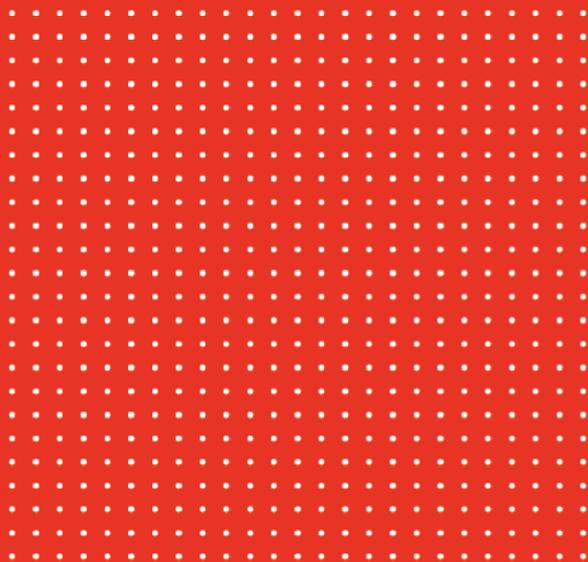# Language Cert

External Validation of LanguageCert's English Language Examinations

Yiannis Papargyris
and
Leda Lampropoulou

Introduction

LanguageCert began administering its own English language examinations in 2017. Since that time LanguageCert has undergone a series of external validations to provide evidence of the robustness of its examinations, and to provide proof of their validity, reliability and their being fit for purpose. Two key external studies referred to below are those conducted by:

CRELLA – the Centre for Research in English Language Learning and Assessment at the University of Bedfordshire.

In 2018, CRELLA was commissioned to investigate LanguageCert 's B2 level test and its relationship to the CEFR.

NARIC – the National Recognition Information Centre for the United Kingdom, the Government agency for the recognition and comparison of International qualifications and skills.

In 2018, UK NARIC was commissioned to conduct an independent assessment of LanguageCert's ESOL examinations.

The current document first provides a brief outline of key concepts to situate the external studies. Following this, the key issues, major conclusions and recommendations from the external studies are presented.

At a basic level, the Framework has been useful in the setup of the LanguageCert website, with the categories driving the ways users navigate through the website. External validation has already been conducted on certain examinations in the IESOL suite (a comprehensive evaluation of the B2 test, with shorter evaluations of other tests); the relevant reports are then located under appropriate Framework headings. The makeup of the Framework also allows for sensible location of internal validation documentation such as marker standardization data and comparability data on different tests forms; similarly, background documentation such as Item Writer Guidelines, Marker Guidelines – which LanguageCert makes transparently available – fit cleanly into the Framework.

Validity and the LanguageCert IESOL Examinations

The LanguageCert International ESOL tests are designed to ensure fitness for purpose and to deliver assessments which take into account contemporary views on validity. Validity is generally defined as the extent to which a test measures the intended purpose. In the case of LanguageCert, this is communicative language ability (see e.g., Bachman & Palmer, 2010) and the foreign language specifications provided by the Council of Europe in such documents as Waystage (Ek & Trim, 1997) and Threshold (Ek & Trim, 1991). The qualities of validity (and reliability) need to be considered together in order to ensure fairness to candidates and to generate trusted result outcomes that will replicate real-world performance of candidates.

The test development process underpinning LanguageCert's English language exams has been established to ensure validity is achieved. Bachman & Palmer (2010) states that language tests should support inference to some domain of 'Target Language Use' (TLU). That is, in order to judge the validity of test results, what a test-taker is expected to be able to do in real-world language use must be laid out. The Common European Framework of Reference (CEFR) has been utilized to help determine the test construct of the LanguageCert exams for this purpose. Its illustrative descriptors across a range of language domains and contexts have been used as a starting point and extensively inform the test development processes employed.

The task types used in the LanguageCert examinations have been selected to ensure they have interactional authenticity and can be related to real-world performance. They directly sample the cognitive skills, strategies and language knowledge that support inference about the potential ability of a candidate in real-world interactional situations.

