the LanguageCert ESOL Communicator B2, people are expected to draw on their personal, social, academic and vocational life.

Example of Part 1

Listening Part 1

You will hear some short conversations. You will hear each conversation twice. Choose the correct answer to complete each conversation.

Number one. Number one.

M: What's the problem?

F: I can't decide where to go on holiday.

M: Why don't you go where you go every year?

- a) I didn't enjoy it so much last year.
- b) I've never thought of that before.
- c) I didn't know what you meant.

B2 Listening

Part 2

I. Rationale

This Listening part 2 of LanguageCert ESOL Communicator B2 involves three recorded dialogues that are heard twice. The person taking the test is presented with two three-option multiple-choice questions for each dialogue, six questions in total for all three dialogues. Each dialogue is in spoken English, with elided forms, idioms and is designed in this way to reflect authentic conversation in real-life situations.

The task is similar to real-life situations in which a learner needs to understand conversations on concrete and familiar topics; issues they might encounter at work, at a doctor's appointment, with friends, with family or similar circumstances.

The instructions are given in English in written mode.

The person taking the LanguageCert ESOL Communicator B2 to complete Listening part 2, has ten seconds after each conversation is heard for the second time, to choose one of the three options before they move on to the next conversation. At the end of all three conversations, there is a further 10-second pause. People taking the test are free to move between the different parts of the test and allocate test time as they wish. At the end of the exam they are given ten minutes to transfer their answers to the answer sheets.

II. What the person being assessed reads/listens to/sees (the input)

The input may include a variety of contexts, such as informal or neutral, but also formal, all accessible to a varied age-range. Dialogues are in spoken English, incorporating intonation features as well as colloquial English, natural responses and interjections. Dialogues portray different relationships between the speakers.

The input mainly relates to the personal and public domain.

Communication themes may include personal identification, daily life, free time, relations with other people, food and drink, shopping, places and weather, entertainment. The speakers' relationships could be teacher-student, schoolmates, work colleagues or a variety of others.

The input is designed for this specific part of the Listening test, with writers following the specifications in the Item Writer Guides, ensuring themes and content are suitable for the B2 level.

Spoken input is read from a script, recorded on the premises under the presence and supervision of the Academic Associate responsible for the material development and production, in case of last-minute corrections and to ensure accurate execution of the script.

The actors recording the materials are highly experienced in working on educational materials and they are all native speakers of English.

To make material suitable for the LanguageCert ESOL Communicator B2, changes might be made, such as rewriting items that might become outdated, re-recording might also take place, phrases might get simplified in case they are too difficult for the level and certain parts might be rephrased if they do not seem to work while recording (e.g. do not sound natural).

The input is mostly concrete and familiar; issues one might encounter at work, at a doctor's appointment, with friends, with family etc.

The input is about six minutes long for Part 2, including the repetition and pauses.

The vocabulary of the input involves a range of everyday life transactions, some professional exchanges and some more formal, understanding gist, context, purpose and key ideas and distinguishing between fact and opinion.

The grammar of the input involves a range of structures and sentence patterns according to B2 Vantage level of the CEFR.

The delivery is at a moderate, natural rate.

With standard UK English and some US English accents.

There is no distortion or background noise used on the recordings.

At B2 level, Listening Part 2 involves two speakers, a male and a female.

The input is likely to be comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do (the expected response).

The response involves choosing one out of three multiple-choice responses. Ten seconds are allowed before the second hearing of the same dialogue.

Responses are required to be of 35 words maximum in total, and the dialogues 130-170 words long each.

The main rhetorical functions expected are: narratives, sequences, instructions, descriptions and explanations on abstract and concrete topics encountered in personal, social, academic and vocational life.

The main purpose of the response is either emotive (it involves recognising the emotional state of the speaker) or metalingual (verifies understanding) and people taking the test are asked to identify response to written input.

The focus of the items is on understanding of the context, gist, attitude, topic, purpose, feelings and opinions, role of speaker and functions.

Example of Part 2

Listening Part 2

You will hear some conversations. You will hear each conversation twice. Choose the correct answers for each conversation.

Conversation One

- M: What do you think of Noel's new girlfriend, Jenny?
- F: She seems very nice, John, and they certainly seem to get on well together.
- M: I'm just a bit worried that they don't seem to spend much time together. They don't really have the same interests at all.
- F: Well in some ways don't you think that's a good thing? They like each other but they also have their own lives.
- M: Yes I suppose you're right. With his last girlfriend they always did the same things together, they were never apart.
- F: And they always did what she wanted, never what he wanted. Eventually it was just too much for him.
- M: Yes, perhaps you're right. It's better if two people are a little bit independent. Sometimes it's good to have a break and do something on your own, and then come back together.
- F: I'm sure they'll be fine. But it's early days yet.

Conversation 1

1. The woman thinks Noel's ex-girlfriend
- a) spent too little time with him.
 - b) always agreed with him.
 - c) made all the couple's plans.
2. How do the speakers feel about Noel's new relationship? They're
- a) hopeful.
 - b) worried.
 - c) excited.

B2 Listening

Part 3

I. Rationale

This Listening part 3 of LanguageCert ESOL Communicator B2 involves a monologue which is played twice.

The person taking the test is asked to write short answers. The monologue is in spoken English, with elided forms, idioms and is designed in this way to reflect authentic speech in real-life situations.

This task is similar to real-life situations in which a learner needs to listen to a monologue to obtain essential information from a radio broadcast, a presentation, a narrative or an interview.

The instructions are given in English in written mode.

The person taking the LanguageCert ESOL Communicator B2 to complete Listening part 3, is free to move between the different parts of the test and allocate test time as they wish. At the end of the exam they are given ten minutes to transfer their answers to the answer sheets.

There is a pause of ten seconds after the recording is played for the first time, and a pause of thirty seconds after it is heard for the second time. This allows learners to check their answers before they move on to the next part of the test. Optionally, one can always revisit their answers, at the end of the Listening part. That is, after all parts are played twice and before the people taking the test move on to the rest of the paper.

II. What the person being assessed reads/listens to/sees (the input)

The input may include monologues based on an authentic text such as a radio broadcast, a guided tour, an interview, instructions, presentations or advertising announcements. The texts contain information of a general nature that do not require an understanding of technical details and are all accessible to a varied age-range.

The input mainly relates to the public or educational domain.

Communication themes may include entertainment, places, shopping, sports, group activities and daily life.

The input is authentic texts, designed for the specific part, taking all specifications into account, ensuring themes and content is on level and that level appropriacy is maintained. Monologues are in spoken English, incorporating intonation features as well as colloquial English, natural responses and interjections.

Spoken input is read from a script, they are recorded on the premises under the presence and supervision of the Academic Associate responsible for the material development and

production, in case of last-minute corrections and to ensure accurate execution of the script. The actors recording the materials, are highly experienced in educational materials and they are all native speakers.

To make material suitable for the LanguageCert ESOL Communicator B2, changes might be made, such as rewriting items that might become outdated, re-recording might also take place, phrases might get simplified in case they are too difficult for the level and certain parts might be rephrased if they do not seem to work while recording (e.g. do not sound natural).

The input is about eight minutes long for Part 3, including the repetition and pauses.

The prompt on the message pad is one to ten words long, phrased as clearly as possible, and the monologue is 460-500 words long. A short heading is given on the message pad to set the context.

The vocabulary of the input involves a range of everyday life transactions, some professional exchanges and some more formal, understanding gist, context, purpose and key ideas and distinguishing between fact and opinion.

The grammar of the input involves a range of structures and sentence patterns according to B2 level of the CEFR.

The delivery is at a moderate, natural rate.

The accent used is standard UK English, with some US English. There is no distortion used on the recording, or background noises.

There is no background noise or distortion on the recordings.

The recording involves two speakers, a male and a female.

People being assessed hear the recording twice.

The input is comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do (the expected response).

The response involves completing notes on a message pad with prompts. Ten seconds are allowed before the monologue is played again for the second time.

