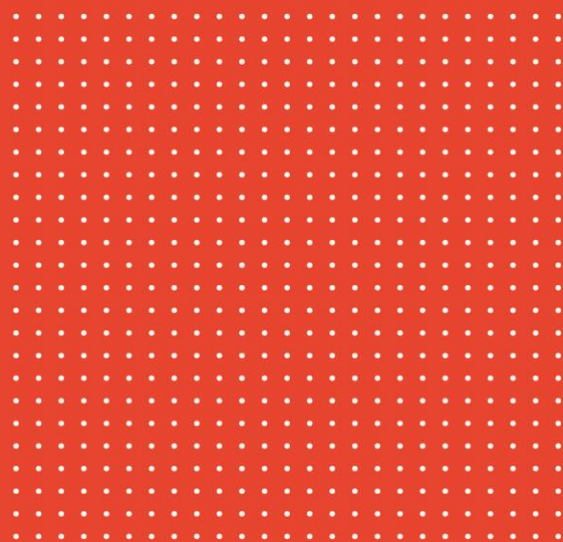


Language
Cert



Cathy Jones

The LanguageCert
General Test:
Assessing
Language in the
Migration and
Employment
Domains



Abstract

Following Jones' (2023) paper which outlined the positioning of the *LanguageCert Academic* IESOL C1 examination to an academic context, the current paper describes the development of *LanguageCert General*, a counterpart qualification to *LanguageCert Academic*, which addresses the migration employment domain. *LanguageCert General*, which is closely based on the pre-existing LanguageCert IESOL B2 test, is a four-skill, multi-level test, aligned to a common underlying measurement scale, derived from a bank of pretested and calibrated assessment material and associated validation research based on an established candidature. The current paper highlights underpinning research, evidence and best practice which have informed the development and definition of a high-stakes relevant, reliable and secure test for migration purposes. It covers test purpose and construct, proficiency levels, task selection, test content, assessment criteria, test delivery, results and an integrated learning ecosystem.

Keywords: test design, test purpose, test content, washback, integrated learning ecosystem

Background

As a leading provider of language examinations and qualifications recognised by universities, employers and governments around the world, LanguageCert designs its examinations such that they assess language skills in a real-world context, using tasks and materials that are relevant to candidates' specific needs and goals. LanguageCert ensures that the CEFR is embedded into the test development cycle and the quality and level of test materials reflect this – providing an international standard for assessing language proficiency.

The LanguageCert English language portfolio includes a range of established, recognised, successful, high-stakes qualifications, including: LanguageCert International English for Speakers of other Languages (IESOL), a level-specific suite of examinations, ranging from levels A1 to C2 in the Common European Framework of Reference for Languages (CEFR) for both occupational and personal use. The portfolio also includes the LanguageCert Test of English, a multi-level linear and adaptive test of English in the workplace, as well as a suite of secure level-specific IESOL SELT (Secure English Language Test) qualifications, using IESOL examination structures, tasks, and items. The IESOL SELT qualifications meet the specific requirements of the UK Home Office as proof of English language competence for visas and immigration for life, work or study visa types (see <https://www.gov.uk/guidance/prove-your-english-language-abilities-with-a-secure-english-language-test-selt>).

As outlined in Jones (2023), in 2020, *LanguageCert General* (LCG) and its counterpart qualification, *LanguageCert Academic* (LCA) were conceived as a dynamic response to changing markets and stakeholder expectations. As a result, work began to extend the portfolio with two high-stakes tests: one for the academic sector and one for those wanting to migrate for work or training in an English-speaking environment. Both tests are derived from the LanguageCert item bank and report scores across relevant levels on the same measurement scale that is used for all LanguageCert – the Global Scale. Global Scale scores are reported for the four skills, Listening, Reading, Writing and Speaking; and the overall result. The focus of LanguageCert General is general English language proficiency for adults. It is designed to measure various aspects of language proficiency to support language policymaking and decision-making by governmental institutions, authorities and employers. One of the main outcomes of the evolution of the existing IESOL B2 and C1 tests into the LanguageCert General and LanguageCert Academic is to enable domain-specific measurement and certification across a broader range of relevant language attainment levels. This meets growing demand from different stakeholders (candidates, recognising institutions, and educational and business authorities) for more breadth in the areas that single level examinations assess. The multilevel format offers practical advantages, particularly in the context of migrants, accommodating an inclusive range of candidates with varied language backgrounds and experiences.