In this manner, validity links performance on the tasks in LanguageCert International ESOL tests to an inference about the test taker's ability in a world beyond the test. The tests are designed to elicit a sample of performance which is interpretable and generalizable to the real world. In order to ensure the test results are generalizable CEFR Can-Do statements have been used as the basis for what test-takers need to be able to achieve at each level.

Achieving Reliability

Reliability relates to consistency in test results. 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 and trialling of test materials and the placement of all items on the LanguageCert Item Difficulty (LID) scale.

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 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.

Historical LanguageCert-related Validation Study

LanguageCert acquired a range of IESOL test materials from City & Guilds, UK in 2015. Prior to being acquired by LanguageCert, the quality of some of the City & Guilds examinations was put through significant external validation, the most prominent study being:

O'Sullivan, B. (2009). City & Guilds Communicator Level IESOL Examination (B2) CEFR Linking Project Case Study Report. Roehampton University, UK.

This study was thorough and extensive, and reported positive outcomes regarding the makeup of the City & Guilds B2 level test, supporting the claims about the test's links to the CEFR.

The Executive Summary may be found in Appendix 1. The full report may be accessed on the LanguageCert website.

Recent External Validation Studies of LanguageCert tests

Since administering its own English language examinations in 2018, two large-scale validation studies of LanguageCert examinations have been conducted. These have been:

Validation Study 1	**Green, A. 2019. Relating LanguageCert Communicator to the CEFR. Centre for Research in English Language Learning and Assessment: University of Bedfordshire, UK.**

This thorough and extensive examination of the B2 test was conducted by CRELLA, the Centre for Research in English Language Learning and Assessment at the University of Bedfordshire.

While there were some recommendations for LanguageCert to consider, the findings from the study strongly support the claim that material throughout the Spoken and Written Exams closely reflected the B2 level.

The Executive Summary may be found in Appendix 2. The full report may be accessed on the LanguageCert website.

Validation Study 2	**National Recognition Information Centre for the United Kingdom. 2019. LanguageCert ESOL International Qualifications: Independent CEFR Referencing - Summary Report. UK NARIC: Cheltenham, UK.**

The National Recognition Information Centre for the United Kingdom (NARIC), the UK Government agency for the recognition and comparison of International qualifications and skills was commissioned in 2018 to conduct an independent assessment of LanguageCert's ESOL examinations. The independent assessment was a mandatory pre-requisite, in order for organisations to be eligible to participate in a UK government procurement for English language testing.

In its evaluation, UK NARIC recognised that LanguageCert's IESOL qualifications had been developed with CEFR as the source document. The extensive evaluation deemed that the item writing process underwent clear technical and content checks. Following successive investigations into standard setting, vetting, statistical analysis, modification, proofreading and finalisation, UK NARIC determined that CEFR alignment was evident at all stages of test development and delivery.

The Executive Summary may be found in Appendix 3.

References

Bachman, L. F., & Palmer, A. S. (2010). Language assessment in practice. Oxford: Oxford University Press.

Van Ek, J. A., & Trim, J. L. M. (1991). Threshold level 1990. Council of Europe.

Van Ek, J. A., & Trim, J. L. M. (1997). Waystage 1990. Cambridge University Press.

Appendix 1

O'Sullivan, B. 2009. City & Guilds Communicator Level IESOL Examination (B2) CEFR Linking Project Case Study Report. Roehampton University, UK.

Executive Summary

Background

This project was a joint undertaking by City & Guilds and the Centre of Language Assessment Research (CLARe) at Roehampton University. The object of the project was to provide evidence of the validity of City & Guilds' Communicator examination, particularly in relation to the central claim that it is aimed at Level B2 in the Common European Framework of Reference for Languages (commonly referred to as the CEFR). In doing this, it was planned that the project would act as a formal review of the existing examination, and it was planned that any areas of concern within the papers would be identified and brought into line with best practice in the area.