Responses are required to be up to five words long.

The main rhetorical functions expected are: understanding narratives, sequences, instructions, descriptions and explanations.

The response is expected to be in the form of note-taking, no full sentences are required.

The main purpose of the response is either referential (gives facts about the world) or metalingual (verifies understanding).

The vocabulary of the response is expected to involve a sufficient range of vocabulary for most general topics with simple, short phrases without any complex structure.

The range of grammar in the response is expected to involve simple grammatical structures.

The level of coherence and cohesion in the response is expected to involve a limited number of devices to link utterances in a clear manner.

In responding to part 3 of the LanguageCert ESOL B2 Communicator, people are expected to draw on a wide range of non-specialised knowledge areas.

Example of Part 3

Listening Part 3

You will hear someone talking. You will hear the person twice. Complete the information. Write short answers of one to five words.

Today, I'm going to give you some information on the college library rules of use and regulations. All right, so here we are in the library. During regular class days the library is open from seven thirty am to seven pm and from eight to twelve midday on Saturday. It's closed on Sundays. Only students with identification cards will be allowed to use the library. Your identity cards are required at all times and must be given to the person in charge of the control desk each time you enter the library. Lending of IDs to other students is strictly prohibited and if you are caught doing that, your library privileges will be suspended. If you lose your ID, report it immediately to the Office of Registrar. Don't confuse IDs with library cards. Only students with library cards will be allowed to borrow books. If you lose your library card, report it and it'll be replaced after a week for the date of application. Now, we have different categories of books here. Reserved books can be borrowed for use inside the library for two hours and may be renewed for another hour unless other students need them. You can also check them out overnight after four pm, but they need to be back before nine am the following morning, unless it's a Sunday. [...]

College Library Rules and Regulations

1. Place to show identity cards

.....

2. When to check out reserves:

B2 Listening

Part 4

I. Rationale

This Listening part 4 of LanguageCert ESOL Communicator B2, involves one long dialogue between two speakers which is heard twice. The person taking the test is presented with six three-option multiple choice questions. The dialogue in part 4 is in spoken English, with elided forms, idioms and designed in this way, to reflect an authentic conversation in real-life situations.

This task is similar to an everyday occurrence between two people, conversing on daily issues at work, school or home.

The instructions are given in English, written mode.

The person taking the Listening part 4 of the LanguageCert ESOL B2 Communicator, has twenty seconds at the beginning of Part 4, and ten seconds before the dialogue is played for the second time.

People taking the test are free to move between the different parts of the test and allocate test time as they wish. At the end of the exam they are given ten minutes to transfer their answers to the answer sheets.

II. What the person being assessed reads/listens to/ sees (the input)

The input may include a discussion, inter-personal dialogues and conversations, a job interview or a phone conversation, all accessible to a varied age-range. The dialogue is in spoken English, incorporating intonation features as well as colloquial English, natural responses and interjections. The input contains information of a general nature that do not require an understanding of technical details.

The input mainly relates to the personal and public domain.

Communication themes may include daily life, free time, social gathering, shopping, food and drink, places and relations with other people.

The input is designed specifically for Part 4, taking all specifications into account, ensuring themes and content is on level and that level appropriacy is maintained throughout.

Spoken input is read from a script, recorded on the premises under the presence and supervision of the Academic Associate responsible for the material development and production, in case of last-minute corrections and to ensure accurate execution of the script. The actors recording the materials are highly experienced in working on educational materials and they are all native speakers of English.

To make material suitable for the LanguageCert ESOL Communicator B2, changes might be made, such as rewriting items that might become outdated, re-recording might also take place, phrases might get simplified in case they are too difficult for the level and certain parts might be rephrased if they do not seem to work while recording (e.g. do not sound natural).

The input is mostly familiar and concrete, every day issues, daily routines and likes/dislikes are addressed.

The input is about eight minutes long for part 4, including the repetition and pauses.

According to the instructions by the LanguageCert International ESOL designers, each multiple-choice item is a maximum of 35 words long including options, phrased as clearly as possible, and the dialogue is 540-600 words long.

The vocabulary of the input involves a range of everyday life transactions, involving familiar situations and topics.

The grammar of the input involves a range of structures and sentence patterns according to B2 level of the CEFR.

The accent used is standard UK English, with some US English.

There is no distortion used on the recording, or background noises.

The recording is always played twice it involves three speakers, either a male and two females, or a female and two males.

The input is likely to be comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do (the expected response).

The response involves six multiple-choice questions, each with three options to select from. The people being assessed have twenty seconds before the recording is played for the first time, and ten seconds before it is played for the second time.

The main rhetorical functions expected are: understanding narratives, sequences, instructions, descriptions and explanations.

The main purpose of the response could be emotive (describes the emotional state of one of the speakers) or referential (gives facts about the world) or metalingual (verifies understanding).

In responding to part 4 of the LanguageCert ESOL B2 Communicator, people are expected to draw on a wide range of non-specialised knowledge areas.

Example of Part 4

Listening part 4

You will hear a conversation. You will hear the conversation twice. Choose the correct answers.

- | | |
|----|---|
| M: | So have you heard about this amazingly generous money donated by one of our ex-students, Penny? |
| F: | Yes, Richard Judd. It's brilliant isn't it, Jeff? |
| M: | Were you on the teaching staff like me when Richard Judd was studying here? |
| F: | Errm, from what I remember he left for art college the summer before I started. Looking forward to meeting him now, though. |
| M: | Oh, such a nice guy, really. I loved teaching him. And this donation is so generous, I mean, he's not as rich as all that. |
| F: | I know, but apparently, he said he wanted to give the school something back for all the encouragement and help he was given when he was here. |
| M: | Yes, which makes me determined to make the best use of the money, not just invest it in the bank but not just spend it on lots of different little things either... You know, so there's nothing really to show for it. |
| F: | Oh, exactly, Jeff. What about the feedback from students' parents – you know the questionnaire you sent them over half – term. |
| M: | Oh well, not too many surprises there. More computers, books, that sort of thing, nothing very exciting, although all very necessary of course. |
| F: | Mmmm... I know what you mean – we want something permanent, that everyone would like, that's what Richard would want I'm sure... |
| M: | ...which is why, Penny, I was wondering about a water fountain. Everyone appreciates the calming effect of water, children especially. We could just watch it and lose all the stress... |
| F: | A fountain! You must be joking! |
| M: | Why not? Not very big or deep or anything – it wouldn't be dangerous, and if we choose the right one, it could be really attractive. |
| F: | No, no, I didn't mean that. But just think of all the coke cans and crisp packets floating about in it. The kids just couldn't resist dropping stuff in, it would just become an alternative bin. |
| M: | Mmmm... hadn't really considered that actually. Another idea we've been asked to consider comes from the English and drama department, a brilliant idea – for a new school theatre... |
-

F: [laughs] Well, I know Richard's been generous but not that much! It would cost a fortune! I agree it would be great and we do have the space for one, but it'll have to wait, I'm afraid.

M: Yes, I know – I'm sure it'll happen one day. And hang on [slight pause] – yes, the final idea, a sculpture comes from the students themselves, who were asked to suggest ideas to their teachers. [...]

1. Who taught Richard Judd when he attended the school?

- a) Only Jeff.
- b) Only Penny.
- c) Both Jeff and Penny.

2. Why does the man suggest a water fountain?

- a) It would encourage wildlife.
- b) It would appeal to Richard.
- c) It would be very relaxing.

B2 Reading

Part 1

I. Rationale

This Reading part 1 of the LanguageCert ESOL Communicator B2 involves one long text of 450-500 words, followed by six multiple-choice items.

This task is similar to a news story, article, review or proposal one would come across in real life, in magazines or online.

The instructions are given in English in written mode.

There is no time limit for this part alone, all four parts of Reading and the two parts of Writing are to be completed in two hours and ten minutes.