A phased roll out of LCG and LCA began in 2022 to ensure that all issues related to the effective delivery of the examinations could be addressed. A gradual roll-out (Phase 1) was planned to ensure not only a smooth introduction of the revised examinations but also to avoid confusion with existing IESOL SELT examinations used for UK visas and immigration (UKVI). LCA and LCG have been designed to replace the four single-level tests, already in use by UKVI, before the end of 2023. Phase 2 of the rollout took place in early 2023 when LanguageCert General and Academic were made more widely available in a large number of test centres managed by Prometric and PeopleCert.

Purpose

This paper describes the methodology for refocusing LanguageCert IESOL B2 responsively as part of the LanguageCert continuous test development and review cycle. It also provides evidence for test users of how ongoing research informs best practice and how it can be applied to test development.

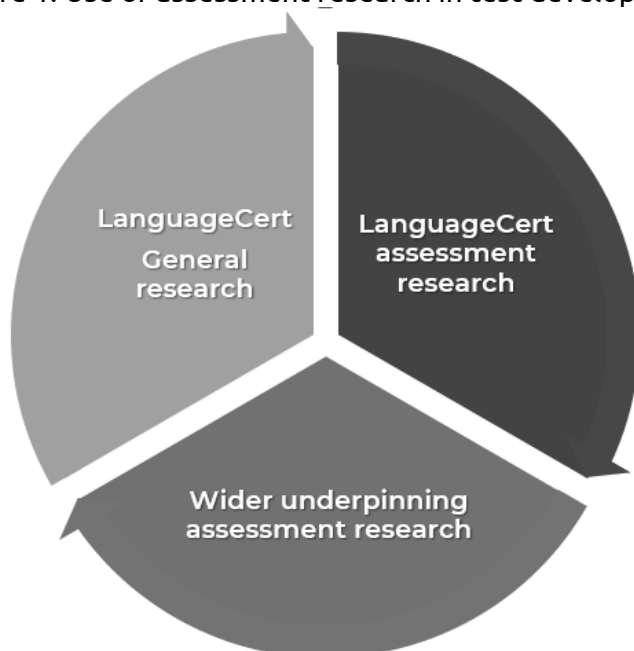
An Evidence-informed Approach

The LanguageCert General test development references a portfolio of research and validation covering three main areas:

1. Wider underpinning research into assessment, learning and teaching – evidence which is referred to below.
2. Research and validation on the wider portfolio of LanguageCert qualifications carried out both by the LanguageCert research team and external research (e.g., conducted by CRELLA, UK NARIC (now UK ENIC), etc).
3. Research undertaken by the LanguageCert research team with specific reference to LCG

Figure 1 below shows how these different bodies of research draw on and feed back into each other in an ongoing reciprocal cycle. Qualification development draws on research undertaken by LanguageCert, as well as the underpinning body of wider assessment research. The qualification-specific research generated for LCG feeds back in turn to the wider assessment landscape and informs future LanguageCert products as well as the wider development of how assessment of this kind can be used to develop products to support international progression and mobility.

Figure 1: Use of assessment research in test development at LanguageCert



Summary of underpinning evidence

The LCG test is designed to measure the English language skills and abilities of individuals who migrate to an English-speaking country. It evaluates language skills for various purposes, including immigration, employment, education and social integration. The LanguageCert B2 test has been widely used since 2017 and was fine-tuned in 2019 based on requirements set by the UK Home Office's Visas and Immigration authority.

In terms of underpinning evidence, for this development LanguageCert drew mainly on the levelled specifications of language needs that complement the Common European Framework of Reference for Languages (Council of Europe 2001, 2018) and the CEFR itself. LanguageCert's development team were also cognisant of publications such as Brindley and Burrows' (2000) *Studies in Immigrant English Language Assessment*.