The Communicator (and the other examinations in the suite) was developed using the CEFR (Council of Europe 2001) as source document to inform the assessment tasks, specifications and assessment criteria. During the development phase, however, the Draft Manual (2003) for relating examinations to the framework was not in existence, so the organisation embarked on a series of internal activities to ensure alignment to the external standards. However, with the publication of the Manual the logical step for the organisation was to register as a case study for operationalising the concepts and processes encapsulated there.

A secondary aim of the project was to provide feedback to the Council of Europe on their Draft Manual (2003) which was used as a basis for the methodology.

Methodology

As mentioned earlier, the methodology used in the project was based on the procedures recommended by the Council of Europe in their Draft Manual of 2003. However, as the project progressed a number of changes were made to facilitate the operationalisation of the process. The project adapted the four-stage approach suggested in the Draft Manual:

1. Familiarisation

2. Specification

3. Standardisation

4. Validation

In terms of the methodology used, a number of important recommendations were made, these related to the nature of the process (which we suggest is iterative rather than linear as implied in the Draft Manual) and the notion of embedding the process in the institution's test development cycle.

Summary of the Main Findings

The main findings of the project can be summarised as follows:

1. It was found that in order to claim a link to the CEFR at Level B2 the cut score for a passing grade for the Communicator Reading paper should be set at 15 (from a maximum of 30). The same cut score was recommended for the Communicator Listening paper. This is actually in line with current practice for Communicator.

2. Passing levels for the Communicator Writing paper were found to be in line with the Council of Europe recommended tasks for CEFR Level B2. The recommendation is that the cut level for this decision should not be altered at this point in time.

3. The linking process is long and demanding, both at the individual and institutional level. The complexity of the design means that it is expensive for any institution to undertake, certainly to the extent undertaken by City & Guilds in this project. While this perhaps explains the reluctance of many examination boards to undertake a full linking project, we nevertheless recommend that the process be extended to as many of the other examinations in the ESOL suite as feasible.

4. Unless the test which is the focus of the linking project is shown to be robust in terms of quality and level, there is no point in even starting a linking project, as the process is unlikely to succeed beyond the standardisation stage without serious issues emerging. In fact, we feel that with a more demanding specification phase, issues should emerge more clearly at this early stage.

5. Limiting the validation evidence to estimates of internal and external validity is far too simplistic a view of validation. The CEFR should be demonstrated to impact on all aspects of the test, from the test taker to the task to the psychometric qualities and relative meaning or value of the test score.

Based on this project, it is the belief of the project team that the evidence presented here supports the claim that the Communicator tests English ability at CEFR Level B2.

We feel that the process of linking the Communicator examination to the CEFR, has resulted in systematic and sustainable improvements to the test and to the system that supports the test.

It is clear to us that the process has resulted in a test that is more clearly at level, is sound from an internal psychometric perspective and is more replicable and of a high quality. However, that is not all. The systems that support the examination have also been systematically improved and more explicitly linked to the CEFR. The item writers' guidelines are, we believe, up-to-date and more robust than in the past. The specifications are now more likely to result in accurate replication of tests on level – one criticism of the old specification was the lack of detail and exemplification, this appears to have led to a tendency to drift away from the level. This is a warning for other test developers, who take time to specify their tests but do not routinely review these specifications (and their use) to ensure that there is no level or construct drift.

We now feel that we are in a position to consider suggesting a number of Communicator tasks to the Council of Europe for use as recommended level indicators in future linking projects.

Appendix 2

Green, A. 2019. Relating LanguageCert Communicator to the CEFR. Centre for Research in English Language Learning and Assessment: University of Bedfordshire, UK.

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.