II. What the person being assessed reads (the input)

The input may include a news story, article, report, review or proposal. Texts are based on authentic sources but are independent of any background or world knowledge. The text has a title to help set the context. The questions follow the order of the text.

The input mainly relates to the personal or public domain.

Communication themes may include the environment, daily life, relations with other people and places.

To make material suitable for the LanguageCert ESOL Communicator B2, texts are designed specifically for this part of the examination, taking all specifications into account, ensuring themes and content is on level and that level appropriacy is maintained.

The input is 450-500 words long, and the multiple-choice items are up to 15 words each with the stem having a maximum of 10 words. The stem is either a question or an incomplete sentence.

The vocabulary of the input involves a broad range of topics and some idiomatic expressions and colloquialisms at B2 level, according to the CEFR.

The grammar of the input involves a range of structures and sentence patterns according to B2 level of the CEFR.

The input is likely to be comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do.

The response involves answering six multiple-choice questions by selecting the correct answer from the three options.

The main rhetorical functions expected are description of events, a process, reporting events, analysis or evaluation of a global issue.

The main purpose of the response is to show detailed understanding of information, ideas and opinions in a longer text.

In responding to Reading part 1 of the LanguageCert ESOL B2 Communicator, people are expected to draw on common, general, non-specialised knowledge.

Example of Part 1

Reading Part 1

Read the text and the questions. Choose the correct answer for each question.

There's a reason why we love computer games

Video computer games have become as much a part of our entertainment world as books and movies. Why is it that so many of us are drawn to the computer monitor or TV screen? A bit of history might help explain the phenomenon. Originally, computers were all work and no play, being expensive to build and operate. But computer engineers and designers, being overgrown kids after all, soon began to get comfortable with their big machines. It occurred to some of them that it might be a good idea to have the computer show some spaceships and planets cruising around the galaxy.

And in the 1960s a group of fun-loving computer geniuses created a game called *SpaceWar* in their idle hours. *SpaceWar* was well received by geniuses and many others as well, which motivated the geniuses to design some more fun stuff. Somehow the visual element of the games activated the imagination. Over time we got *Pong*, *PacMan* and a continually evolving string of games.

1. *SpaceWar* was developed
 - a) as part of a serious computer project.
 - b) by computer experts in their spare time.
 - c) to show people what space travel is like.

2. After *SpaceWar* more games appeared because
 - a) computer design improved.
 - b) it wasn't exciting enough.
 - c) people enjoyed playing it.

B2 Reading

Part 2

I. Rationale

The Reading part 2 of the LanguageCert ESOL Communicator B2, involves one long text with sentences removed. A list of seven sentences are placed below the text, labelled A to G, six of them to be used in the gapped parts. One serves as a distractor and is not used.

This task is similar to reading a general, informative text, placing particular emphasis on coherence and cohesion.

The instructions are given in English in written mode.

There is no time limit for this part alone, all four parts of Reading and the two parts of Writing are to be completed in two hours and ten minutes.

The task is given a title to help set the context. It is 340 – 400 words long, including the six removed sentences.

II. What the person being assessed reads (the input)

The input may include books, fiction, journals, magazines, blog entries, instructions or articles.

The input mainly relates to the public domain.

Communication themes may include environment, entertainment, places, environment or personal relations.

The input is either adapted from sources or prepared especially for this part of the exam. All materials are checked by editors and expert item writers to ensure accordance with the CEFR.

To make materials suitable for Reading Part 2, vocabulary and grammar in the authentic texts may be amended to match the level. Names and locations may also change and topics that could be answered through world knowledge are not tested.

The input is less familiar and somewhat abstract: contemporary issues and topics.

The vocabulary of the input involves a range of idiomatic expressions and topics, all in accordance with B2 level of the CEFR.

The grammar of the input involves a limited range of complex structures

The input is comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do (the expected response).

The response involves filling the six gaps of the text with one of the seven sentences given below the text. There is only one suitable sentence for each gap and one extra sentence that does not need to be used.

The response is restricted and controlled. The main purpose of the response is to show understanding of how meaning is built up in discourse.

In responding to Reading part 2 of the LanguageCert ESOL B2 Communicator, people are expected to draw on a wide range of non-specialised knowledge areas.

Example of Part 2

Reading Part 2

Read the text. Use the sentences to complete the text. Choose the correct sentence for each gap. There is one extra sentence you will not need.

Playing and learning

Throughout history children have played in groups and taken part in imaginative games. They've pretended to be pirates and princesses, heroes and villains. 1. However, the way in which children play has changed during recent times. Children now have a much wider range of toys to choose from and as a result spend less time playing pretend games nowadays. 2.

Pretend games actually help children to develop an important learning skill called 'executive function',

- A Parents created secure environments to play in.
 - B They've improvised and regulated their play by making up their own rules and characters.
 - C Researchers recently repeated this experiment, and found very different results.
 - D This recent trend has been shown to have an impact on their imagination.
 - E As a result some teachers think playing is a waste of time.
 - F This means that children talk to themselves about what they are going to do and how they are going to do it.
 - G By comparing two studies, we can now prove that children's ability to self-regulate has been reduced in the past sixty years.
-

B2 Reading

Part 3

I. Rationale

This Reading part 3 of LanguageCert ESOL Communicator B2 involves four short texts with a linked topic. The person taking the test must read the four texts and answer seven questions.

This task is similar to the information one would find in a brochure or leaflet or a letter.

The instructions are given in English in written mode.

There is no time limit for this part alone, all four parts of Reading and the two parts of Writing are to be completed in two hours and ten minutes.

II. What the person being assessed reads (the input)

The input may include part of an email, an advertisement, a leaflet, a letter or an announcement. The first three questions begin with the stem, "Which text: ". They test awareness of purpose of texts or intended audience.

The last four questions begin with the stem, "Which text provides the answers to the following questions?". They test the ability to scan for information.

Each of the four texts is 90-110 words long. They are labelled A to D. The total length of the four texts is 380-420 words. Each of the seven questions are maximum 15 words long.

The input mainly relates to the public domain.

Communication themes may include the environment, daily life, free time, shopping, food and drink or places.

The input is either adapted from sources or prepared especially for this part of the exam. All materials are checked by editors and expert item writers to ensure accordance with the CEFR.

To make materials suitable for Reading Part 3, vocabulary and grammar in the authentic texts may be amended to match the level. Names and locations may also change and topics that could be answered through world knowledge are either not tested.

The input is mostly familiar and concrete.

The vocabulary of the input involves a sufficient range, for most general topics with varied formulation.

The grammar of the input involves a limited range of complex structures.

The input is likely to be comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do (the expected response).

The response involves answering seven questions by choosing the text that represents the answer.

Responses are expected to be either A, B, C or D. Each point is targeted only once, and people taking the test must employ skimming and scanning to choose the correct text for each question.

The response is controlled by the instructions.

The main purpose of the response is to verify understanding.

In responding to Reading part 3 of the LanguageCert ESOL B2 Communicator, people are expected to draw on common, general, non-specialised knowledge, showing awareness of the purpose of different texts and the ability to locate specific information.

Example of Part 3

Reading Part 3

Read the four texts. Which text gives you the answer to each question? Choose the correct text (A-D) for each question.

A

Ideal for the weekend sailor, this boat can be towed by car and trailer to and from your house to wherever you would like to sail. Sailing areas have slipways where you can reverse your boat into water and then park your car. You may decide to store your boat near the water, in a boatyard or a caravan park. The boat can also be moored on the water. The safest and most secure way is to use a marina. These can easily be found on the coast, inland waterways and lakes.

B

When I was thirteen and my brother was ten, this farmer on a farm where we were staying on holiday took us a long way out to sea in his boat to do some fishing. It was a fantastic day. We had a great time, caught some fish and then very quickly the skies turned dark grey and we were in a race against time to get back to the coast as the winds got up and the rain poured down. It's only thinking back that I realise the danger we were in.