Publications related to the CEFR used in the development of the LanguageCert General test include:

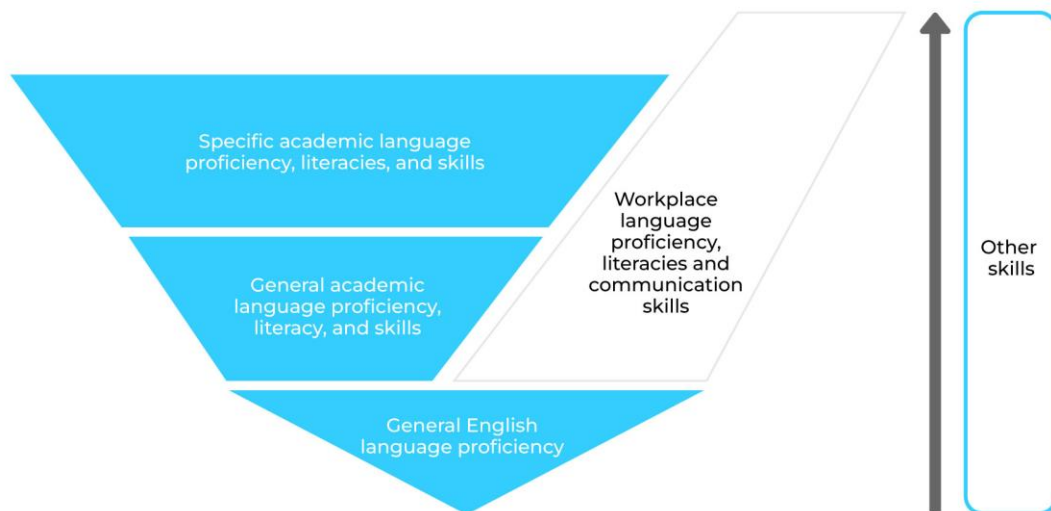
- The Common European Framework of Reference for Languages, Council of Europe (2001, 2018)
- A2: van Ek, J.A. and Trim, J.L.M. (1990b/1998b) *Waystage* 1990. Cambridge University Press.
- B1: van Ek, J.A. and Trim, J.L.M. (1990a/1998a) *Threshold* 1990. Cambridge University Press.
- B2 and above: van Ek, J.A. and Trim, J.L.M. (2001) *Vantage*. Cambridge University Press.

Consideration was also given to the performance descriptors as well as findings and recommendations of the ALTE reports *Linguistic integration of adult migrants: requirements and learning opportunities* (2018). as well as *Language tests for access, integration and citizenship: an outline for policy makers* (2016).

What is General English and Why is it Important?

Knoch (2021) proposes a useful model of what language proficiency entails in a post-secondary context. At its foundation, the model includes general English language proficiency (e.g. Bachman, 1990; Bachman and Palmer, 2010). Alongside these general and specific academic English proficiencies sits a component that Knoch (2021) labels workplace proficiency, literacies and communication skills.

Figure 2: Language proficiency in a post-secondary context (after Knoch, 2021)



The LCG is a test of the underpinning building blocks of general English language proficiency which include some elements of general workplace language, proficiency, literacies and communication skills as well as some elements of language for further study and training.

Defining the Target Language Use Domain

The conceptual model in Figure 2 above illustrates the connections that shape LanguageCert's approach to language assessment, and the position of learning and preparation materials within these connections.

International migration takes place in many different contexts for many different reasons, from entry for work and study purposes, family reunion and entry as an asylum seeker or refugee. The linguistic requirements for migrants may vary, depending on the migrant "journey" – a journey whose stages are described by Saville (2009) in the context of assessment in the management of international migration.

In the context of migration, language skills are vitally important in securing some fundamental human rights: fostering social inclusion, access to education, employment, healthcare and housing. Defining the target language use domain involves establishing the real-life linguistic demands on migrants and deciding if and how these be measured in a valid test design.

Definition of the LanguageCert General target language test domain must contain more detail than a test solely designed “for migration” and this is intrinsically linked to the test purpose. The LanguageCert General test is for candidates seeking to migrate for work or vocational purposes. It can also be used to measure the language competence required for effective social interaction. As a high-stakes test, it can also be used to make decisions regarding immigration, right to remain and the acquisition of citizenship.

The test measures a range of skills and competences appropriate for personal, occupational and vocational contexts: reading and listening for gist and detailed understanding of a range of written and audio sources including adverts, articles, websites, diaries, radio programmes, and podcasts. The test includes writing in formal and informal registers, expressing viewpoints and taking part in role plays in real-life scenarios. In total, LanguageCert General is tailored to those wanting to live, work, study or train in an English-speaking context.

The focus on domains, and the target language use within them, permeates all aspects of test design, development, and delivery. This includes how LanguageCert ensures that candidates are supported with examination-specific practice tests and learning materials. LanguageCert does this “by design”, with all aspects of each qualification being fully integrated and aligned.