LanguageCert ESOL International Qualifications: Independent CEFR referencing

UK NARIC's independent review of the LanguageCert ESOL International Qualifications has found that the tests align with the Common European Framework of Reference for Languages (CEFR) at the following scores:

CEFR level	LanguageCert Qualification	Reading	Writing	Listening	Speaking
A1	LanguageCert Entry Level 1 Certificate in ESOL International (Listening, Reading, Writing) (Preliminary A1)	33	33	33	
A1	LanguageCert Entry Level Certificate in ESOL International (Entry 1) (Speaking) (Preliminary A1)				33
A1	LanguageCert Entry Level Certificate in ESOL International (Entry 1) (Speaking & Listening) (Preliminary A1)			18	
A2	LanguageCert Entry Level 2 Certificate in ESOL International (Listening, Reading, Writing) (Access A2)	33	33	33	
A2	LanguageCert Entry Level Certificate in ESOL International (Entry 2) (Speaking) (Access A2)				33
A2	LanguageCert Entry Level Certificate in ESOL International (Entry 2) Speaking & Listening) (Access A2)			18	
B1	LanguageCert Entry Level 3 Certificate in ESOL International (Listening, Reading, Writing) (Achiever B1)	33	33	33	
B1	LanguageCert Entry Level Certificate in ESOL International (Entry 3) (Speaking) (Achiever B1)				33
B1	LanguageCert Entry Level Certificate in ESOL International (Entry 3) (Speaking & Listening) (Achiever B1)			18	
B2	LanguageCert Level 1 Certificate in ESOL International (Listening, Reading, Writing) (Communicator B2)	33	33	33	
B2	LanguageCert Level 1 Certificate in ESOL International (Speaking) (Communicator B2)				33
C1	LanguageCert Level 2 Certificate in ESOL International (Listening, Reading, Writing) (Expert C1)	33	33	33	
C1	LanguageCert Level 2 Certificate in ESOL International (Speaking) (Expert C1)				33
C2	LanguageCert Level 3 Certificate in ESOL International (Listening, Reading, Writing) (Mastery C2)	33	33	25	
C2	LanguageCert Level 3 Certificate in ESOL International (Speaking) (Mastery C2)				33

In November 2018, UK NARIC completed an independent evaluation and CEFR referencing exercise of fifteen LanguageCert IESOL tests – presented in the figure above. The English language tests, offered by PeopleCert Qualifications Ltd, have been developed in order that applicants can demonstrate skills for further study, settlement and / or employment.

EXECUTIVE SUMMARY

Qualification Overview and Test Format

The qualifications target different skills. Each qualification has been designed to test at a particular CEFR level covering reading, writing, and listening. Candidates can also take an additional paper in speaking available at each CEFR level in order to demonstrate competency across all four skills. For CEFR A1, A2, and B1 there is also the option of taking a standalone speaking and listening qualification.

Reading and Writing

In the three skills tests, reading and writing are combined, and the test is divided into four parts. Candidates are expected to handle a range of text types. At the lower CEFR levels candidates are tested on their reading skills using shorter texts such as postcards or emails relating to everyday tasks and real-life communication, such as the arrangement of a social event (A1, A2, B1) with longer texts such as factual based articles introduced at B1 with increasing complexity of content and abstractedness of topics seen at higher levels (B2, C1, C2).

Topics increase in complexity at the higher levels, including abstract ideas (B2, C1, C2). Input texts cover a range of grammatical structures, vocabulary, and functions appropriate to the levels. Multiple choice, gapfill, and short answer questions are used to test understanding. Candidates need to demonstrate the ability to understand general meaning, as well as specific details. Inferencing, synthesis, and understanding of structures such as contrasting arguments, cause-effect and problem/solution, paraphrasing and use of some literary devices are required at the higher levels with language increasing in abstractedness and technicality in extended texts (B2, C1, C2).

The writing is typically divided into two parts. Some tasks test integrated skills such as requiring a candidate to read an input text and respond to it. Expected output texts reflect different genres and therefore elicit different subskills. Some creative texts are introduced at higher levels (B2, C1, C2), such as descriptive composition.

At A1, output texts are typically around 20-30 words, increasing to c.70-120 (B1), c.100-150 (B2), c.150-200 (C1), to 200+ (C2). Candidates are expected to demonstrate skills appropriate to the levels, such as use of a range of grammar, vocabulary, functions, and appropriate structuring. Higher levels require students to more specifically select informal or formal language, demonstrate the use of colloquial or idiomatic language, to develop and justify arguments, complain, exemplify and respond to counterarguments.