C

Before you start, ask someone who is experienced to show you the equipment on your boat. All boats

Which text:

- | | | |
|----|--|--|
| 1. | gives detailed instructions? | |
| 2. | encourages people to join an organisation? | |
| 3. | is aimed at potential purchasers? | |

Which text provides the answers to the following questions?

- | | | |
|----|---|--|
| 4. | Where's the best place to store a boat? | |
| 5. | What sort of weather did someone have to cope with? | |
| 6. | Who wants to avoid accidents at sea? | |
| 7. | Where's the best place to learn to sail? | |

(Total: 7 marks)

B2 Reading

Part 4

I. Rationale

In Reading part 4 of the LanguageCert ESOL Communicator B2, there is a long-paragraphed text followed by seven open questions.

The task is similar to reading an informative, semi-academic text, an encyclopaedia entry, a literary text or a biography.

The instructions are given in English in written mode.

There is no time limit for this part alone, all four parts of Reading and the two parts of Writing are to be completed in two hours and ten minutes. The person taking the test must read the text and answer seven questions.

II. What the person being assessed reads (the input)

The input may include essays, books, fiction, literary journals or magazines.

The input mainly relates to the public domain.

Communication themes may include the environment, historical facts, literature excerpts, scientific issues or biographies.

The input is prepared and designed especially for this part of the assessment.

To make material suitable for Reading part 4 of the LanguageCert B2 Communicator, authentic texts are adapted to B2 level according to the CEFR. The texts are given a title to help set context. Topics, vocabulary and grammar of the text have been adapted and checked by editors and expert item writers to ensure accordance with B2 level of the CEFR. Items are not answerable through world knowledge and the topics are selected carefully, so as not to become outdated.

The input is somewhat abstract, topics normally encountered in academic or vocational life.

This is linked to reading a literary text or an academic article.

The input is 400-450 words long. The questions are no more than 12 words long each and answerable in one to five words. Answers follow the order of the text, which is laid out as an article with paragraphs, without pictures or artwork.

The vocabulary of the input involves a broad range, including some idiomatic phrases and colloquialisms of B2 level.

The grammar of the input involves a range of structures on B2 level.

The input is comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do (the expected response).

The response involves answering all seven open-ended questions.

Responses are required to be one to five words long, to show understanding of the text.

The main purpose of the response is to show the person taking the test's ability to locate and obtain specific information; the correct written response will be a close approximation of what is written in the text.

The main rhetorical functions expected are description of events/data/objects, presentation, explanation, reporting of events and exemplification.

The main purpose of the response is to clarify or verify understanding.

The vocabulary of the response is expected to involve a sufficient range of vocabulary for most general topics with varied formulation.

The range of grammar in the response is expected to involve a limited range of complex structures.

In responding to Reading part 4 of the LanguageCert ESOL Communicator, people are expected to draw on a wide range of non-specialised knowledge areas.

Example of Part 4

Reading Part 4

Read the text and answer the questions. Use a maximum of five words for each question.

History of the Panama Canal

The Panama Canal is a vital shipping lane linking the Atlantic and Pacific Oceans. Opened in 1914, it was one of the greatest engineering projects of the modern age. However, the concept of a canal through the Isthmus of Panama, a comparatively narrow strip of land linking North and South America, had been thought of as far back as the 16th century. When the Spanish explorer Vasco Núñez de Balboa led an expedition across the Isthmus of Panama in 1513, he saw the potential for a passage that would allow access from sea to sea without having to navigate around Cape Horn at the southernmost tip of South America.

In 1534, Charles V, King of Spain, formally ordered further investigation into the possibility of the construction of a ship canal across the Isthmus. The surveyed route followed, to a large extent, the course of the present day Panama Canal. However, the governor politely informed Charles V that the undertaking of such an engineering operation was impossible.

1. Where, specifically, did Balboa not want to sail?

.....

2. What was Charles V told about his plan for a canal?

.....

B2 Writing

Part 1

I. Rationale

This part of the paper involves people taking the test responding to a rubric and an input text in order to produce a neutral/more formal response for an intended audience.

The task is similar to real-life situations. It is intended to simulate a response the person taking the test would give to an informal letter/email to a friend or a competition, a blog entry etc.

The instructions are given in English, written mode.

There is no time limit for this part alone, all four parts of Reading and the two parts of Writing are to be completed in two hours and ten minutes.

II. What the person being assessed sees/reads (the input)

Input text is given as e.g. letter, poster, diary, timetable, leaflet or other suitable form. There are three required points in the rubric in bullets. The topics are accessible to a wide range of learners.

The input mainly relates to personal, public or educational domains, all familiar to the person taking the test.

The input text is 50-65 words long and the rubric is maximum 35 words long, excluding the phrase, "Write between 100-150 words."

The input text and topic are comprehensible to a language learner at CEFR level B2.

Topics, and vocabulary and grammar of the rubric have been adapted and checked by editors and expert item writers to ensure accordance with B2 level of the CEFR.

Vocabulary and grammar in the input is at B1/B2 level of the CEFR in order to ensure that the task is clear and not dependent upon reading ability.

The input is familiar and mostly concrete, dealing with personal preferences, routines, leisure, all topics within the person taking the test's knowledge scope.

III. What the person being assessed needs to do (the expected response)

The response involves writing an article, personal email or letter of 100-150 words. They need to cover all three bullet points found in the rubric. They need to produce appropriate content for the intended audience that is public and distant.

Appropriate use of a variety of organisational patterns and a wide range of cohesive devices are expected from the person taking the test.

The main rhetorical functions expected are persuasion letter, letter/email (personal), description of events (article) etc.

The response is controlled by the instructions and the main purpose is either referential or persuasive.

The vocabulary of the response is expected to involve a sufficient range of vocabulary for most general topics with varied formulation.

The range of grammar in the response is expected to involve a limited range of complex structures.

The level of coherence and cohesion in the response is expected to involve clear, smoothly flowing, well-structured production, connectors and cohesive devices.

In responding to Writing part 1 of the LanguageCert ESOL communicator, people are expected to draw on common, general, non-specialised knowledge areas.

Example of Part 1

Writing Part 1

Write a **letter** in order to enter the competition advertised in the notice below. Mention

- when you use the centre
- why you want membership
- how to improve facilities

Write between 100 and 150 words.

Free Leisure Centre Membership: Competition

If you sometimes use our leisure centre, but are not a member, why not win a year's free membership? All you have to do is write about yourself and our centre.

We will publish the winning letter in our newsletter and your free membership card will follow.
Write to Joanne Blaire, Centre Manager

B2 Writing

Part 2

I. Rationale

The Writing part 2 of the LanguageCert ESOL Communicator involves producing one of the text types below: an informal email / letter to an English friend, a narrative, a descriptive composition, a report or a review.

This task is similar to a real-life communication with an organisation, a publisher, an editor or a friend.

The instructions are given in English, in written mode.

There is no time limit for this part alone, all four parts of Reading and the two parts of Writing are to be completed in two hours and ten minutes.

II. What the person being assessed reads (the input)

The input may include books, magazine articles, competition entries, professional letters, personal accounts.

The input mainly relates to personal and public domain or educational domains, all familiar to the person taking the test.

Communication themes may include personal identification, home, environment, social issues.

To make material suitable for the Writing part 2 of the LanguageCert ESOL Communicator, topics, vocabulary and grammar of the rubric have been adapted and checked by editors and expert item writers to ensure accordance with B1-B2 level of the CEFR in order to ensure the task is clear and not dependent upon reading ability.

The input is mostly familiar and concrete asking for matters encountered at work, school or leisure.

The input is maximum 40 words long.

Two content points are given that need to be covered by the person taking the test.

The person taking the test is asked to respond to the topic using between 150-200 words.

The vocabulary of the input involves a sufficient range of vocabulary to express most topics pertinent to everyday life such as personal preferences, work, travel, family and leisure.

The grammar of the input involves a limited range of complex structures. Coherence and cohesion are expected at B2 level.

The input text and topic are comprehensible to a language learner at CEFR level B2.