The LanguageCert System of examinations has been developed using a range of language models addressing different language sub-skills and competencies. This includes the models from authors such as Bachman (1990), Canale and Swain (1980), and Weir (2005), as well as the model proposed by the CEFR (2018, 2019) which is the recognised international standard. These sources are used to ensure that LanguageCert’s tests are valid, reliable, and authentic for the targeted domains.

Washback by Design

Washback by design refers to the intentional and systematic incorporation of the potentially positive impact of an assessment on teaching and learning into the test development process. Green (2007) has examined the effects of high-stakes qualifications such as IELTS on teaching and learning, exploring the effect of assessment and evaluation criteria on development of test-taking strategies and development of critical thinking and analytical skills alongside communicative language competence. Cheng and Sultana (2021) provide a comprehensive review of washback research in language testing and the potential for assessment to promote positive washback in teaching and learning. They highlight a need for continuing research and assessment policies that promote positive washback and support teaching and learning.

Designing assessments that promote positive washback and measuring their intended impact is complex and challenging and yet, emphatically, non-negotiable. To deliver an assessment without attempting to understand or measure its intended (and unintended) consequences and its impact on the lives and life chances of candidates would be morally and ethically questionable.

The area of washback by design is one in which LanguageCert is poised to make a contribution, adding to the corpus of work already undertaken by Cheng, Green and others in the field.

Washback by design is explicit in LanguageCert assessment services and processes and is a fundamental consideration in developing tests and preparatory learning materials. LanguageCert supply learning and preparation materials to encourage candidates and their tutors not to prepare for the tests blind to the language skills necessary to succeed, and unclear on how they will be tested. "By design" means the recognition and response to the need for positive washback in all processes for developing tests and their related learning materials. This approach ensures alignment between what language learners experience as they prepare for LanguageCert tests, and what they experience in the examinations. It also ensures that the skills learners practice for the tests have real-world validity and maximise learners' quality of life in terms of personal, occupational, social and economic wellbeing.

English-language learning for migration to an English-speaking context leads to a wealth of individual benefits and societal advantages for test takers. These include increased access to opportunities and resources in terms of education, employment and personal growth and social services including education, healthcare and housing; improved ability to understand laws, customs and practices that influence interaction, communication and behaviour in the host context, smoothing integration into daily life and community activities; enhanced communication with a range of acquaintances, employers, colleagues and service providers. More effective communication skills lead to increased social interaction and foster a sense of belonging as well as mutual cultural exchange, understanding and appreciation.

An overarching intention is to contribute to understanding how assessment might be used to improve long-term outcomes. An underpinning principle to LanguageCert's approach to test development is that if the test is not fit for purpose, it is understandable that teaching (or learning) to the test can constitute negative washback and a focus on skills or knowledge – nothing more than hurdles to clear in an examination scenario – which will not enable personal, occupational, social or economic success and wellbeing. If the test is designed consultatively to meet the specific needs of stakeholders – including migrants, employers, authorities and policy makers – then LCG may be viewed as a test which accurately encapsulates curriculum objectives and as such reflects practical language use and therefore exerts positive impact. By promoting the honing and development of relevant skills in the realm of teaching and learning, assessment can be seen as the portal to opportunities to use the same skills in the real world as enablers of success, progression and transformation.

Designing Tests that Measure Language Competence Across the Four Skills

This section outlines how language competence is measured across the four skills test.

Developing Domain Relevance in the Listening Tests

The LCG Listening test consists of 30 items across four parts. The range of content types are appropriate for the targeted domain in terms of relevant task types underpinned by robust statistical measurement that allow candidates to focus on content rather than familiarity with too many different activity requirements.

In one part of the Listening Test, candidates hear a range of dialogues in a range of situations and contexts in which migrants might find themselves. An awareness of the appropriacy of language depending upon who the interaction is with - a formal interview, a boss at work, a co-worker, or a neighbour - enables successful communication, helps achieve desired outcomes and derive value from social networks, relationships and interactions within a community or society.

In another listening task, candidates take notes while listening to a monologue. The ability to listen to an extended monologue and take notes is an essential skill for a migrant in many scenarios in everyday life. Listening and taking accurate notes is important in social, educational and occupational scenarios. The ability to note factual details when presented with information is important, for example, when dealing with the administrative requirements associated with migration, e.g., noting required documentation or addresses of offices. In a personal context, migrants might wish to take notes to record medical arrangements or details provided by friends and acquaintances in a more social setting. Additionally, in an educational setting, the ability to take accurate notes is a prerequisite for supporting educational development not only in English, but across a complete range of professional, vocational, and academic study.