Listening

The listening test is approx. 20 mins (A1, A2) increasing to 30 mins for those taking the B1, B2, C1 and C2 certificates. It is divided into four parts, assessed using multiple choice questions, note-taking, and short answer responses. A range of text types are used, increasing in complexity to reflect proficiency levels, such as including a greater level of technicality in terms of topics and language used. At lower levels, candidates may need to identify overall meaning, related to personal or everyday topics, or specific details (A1, A2, B1). More abstract topics and extended texts are a feature of the higher levels (B2, C1, C2), which may require candidates to identify lines of argument, attitudes, recognise hedging, consistently identify implicit meaning, or follow circumventions (B2, C1, C2). Input texts cover a range of grammatical structures, vocabulary and functions appropriate to the levels.

At the lower levels, the audio texts are slower than natural speed though with some natural features of speech such as hesitations (A1, A2, B1). Standard accents are used, with the variety of accents increasing with the higher levels (B2, C1, C2). Complex, idiomatic, academic, technical, and colloquial language is seen most consistently at higher CEFR levels (B2, C1), with a specific increased focus on connotation at C2.

Speaking

The speaking test is divided into four parts and conducted with an interlocutor - as a result, candidates also need to demonstrate interactive skills. At lower levels this takes 6 minutes, increasing to 12 minutes (B1), 13 (B2), and 15+ (C1, C2). Tasks range from some short questions about the candidate, role plays, a discussion and a short monologue on an unseen topic with follow up questions. At lower levels, the focus is on personal or everyday life topics and requires students to take turns, describe, and ask / answer factual or concrete questions.

Tasks allow candidates to demonstrate their ability to initiate conversation, take turns, develop points, and use a range of grammar, vocabulary and functions across different topics, appropriate to the CEFR levels tested. At higher levels, more abstract questions are used (B2, C1, C2). Within these tasks, candidates need to be able to hypothesise, express abstract ideas, speculate, emphasise, negotiate, and express shades of opinion with flexibility with an understanding of sociolinguistic context in increasingly extended turns.

Speaking & Listening qualifications

Available at A1, A2 and B1, these standalone tests include integrated skills and last approx. 9 minutes (A1), 12 minutes (A2) and 15 minutes (B1). They are divided into four parts. The test covers question and answer interaction, role plays, and a discussion-based task leading to a monologue/presentation style task. Topics relate to personal, public or educational domains. Interlocutors support interaction, including through use of follow-up questions. A range of grammatical structures, vocabulary, and functions are used and expected, appropriate to each CEFR level.

Grading and Certification

Reading and listening questions are typically machine marked with some human marked to a mark scheme. Writing tasks are marked across task fulfilment; accuracy and range of grammar; accuracy and range of vocabulary; organisation. Speaking is marked against criteria including task fulfilment and coherence; accuracy and range of grammar; accuracy and range of vocabulary; pronunciation, intonation and fluency.

Speaking and Listening standalone qualifications are marked across listening and responding; interactive communication and task fulfilment; accuracy and range of grammar; accuracy and range of vocabulary; pronunciation, intonation, and fluency.

Candidates receive a statement of results (SoR) including exam level, overall achievement (high pass, pass, fail) and numerical results per skill. A results and certification verification service is available.

CEFR global scale[1]

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

---

[1] The Common European Framework of Reference for Languages Global Scale. Available at: https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale.
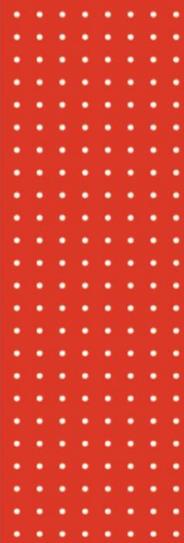
Language
Cert
languagecert.org