III. What the person being assessed needs to do (the expected response)

Response is expected to be 150-200 words long. Two content points are given that need to be covered by the person taking the test.

Cohesive devices must be employed. Both content points must be covered in an organized text, as directed by prompt, and are expected to be clearly structured.

The main rhetorical functions expected are descriptions of events (essay), description of scenes, explanation, giving opinion or writing personal emails.

The response is expected to be in the form of a letter/email, a narrative, a descriptive composition, a report or a review.

The response is controlled by the prompt (two points to be covered).

The main purpose of the response is referential, conative or emotive.

The vocabulary of the response is expected to involve a sufficient range with some idiomatic expressions.

The grammar of the response is expected to involve a wide range of structures.

The level of coherence and cohesion in the response is expected to involve linking words, organisational patterns, and smoothly flowing production.

In responding to Writing part 2 of the LanguageCert ESOL Communicator, people are expected to draw on personal/daily life, and a wide range of non-specialised knowledge areas.

Example of Part 2

Writing Part 2

Write an **essay** for your college magazine entitled 'My earliest childhood memories'. Describe your clearest memory and say why you remember it so clearly.

Write between 150 and 200 words

LanguageCert International ESOL (Speaking)

B2 Speaking

Part 1

I. Rationale

Part 1 of the Speaking examination involves five questions asked by the Interlocutor on five different topics. The person taking the test spells their name and responds to questions about him/herself. The interlocutor has a bank of questions to select from.

The task is similar to everyday situations: getting acquainted, talking about preferences, communicating personal information etc.

The instructions are given in English, in spoken mode.

The time limit for this part of the exam is three minutes.

Interlocutors have native-like fluency and go through rigorous training before placement. They get additional training every year and are regularly evaluated on their performance by the Chief marking examiner.

II. What the person being assessed listens to (the input)

In this task, the person taking the test is mainly asked about every day or general personal/life preferences.

Communication themes include: daily life, free time, shopping, food, films, places, weather etc.

Input is familiar and mostly concrete. The vocabulary and grammar of the input are adjusted to the B2 level of the CEFR.

The vocabulary of the input involves a sufficient range of lexis for general topics.

There are no visuals for this part.

Input is read from script to ensure standardization.

III. What the person being assessed needs to do (the expected response).

The people taking the test are expected to answer the Interlocutor's questions in a few sentences. It should not take more than three minutes to answer all five questions.

The main purpose of the response is to talk about personal preferences and opinions, and briefly justify their responses.

The vocabulary and grammar used is expected to be of B2 level according to the CEFR.

There needs to be coherence and cohesion in the responses with clear and smoothly flowing speech.

In responding to part 1 of the Speaking exam, the person taking the test is expected to draw on general, non-specialised knowledge.

Example of Speaking part 1

Topics

Television

- How much television do you watch in a week?
- What kinds of programmes do you like and dislike?
- What are the advantages and disadvantages of watching television?
- What are a parent's responsibilities regarding television?

Friendship

- What qualities do you look for in a good friend?
- Would you prefer a wide circle of friends or just a few close friends? Why?
- How can friends help you in times of difficulty?
- Can you tell me about ways friends have helped you?

Leisure facilities

- What leisure facilities do you and your friends use?
- What kinds of leisure facilities are there in your town?
- What leisure facilities are popular with older people?
- If **you** were responsible, what leisure facilities would you build in your area?

(continued)

Public transport

- Why do you think some people pay lots of money to fly first class?
- What's the best thing about public transport where you live?
- If you had a lot of money, how would you travel?
- What would improve public transport for you in your area?

Foreign language ability

- Why do you think it is important to speak a foreign language?
- Why do you think some people learn a language and then forget how to speak it?
- Which languages would you like to learn and why?
- How useful have you found speaking English when you travel?

B2 Speaking

Part 2

I. Rationale

Part 2 of the Speaking examination consists of two or three situations that are presented by the Interlocutor and people taking the test are required to respond to and initiate interactions.

This task is similar to an interaction that could take place in real life.

The instructions are given in English, in spoken mode.

This task is three minutes long in duration. The person being assessed is presented with situations and initiates or responds, in order to have a dialogue with the Interlocutor.

II. What the person being assessed listens to (the input)

The input relates to real-life situations, using a range of functional language to elicit or respond as appropriate.

The Interlocutor presents the situation to the person taking the test, the relationship of the speakers, and either tells the person taking the test to start or respond. They both take turns to create a short conversation during which, they resolve an issue related to everyday life.

Each situation should be no more than two turns for each person.

Input could be either formal, neutral or informal interaction.

Input is familiar and mostly concrete. The vocabulary and grammar of the input are adjusted to B1-B2 level of the CEFR in order to ensure the task is clear and impacted by reading ability.

III. What the person being assessed needs to do (the expected response).

Utterances need to be three minutes long in total, two turns are expected for each speaker.

The vocabulary and grammar used is expected to be of B2 level according to the CEFR.

The response needs to have coherence and cohesion, with clear and smoothly flowing speech.

In responding to part 2 of the Speaking exam, the person taking the test is expected to draw on general, non-specialised knowledge.

Example of Speaking part 2

A

- I'm your new neighbour. I start:
Hello. My name's _____. I've just moved in next door.
- We're flatmates. You don't really want to go out tonight. I start:
Come on, we'll be late if we don't go soon.
- We're in a queue in a cafeteria. I start:
Excuse me, have we met before?
- We're colleagues. I start:
You know the boss is retiring at the end of the month... what shall we do?

C: *(Responds.)*

I: *(Role-play the situation with the candidate – approximately two turns each.)*

I: *Second situation (choose one situation from B).*

B

- You're just leaving my party and realise you don't have enough money to get a taxi home. You start.
 - We're friends. You want to persuade me to go to the cinema with you this evening. You start
 - I am your teacher/boss. You have just arrived very late at school/work yet again. You start.
 - We're colleagues. You want me to cover for you this weekend. You start.
-

B2 Speaking

Part 3

I. Rationale

In Part 3 of the Speaking examination, the focus of the task is to co-operate in order to reach an agreement or a decision, using written text as a prompt.

The task is similar to a real-life situation where a person would interact with someone, in order to reach a conclusion about an everyday issue.

The instructions are given in English, in Spoken mode.

The interlocutor exchanges information with the person taking the test in order to perform a task.

The goal of the task is to make a plan or arrangement after they discuss and reach a conclusion.

II. What the person being assessed listens to (the input)

A task sheet is handed over to the person taking the test by the Interlocutor. The person taking the test is then given twenty seconds to think about what they want to say after being presented with the task. The Interlocutor has some different information on the same task.

The person taking the test's task sheet has, on average, four different details to the Interlocutor's task sheet. The person taking the test uses their prompt sheet and this 'information gap' in order to hold a discussion, collaborate and reach an agreement with the interlocutor.

There are about three decisions to be made between the speakers. Task sheets bear information in the form of notes and not full sentences, to keep reading to a minimum. People taking the test do not see the interlocutor's sheet.

This task, Speaking part 3, is three minutes long.

III. What the person being assessed needs to do (the expected response)

The response involves taking the prompted sheet into account, taking turns and respond/discuss/collaborate and reach an agreement with the interlocutor on the specific task.

The task does not require the person taking the test to adopt a different persona but draw on their own knowledge or opinions while using the information provided at the same time.

The grammar and vocabulary used and elicited must be at B2 level according to the CEFR. The grammar range in the response is expected to involve a wide range of grammatical structures and vocabulary must be sufficient for a range of topics with varied formulation.

Responses are required to have coherence and cohesion, with clear and smoothly flowing speech.

In responding to Speaking part 3 of the LanguageCert Communicator, people are expected to draw on personal/daily life, basic communication needs and a wide range of non-specialised knowledge areas.

Example of Speaking part 3

PART 3 (3 minutes)

I: Now, Part Three. We're going to discuss something together. We are planning to go on holiday with some friends. Let's talk about what kind of holiday we want and make some decisions. Here are some ideas. (*Hand over candidate's task sheet.*) I have some different ideas.