Candidates also have the opportunity to hear an extended conversation. Understanding and following an extended dialogue is an essential communication skill in a range of settings for a migrant, including occupational, social, and educational. There are multiple challenges when listening to extended conversations and discussions between colleagues, friends, acquaintances, or when listening to the news, or enjoying forms of entertainment for relaxation. Speed of delivery, lack of visual cues, extended range of vocabulary and topic knowledge are all in play. The skills a migrant needs to overcome potential barriers to communication include listening for key vocabulary and linguistic signposts to ascertain gist, what the speaker is talking about and why. Candidates also need to be able to focus on the details of extended discussion to identify opinion, purpose, agreement, disagreement, feelings, and emotions in range of social, cultural and economic contexts.

Range of Accents

Each Listening Test uses a range of accents across the various parts of the examination, to ensure a candidate does not experience just one type of accent during their test.

The Listening Test includes a range of accents drawn from the UK national and regional and other English-speaking communities, including North American, Australian, Irish and South African.

The balance and proportion of accent representation also relates to the lengths of time different accents are heard during the tests.

The balance of accents also reflects the current markets for LanguageCert's test products. LanguageCert responds to target geographies where the candidates study or migrate to. It also recognises where institutions reside. As the market is dynamic, this balance is continually reviewed and integrated within the test development and maintenance programme.

There are checks and balances in LanguageCert's documented test creation procedures to ensure that an appropriate balance is achieved across test forms, and this is kept under review. As a global examination board, working with international teams of test developers and writers, LanguageCert avoids a UK-centric bias (in terms of accent, topics, vocabulary, cultural context and socio-economic or educational bias), which could lead to advantages or disadvantages for certain groups of candidates.

Developing Domain Relevance in the Reading Tests

Reading skills enable migrants to access a range of information including services, employment opportunities and learning resources. For migrants with families, reading skills are important both for their children's education and to help with homework. The Reading Test consists of 30 items across four parts. The Reading Test includes a range of content types, including multiple-choice questions, gap filling and multiple matching. The tasks include a range of source texts of different lengths relevant to the domains of the test. These include newspapers, websites and public notices. Two of the IESOL SELT content types are unchanged and two new content types have been included to target level and domain more effectively.

LCG includes a new (in relation to IESOL B2) Reading Part 1, divided into Part 1a and Part 1b, both of which are vocabulary tasks. Part 1a is a multiple-choice task in which candidates read six sentences and replace a highlighted word in each sentence with a synonym without changing the meaning. There are four options to replace each word. Part 1b is a multiple-choice cloze task in which candidates select the correct word or phrase to fill gaps in a short text. The focus of the new Part 1 tasks is on lexico-grammatical awareness of vocabulary and structures. In everyday life, migrants are likely to encounter unfamiliar vocabulary. The ability to deploy reading strategies to work out the meaning of unfamiliar words, for example by using the surrounding language, is essential to support understanding in reading. Migrants will need to understand unfamiliar words so they can interpret the overall meaning of a sentence, and thereby understand wider meaning of whole texts including books, magazines, work-related documentation, legal documentation, forms, newspapers, letters, and emails. Without understanding how to use these strategies, reading is disjointed, frustrating and unpleasurable (impacting negatively on confidence and reading for pleasure). In addition, communication is impaired.

The Reading Test includes a range of different genres. Understanding how texts are structured is an important skill for a migrant who needs to follow longer texts for a range of occupational, educational, and personal purposes, e.g., reports, instructions, articles, and training documents. Different types of writing are structured according to specific conventions using specific cohesive devices and the ability to identify and use these markers supports fluent and active reading that will also support the development of writing skills.

The Reading Test also includes opportunities to process information from a range of sources. Migrants will need to assimilate information from a range of texts on a related theme or for an overall purpose; e.g., a range of product reviews on a website, comments on a topical matter or workplace issue, or other written material to support understanding of an issue, instruction, or question. It is important and helpful for migrants to be able to identify meaning, opinions, facts, and attitudes in a range of texts and to be able to compare and contrast these in their reading.