Take twenty seconds to think about what you want to say. (*20 seconds.*) Please start.

Interlocutor's Task Sheet

Holiday plans

<i>Time:</i>	<i>First two weeks in August</i>
<i>Location:</i>	<i>Coastal resort</i>
<i>Transport:</i>	<i>Hire car</i>
<i>Accommodation:</i>	<i>Self-catering apartment</i>
<i>Activities:</i>	<i>Swimming, sunbathing,</i>
<i>Reserve through:</i>	<i>Internet</i>

Candidate's Task Sheet for Part Three (Interlocutor's copy)

Holiday plans

<i>Time:</i>	<i>First two weeks in June</i>
<i>Location:</i>	<i>Mountains</i>
<i>Transport:</i>	<i>Hire car</i>
<i>Accommodation:</i>	<i>Family hotel</i>
<i>Activities:</i>	<i>Walking, photography</i>
<i>Reserve through:</i>	<i>Internet</i>

B2 Speaking

Part 4

I. Rationale

Part 4 of the speaking exam involves a task that requires the person taking the test to speak for two minutes and then respond to follow-up questions on the same topic. The person taking the test has thirty seconds to think about the topic before they start talking.

The task is similar to a real-life situation where a person would express their personal opinion on a general topic.

The instructions are given in English, in Spoken mode.

The aim of the task is to narrate, describe, communicate ideas and express opinions.

II. What the person being assessed listens to (the input)

The person taking the Speaking part 4 of the LanguageCert Communicator, is given a topic, they have thirty seconds of preparation time and then they talk about the task for two minutes. The interlocutor asks some follow-up questions on the same topic, for another two minutes.

The topic provides enough scope for people taking the test to expand on.

Spoken input is of moderate natural rate in a range of grammatical structures and vocabulary for most general topics mostly expressed in wh- questions.

Part 4 is four minutes long.

III. What the person being assessed needs to do (the expected response)

The response involves speaking for about two minutes and expanding on the topic given by the interlocutor, after thirty seconds of preparation time.

The grammar and vocabulary used and elicited is at B2 level according to the CEFR. The grammar range in the response is expected to involve a wide range of grammatical structures and vocabulary must be sufficient for a range of topics with varied formulation.

The response is expected to have coherence and cohesion in the responses with clear and smoothly flowing speech.

In responding to Speaking part 4 of the LanguageCert Communicator, people are expected to draw on personal/daily life, a wide range of non-specialised knowledge areas and social-related knowledge areas.

Example of Speaking part 4

PART 4 (4 minutes including follow-up questions)

I: In Part Four you are going to talk on your own for about two minutes. Your topic is *(choose topic for candidate)*.

Topics

A Care for elderly people

B Important inventions

C Some festivals in your country

I: You now have thirty seconds to write some notes to help you. *(Hand over piece of paper and pen/pencil.)* So your topic is *(repeat topic)*. *(Withdraw eye contact for thirty seconds. Leave recorder running.)*

I: *(Candidate's name)*, please start.

C: *(Talks.)*

I: *(When candidate has talked for a maximum of two minutes, say, 'Thank you', and then ask some follow-up questions.)*

Follow-up questions**Care for elderly people**

- What do you think of special homes to look after old people?
- What problems do a larger population of very old people present?
- Do you think people should work for longer?
- How do people spend their retirement in your country?

Important inventions

- Which modern invention do you use most?
- Do you think any modern invention has had a bad effect on society?
- What invention would make the world a better place?
- Which modern invention has changed life dramatically?

Some festivals in your country

- How have these festivals changed in your country in the past 50 years?
- How important is it for children to learn about traditions?
- Which festivals always have traditional food?
- Which festival would a visitor enjoy the most?

Scoring Assessment Tasks

Objectively-marked subtests (Listening Parts 1, 2, 3 & 4, Reading Parts 1, 2, 3 & 4)

In objectively scored tests (Reading and Listening), responses are marked automatically by LanguageCert's PASSPORT Marking Application. Some task types require human markers to check the accuracy of very short answers against very tight mark schemes. The key in the mark schemes is developed as part of the item production process, and this is analysed rigorously at the item editing and pretesting stage to ensure the key is as tight and unambiguous as possible.

Subjectively-marked subtests (Writing Tasks 1 and 2)

In the LanguageCert standard model for Writing tests, people taking the test complete two writing tasks. Writing responses are marked using an analytic mark scheme which is tied to the CEFR descriptors. The assessment criteria for Writing are,

- a) Task Fulfilment,
- b) Accuracy and Range of Grammar,
- c) Accuracy and Range of Vocabulary, and
- d) Organisation.

For Writing, 12 marks may be awarded per Writing Task (3 per assessment criterion). Each writing task is assessed separately and then marks are added up to give a total maximum of 24 raw marks for Writing.

Subjectively-marked subtests (Speaking)

In the LanguageCert standard model for Speaking tests, people taking the test participate in a short interview (12 minutes). Speaking responses are recorded by the Interlocutor and marked – by a trained marking examiner – using an analytic mark scheme which is tied to the CEFR descriptors. The assessment criteria for Speaking are,

- a) Task Fulfilment and Coherence,
- b) Accuracy and Range of Grammar,
- c) Accuracy and Range of Vocabulary, and
- d) Pronunciation, Intonation and Fluency.

For Speaking 12 marks may be awarded per candidate interview (3 per assessment criterion).

The accuracy of scoring is assured and checked by LanguageCert's Academic Associate for Marking. S/he second marks at least 10% of all Writing scripts and/or Speaking tests per marker per exam session. Second making has two objectives. Firstly, to ensure the accuracy of marking, secondly to monitor marker performance.

Where there are discrepancies between the marks awarded by the marking examiner and those awarded by the Academic Associate for Marking, the latter stand. Discrepancies are for the most part minor and reasonable (one mark per Writing task or Speaking test). Where more sizeable and extensive discrepancies are identified, the level of inaccurate or inconsistent marking may be identified as unacceptable by the Academic Associate for Marking. In such a case, the marking examiner is stopped temporarily, and all scripts already marked by him/her will either be allocated to a different marker or remarked by the Academic Associate for Marking.

That marker is given additional support (task-specific training or instructions and/or a standardisation exercise) until the Academic Associate for Marking is satisfied that s/he can mark in line with the standard.

Marking examiners (markers) are extensively trained in applying the LanguageCert assessment criteria, closely monitored by the Academic Associate for Marking, and retrained during annual standardisation events which ensure accuracy in the application of assessment criteria.

B2 Markscheme for Writing

	Task Fulfilment, Task1	Task Fulfilment, Task 2	Accuracy and range of grammar	Accuracy and range of vocabulary	Organisation
3	<ul style="list-style-type: none"> - fully addresses and communicates all 3 content points - genre and tone appropriate 	<ul style="list-style-type: none"> - fully addresses and communicates both content points - genre and tone appropriate 	<ul style="list-style-type: none"> - uses a range of simple and complex forms with control and flexibility - errors do not impede communication 	<ul style="list-style-type: none"> - uses a range of vocabulary, including less common items, appropriately - errors do not impede communication 	<ul style="list-style-type: none"> - text is well-organised and coherent, using a variety of cohesive devices - organization is fully appropriate to text type - few, if any, punctuation errors
2	<ul style="list-style-type: none"> -covers at least 2 content points with some expansion and communication mainly achieved - genre and tone mostly appropriate 	<ul style="list-style-type: none"> - covers both points with some expansion and communication mainly achieved - genre and tone mostly appropriate 	<ul style="list-style-type: none"> - uses simple and some complex forms with a good degree of control - errors do not impede meaning, but may cause re-reading 	<ul style="list-style-type: none"> - uses a range of everyday vocabulary accurately, with occasional misuse of less common items - errors do not impede meaning, but may cause re-reading 	<ul style="list-style-type: none"> - text is generally well-organised and coherent using a variety of linking words and cohesive devices - organization mainly appropriate to text type - some punctuation errors that don't impede communication
1	<ul style="list-style-type: none"> - communication of 2/3 points is minimally achieved OR - communication of only one point 	<ul style="list-style-type: none"> - communication is minimally achieved OR - communication of only one point 	<ul style="list-style-type: none"> - uses limited range of simple forms with control - errors impede meaning at times 	<ul style="list-style-type: none"> - uses everyday vocabulary generally appropriately, while overusing certain common items - errors with vocabulary and/or spelling impede meaning at times 	<ul style="list-style-type: none"> - text is connected using basic linking words and a limited range of cohesive devices - organization and/or paragraphing inappropriate - punctuation errors

0	- communication fails OR - off topic	- communication fails OR - off topic	- errors so serious that communication fails	- Vocabulary usage and/or spelling so poor that communication fails	- little, or no, organization or coherence
----------	--	--	--	---	--

Note: Dealing with 'problem' scripts

1. Short answers – if a response is under 80 words in Task 1, subtract one mark from Task Fulfilment.
2. Short answers – if a response is under 120 words in Task 2, subtract one mark from Task Fulfilment.