Finally, in occupational, educational, and personal domains, migrants are required to read longer texts. The ability to read and understand longer texts is a foundation skill which empowers people to grow and succeed as employees, students, individuals in society and as prospective citizens. Understanding the key features and content of a range of longer texts will enable migrants to develop and consolidate new skills, learn, and grow their imagination, as well as improve their other English language skills.

Developing Domain Relevance in the Writing Tests

Writing skills allow migrants to express themselves, self-advocate and access opportunities. LCG contains two writing tasks set in contexts that are appropriate for the nature of the candidature and the desired outcomes of the test. Tasks revolve around neutral/formal and informal communicative writing for a specific purpose and intended audience. In the first task, candidates produce a short letter email or report in response to a short input text covering three required pieces of information. In the second task candidates compose an informal email, a narrative or descriptive text, or an article which addresses an experience, ideas on a topic, future plans or explaining feelings.

Developing Domain Relevance in the Speaking Tests

Migrants need to be able to engage in interactions giving personal information, opinions and describing feelings. Effective verbal communication skills enable migrants to actively engage in life in a new country. Speaking skills enable integration in terms of employment, education and social interaction. The LCG Speaking Test includes opportunities for candidates to engage in interaction about themselves and their opinions, exchange views and state advantages and disadvantages. Candidates also initiate and respond in role plays designed to replicate a range of workplace scenarios or situations in everyday life. In another part of the test, candidates read a short text aloud and answer some follow-up questions. Such a task is intended to replicate reading aloud in a workplace, study or social context. Follow-up questions require the ability to report, paraphrase and recommend. Finally, candidates prepare and deliver a short presentation on a given topic in an occupational or personal context. Candidates have an opportunity to express and justify their thoughts, view and opinions in the presentation.

Developing Domain Relevance in the Marking Criteria

LCG and LCA both use an analytical mark scheme for all tasks in the Speaking Test and individual task-based mark schemes for the two tasks in the Writing Test. In the Writing test, the two examiners use the same markschemes and the same analytical criteria. In Speaking, the interlocutor awards marks for 'Task Fulfilment and Communicative Effect' which is, in effect, a holistic 'global achievement' scale while the second examiner who listens to the recording retrospectively awards marks against analytical criteria. The application of the marking criteria to each respective domain reflects the nature of the domain-specific tasks in the examinations and outlined in this paper. For example, under task fulfilment in an LCG writing task, examiners are looking for an appropriate genre and tone when candidates respond to a task requiring an email in a work context. This differs from the LCA test, where the writing tasks require the ability to present relevant information, develop arguments, as well as expand upon and support key points, using a different style and tone. This approach flows across to the organisation, grammar, and vocabulary criteria, where a marking and rating is based on the ability to create and sustain a logical flow, to convey meaning effectively, and use correct punctuation. This difference in focus is operationalised through the training of examiners using sample candidate scripts which illustrate the features referred to above, and in the mark schemes.

Reliability and Scoring

LCG is a four-skill test that reports performance across multiple levels (the IESOL SELT tests are single-level). This extension in the reporting capability is in response to demands (from both candidates and recognising institutions) for a practical and effective multi-level test. LCG is focused on the B1 and B2 levels but also measures at A2 and C1. Compared to the original IESOL B2 test, LCG has an increased number of items (from 26 to 30) to facilitate a greater spread of difficulty and improve the ability to report with confidence across a range of CEFR skill levels.

Results are reported against the CEFR levels and on the LanguageCert Global Scale (Milanovic et al, 2023). The Global Scale score (which is provided by language skill and overall result) gives finer gradations of performance within the CEFR levels but is also a standalone measure that can be aligned with any relevant external scale.

The Global Scale for reporting results has been established through the pretesting and live calibration of test materials by LanguageCert, and through the mapping of the Academic and General tests against other examinations in the same domains (for example IELTS) via the CEFR. The accuracy of these measures is determined and verified by a concordance study which is currently in progress. The study examines the extent of overlap in content and performance between LCA and LCG and IELTS Academic and General Training tests.

The LCG test is a multi-level assessment, as mentioned, and measures across levels. LanguageCert research (Lee et al., 2023) has shown that, while the IESOL SELT level-based tests assess at their target CEFR levels, they contain an appropriate number of items to allow assessment across levels. The IESOL SELT B2 examination, for example, has items which assess above and below B2. The ability to measure and report candidate ability across a range of levels is useful for candidates and stakeholders who make decisions informed by candidate results.