B2 Markscheme for Speaking

	Task fulfilment and coherence	Accuracy and range of grammar	Accuracy and range of vocabulary	Pronunciation, intonation and fluency
3	<p>Handles the interaction with ease and confidence.</p> <p>Relates own contributions, which are fully relevant to the conversation/topic, skillfully. Takes turns spontaneously and integrates relevant detail into the conversation.</p> <p>Addresses the topic well, appropriately highlighting significant points, and supporting with relevant detail. Produces clearly organised and coherent speech, in an appropriate style.</p>	<p>Wide range of the grammatical structures expected at B2, a consistently high level of grammatical accuracy.</p> <p>Occasional minor syntactical problems and non-systematic errors occur, which are usually corrected.</p>	<p>Wide range of vocabulary and a consistently high level of accuracy.</p>	<p>Pronunciation and intonation clear and natural.</p> <p>AND</p> <p>Language use is fluent and effective. Does not need to search for language.</p>
2	<p>Handles the interaction with relative ease, managing the conventions of turn-taking, using appropriate phrases, well, though not always very 'elegantly'. Expresses his/her message clearly, with relevant supporting detail where appropriate.</p> <p>Content/contributions are mostly relevant to the conversation/topic, and it is rare that any misunderstanding occurs.</p> <p>Uses a limited range of cohesive devices to link his/her utterances into clear, coherent discourse but there may be some</p>	<p>Sufficient range of the grammatical structures expected at B2, a relatively high degree of grammatical control.</p> <p>AND/OR</p> <p>Errors occur, but they do not impede communication.</p>	<p>Sufficient range of vocabulary to deal with the tasks at B2 with a degree of formality appropriate to the circumstances.</p> <p>AND/OR</p> <p>Some errors occur, but they do not impede communication.</p>	<p>Pronunciation reasonably clear and is understood without difficulty.</p> <p>AND</p> <p>Intonation is natural, has a fairly even tempo and contributes to the content.</p> <p>AND</p> <p>Few noticeably long pauses, although there may be some hesitation when searching for patterns and expressions.</p> <p>No undue strain on the listener.</p>

	'jumpiness' in long contributions.			
1	<p>Fails to meet most of the task requirements. Has difficulty in conveying the message, often relying heavily on the interlocutor to keep the interaction going or for support. Contributions are often irrelevant to the conversation or topic. Often misunderstands what is being said. Connects ideas relatively simply and may be difficult to follow in places.</p>	<p>Range of grammar too limited to deal with the variety of tasks at B2. AND/OR A number of errors, some of which impede communication.</p>	<p>Range of vocabulary too limited to deal with the variety of tasks at B2. This results in lack of detail and/or clarity. AND/OR Errors occur, some of which impede communication.</p>	<p>Undue strain on the listener because: influence of L1 is strong on pronunciation, intonation and/or stress patterns. OR A substantial part of the message is unclear. OR Frequent hesitation, repetition and attempts to repair language.</p>
0	<p>Clearly fails to meet the task requirements. Has difficulty linking ideas into connected speech. Is often difficult to follow. OR Insufficient sample of language.</p>	<p>The range of grammatical structures is extremely limited for B2 level. There are a lot of errors, which interfere with the meaning and most of the message is affected. OR Insufficient sample of language.</p>	<p>The range of language is so restricted that the person taking the test consistently finds it difficult to express him/herself. Errors impede communication and most of the message is affected. OR Insufficient sample of language.</p>	<p>Extreme strain on the listener because: Influence of L1 impedes most of the message. OR Continuous hesitation and/or reliance on the interlocutor. OR Insufficient sample of language.</p>

In both Speaking and Writing, people taking the test are awarded 0 to 3 marks for each of the four assessment criteria. Therefore, the maximum mark available across the two tasks of the writing section is 24, while the maximum mark available for the Speaking test is 12.

Marks are finally reported on a scale of 50 marks per skill (please see table below) to ensure equal weighing of all skills. Scaled scores are achieved through a simple conversion of the raw marks achieved per skill.

In order to pass the Listening, Reading, Writing examination, candidates need to achieve a total converted score of at least 75 out of 150 as per the table below. No minimum score per individual skill is required.

In order to pass the Speaking examination, Candidates need to achieve a total score of at least 25 out of 50 as per the table below.

Please note that the below are administered as two separate examinations and that separate certificates are issued for *Listening, Reading, Writing* and for *Speaking*.

IESOL skill	Maximum part marks	Total scaled score per skill	Total scaled score	Grades
Listening	26	50	150	Fail 0-74
Reading	26	50		Pass 75-100
Writing	24	50		High Pass 101-150

IESOL skill	Maximum marks	Total scaled score	Grades
Speaking	12	50	Fail 0-24 Pass 25-33 High Pass 34-50

Reporting scores on LanguageCert International ESOL as a whole

Scores on LanguageCert International ESOL Communicator B2 are reported to centres and people taking the test in the form of a certificate and a Statement of Results (SoR). The Certificate demonstrates the final grade achieved by the candidate (e.g. Pass), whereas the SoR provides feedback per skill assessed (e.g. Listening: Standard not met).

The overall grades awarded are either High Pass, Pass or Fail.

Overall scores on LanguageCert International ESOL Communicator B2 (Listening, Reading, Writing) are calculated by adding the raw marks achieved by the candidate per skill and converting them into the scaled score as per the table above.

Overall scores on LanguageCert International ESOL Communicator B2 (Speaking) are calculated by converting the total raw marks achieved by the candidate into the scaled score as per the table above.

Pass marks on LanguageCert International ESOL Communicator B2 (Listening, Reading, Writing) were initially set using a modified Angoff standard setting procedure. Judges were asked to estimate the cut score based on a definition of the minimally competent candidate at level B2 (defined for each skill area prior to the event). Data were used to support the judgements (made over two rounds). The cut score for a High Pass was set at a point where the candidate can be said to have met some of the criteria for the next highest level. The cut off scores were initially provisionally calculated by expert validation and confirmed with a concurrent validity test during the City & Guilds B2 Communicator Linking Project⁵. The level of the cut off scores will continue to be monitored through ongoing analysis of the live tests and once the LanguageCert Item Difficulty (LID) scale is fully developed.

Pass marks on LanguageCert International ESOL Communicator B2 (Speaking) were initially set using a modified Angoff standard setting procedure. Judges were asked to estimate the cut score based on a definition of the minimally competent candidate at level B2 (defined for each skill area prior to the event). Data were used to support the judgements (made over two rounds). The cut score for a High Pass was set at a point where the candidate can be said to have met some of the criteria for the next highest level. The cut off scores were initially provisionally calculated by expert validation and confirmed with a concurrent validity test during the City & Guilds B2 Communicator Linking Project. The level of the cut

⁵ O' Sullivan, Barry, *The City & Guilds Communicator examination linking project: a brief overview with reflections on the process* in Martyniuk, W. (ed.) *Studies in Language Testing. Aligning Tests with the CEFR: Reflections on using the Council of Europe's draft Manual*. Cambridge University Press: 2010

off scores will continue to be monitored through ongoing analysis of the live tests and once the LanguageCert Item Difficulty (LID) scale is fully developed.

Official results are issued within 10 business days of receipt of the examination materials and Close-out Report for an exam session from the exam centre. Upon issuance of results, LanguageCert sends an e-mail to people taking the test with their result and an e-certificate. People taking the test can also access their Statement of Result and download their e-certificate if they wish through their LanguageCert account via LanguageCert's exam management system, PASSPORT. LanguageCert also sends a full report of each individual's test results to the Centre, including results analysis. The report is also available through LanguageCert's exam management system, PASSPORT.

Certificates

Once the official exam results are issued, people taking the test can log in to their profile, view their results and download their e-certificate.

Hard-copy certificates that have been ordered will be issued and shipped via a standard postal or UPS service to either the people taking the test, the venue or the Centre (the Centre is responsible for distributing the certificates to people taking the test) within five (5) business days from the date results were issued. All information on the certificates is in English, while the name of the person taking the test is in their native language.

In addition, and to ensure the security of its certificates, LanguageCert offers a unique online certification verification service, through which third parties can verify the validity of a certificate through the LanguageCert website.

Feedback reports

People taking the test unsuccessfully will automatically receive a short feedback report designed to prepare them to retake an examination. This consists on feedback on their performance in each skill and assessment criterion.

Assessment Results and Analysis

Test response data is routinely analysed for LanguageCert International ESOL B2 Communicator.

Reliability relates to consistency in test results and this is achieved in the LanguageCert International ESOL tests by ensuring test forms are comparable in terms of content and difficulty, and through robust item-banking techniques, involving the pretesting, live testing and trialling of test materials and the placement of all items on the LanguageCert Item Difficulty (LID) scale. The Language Assessment Development Manager is responsible for collecting all data.

Indicatively, for the three most recent exam sessions with a test population of over 80 candidates, the mean scores were:

Session 1 (109 examinees)

Mean scores

All (objectively scored) items: 39.725 (raw marks) [SD: 9.426] 76.391 (converted score/100); mean P: 0.794

Listening: 20.257(raw marks) [SD: 6.971], 38.954 (converted score/50); mean P: 0,778

Reading: 19.468 (raw marks) [SD: 6.160], 37.436 (converted score/50); mean P: 0,755

Writing: 18.461 (raw marks) [SD: 4.654], 38.472 (converted score/50); mean P: 0,750

Reliability Figures (alpha)

All (objectively scored) items: 0.921

Listening: 0.758; Reading: 0.905

Session 2 (91 examinees)

Mean scores

All (objectively scored) items: 31.868 (raw marks) [SD:10.042], 61.282 (converted score/100); mean P: 0,635

Listening: 16.846(raw marks) [SD: 5.381], 32.394 (converted score/50); mean P: 0667

Reading: 15.022 (raw marks) [SD: 5,364], 28,887 (converted score/50); mean P: 0,603

Writing: 17,021 (raw marks) [SD: 4,125], 35,471 (converted score/50); mean P: 0,653

Reliability Figures (alpha)

All (objectively scored) items: 0.904

Listening: 0.845; Reading: 0.824

Session 3 (249 examinees)

Mean scores

All (objectively scored) items: 35.494 (raw marks), 68,254 (converted score/100); mean P: 0,728

Listening: 20.257(raw marks) [SD: 6.971], 38.954 (converted score/50); mean P: 0.740

Reading: 19.468 (raw marks) [SD: 6.160], 37.436 (converted score/50); mean P: 0.716

Writing: 17,189 (raw marks) [SD: 4.654], 38.472 (converted score/50); mean P: 0.708

Reliability Figures (alpha)

All (objectively scored) items: 0.866

Listening: 0.786; Reading: 0.774

Reliability is crucial for all test stakeholders who need to be sure that different administrations of the test deliver very similar results. This is essential for fairness to test-takers and to ensure that receiving institutions such as universities and employers can be guaranteed that the same ability level is required to pass the same examination at different administrations.

The start of the process of ensuring reliability of results is to ensure standardisation of the test-taking experience. This begins with test specifications that ensure tests can be replicated over years of administrations, through standardised test-taking conditions and finally through the difficulty of the test materials and the way tests are graded.

Specifications and robust standardised item-production techniques permit a constant supply of new test items into the item bank. Harmonised procedures for test day administration are provided to test centres. Item-banking techniques, using statistical analysis provided from pre-tests and live tests ensure empirically that tests are always of very similar difficulty.

There are always some sources of 'error' in a test score. Such sources of error can include, for example, the test version, the test-taker and how they respond on the day of the test, the test-taking environment and the test marker etc. LanguageCert ensures these sources of error are kept to the absolute minimum.

The use of item-banking techniques is central to ensuring reliability of test results. The LanguageCert item-banking system is an electronic system that permits the storage of large amounts of test items which all have attributes relating to difficulty level, discrimination, testing focus, topic etc. Rigorous quality assurance procedures include the pretesting of all objective items for the Reading and Listening tests. Response data from Pretesting is captured and analysed using advanced statistical techniques including Rasch analysis (a branch of Item Response Theory). Through the utilisation of 'anchor' materials which have known measurement characteristics, all new items in the pre-test can be measured

and given a difficulty value. This process of calibration is one way in which LanguageCert ensures reliability of test results and fairness to students.

Having an item-difficulty scale enables LanguageCert to produce tests of the same, or very similar, difficulty across multiple test administrations. If there are any slight differences in difficulty of test forms this is accommodated in the grading process by, for example, setting the pass mark slightly lower for a slightly more difficult test. Ability estimates are used in this grading analysis process. All LanguageCert objective tests can be pre-graded in this way.

For Writing and Speaking, assessment scales have been developed to ensure the tests are referenced to the CEFR. These are criterion-referenced tests and are calibrated against the CEFR descriptors for each CEFR level. The marking scales across the levels A1-C2 ensure a common scale of increasing difficulty mapped against CEFR can-do statements.

Inter-Rater Reliability Alphas for IESOL Examiners

Reliabilities were calculated for each of the IESOL levels using a free online tool called Recal available here <http://dfreelon.org/recal/recal-oir.php> to calculate Krippendorff's alpha for each of the data sets.

Krippendorff's alpha (α) was chosen because it "is a reliability coefficient developed to measure the agreement among observers, coders, judges, raters, or measuring instruments drawing distinctions among typically unstructured phenomena or assign computable values to them. α emerged in content analysis but is widely applicable wherever two or more methods of generating data are applied to the same set of objects, units of analysis, or items."

Cohen's kappa coefficient (κ) is often used for inter-rater reliability studies but is typically used when only two raters are involved. As multiple raters were used in the IESOL study, Krippendorff's alpha was considered more suitable. Alphas above .8 are considered excellent, and above .7 to be acceptable. Given the variability of the alphas reported below, the team may wish to consider some qualitative research at a later stage using think-aloud protocols where raters say what they are doing as they rate tasks. These think-aloud protocols are recorded and transcribed and analysed to identify rater strategies. e.g. of consistent versus aberrant raters.

Results

A1

Raters: 26

Scripts: 6

N decisions:156

Krippendorff's alpha (interval): 0.839

A2

Raters: 26

Scripts: 6

N decisions:156

Krippendorff's alpha: 0.752

B1

Raters: 22

Scripts: 6

N decisions:132

Krippendorff's alpha:0.84

B2

Raters: 22

Scripts:6

N decisions:132

Krippendorff's alpha: 0.707

C1

Raters: 22

Scripts: 6

N decisions:132

Krippendorff's alpha: 0.74

C2

Raters: 22

Scripts: 6

N decisions:132

Krippendorff's alpha: 0.77