Lee et al.'s (2023) study explored the alignment of LanguageCert IESOL SELT tests in relation to the two objectively marked components of Listening and Reading. The use of externally referenced anchoring demonstrated the robustness of the CEFR test levels. For example, in the case of LanguageCert IESOL SELT B2 test, most accurate measurement was observed across two CEFR levels (B1 and B2) and reasonable measurement was observed at the lower end of C1 (see Lee et al., 2023).

The value and utility of a test that measures across multiple levels on a common scale are heightened in LCG (and LCA). Both tests' multi-level assessment capability has been enhanced by increasing the number of items in each test form. This has been done in the knowledge that the original IESOL tests supported accurate measurement across the two levels that each targeted, and reasonable measurement across four levels. By increasing the number of items in each of the General and Academic tests, accuracy has increased across levels. This enhancement also included refining the content types in the Reading test – in particular the replacement of the True/False task. This refinement ensures that the full range of levels is tested effectively, and that all items discriminate well.

New materials target specific levels as defined in the Item Writer Guidelines (IWGs). The materials are created by experienced LanguageCert writers and reviewers. Used in combination with calibrated anchor items, LanguageCert is confident that both tests assess across the stated ability range effectively. This is reinforced through ongoing internal and external validation research to locate all LanguageCert assessment products on its underpinning measurement scale, and aligning all LanguageCert products to the CEFR through which equivalence with other qualifications can be drawn.

LanguageCert estimates the standard error of measurement (SEM) for all tests and reports this both overall and for the individual Listening, Reading, Speaking and Writing skill tests.

Measurement Scale

The Global Scale is used to measure each candidate's performance – see Milanovic et al. (2023). The Global Scale reports scores on a 0 to 100 scale. Candidates receive a score for each skill on the Global Scale, as well as a CEFR level based on the alignment of their total score with the Global Scale. The Global Scale corresponds directly to LanguageCert's internal LID (LanguageCert Difficulty) scale.

The LID scale has been in use since 2016. It is a scale of difficulty used for internal item banking and test construction purposes. The LID scale was developed using a combination of expert judgement and statistical analyses. Up to ten expert consultants, each of whom had over 20 years' experience writing, editing and vetting test materials to measure directly against the CEFR, completed a standards-setting exercise which generated anchor material to enhance and validate the scale. These anchor items then underwent trials and live tests, with all other items measured against them, thereby giving each a difficulty value on the LID scale (see Lee et al, 2023).

An in-depth analysis was conducted on all anchor items and adjustments made where necessary. Rasch and Classical Statistics analyses were then conducted on all live and trial tests, leading to the majority of test items in the bank now considered as being fully calibrated. Research and validation studies in this area are provided in Coniam et al. (2021a) and Coniam et al. (2021b).

The Global Scale links to the LID scale and thereby the CEFR levels. In turn, this means that performance on LanguageCert tests may be seen to be directly comparable to examinations provided by other English language testing organisations, such as IELTS, Cambridge Advanced and the China Standards of English scale. Figure 3 illustrates how the Global Scale reports against the CEFR levels. These findings are under ongoing review in the LanguageCert concordance study which is currently underway.

Figure 3: The LanguageCert Global Scale

LanguageCert Global Scale	CEFR	LanguageCert General	LanguageCert Academic	LanguageCert Global Scale
100	C2			100
99				99
98				98
97				97
96				96
95				95
94				94
93	93			
92	92			
91	91			
90	90			
89	C1		90	89
88				88
87				87
86				86
85				85
84				84
83				83
82	82			
81	81			
80	80			
79	79			
78	78			
77	77			
76	76			
75	75			
74	B2	75	75	74
73				73
72				72
71				71
70				70
69				69
68				68
67	67			
66	66			
65	65			
64	64			
63	63			
62	62			
61	61			
60	60			
59	B1	60	60	59
58				58
57				57
56				56
55				55
54				54
53				53
52	52			
51	51			
50	50			
49	49			
48	48			
47	47			
46	46			
45	45			
44	44			
43	43			
42	42			
41	41			
40	40			
39	A2			39
38				38
37				37
36				36
35				35
34				34
33				33
32	32			
31	31			
30	30			
29	29			
28	28			
27	27			
26	26			
25	25			
24	24			
23	23			
22	22			
21	21			
20	20			
19	A1			19
18				18
17				17
16				16
15				15
14				14
13				13
12	12			
11	11			
10	10			
9	Pre A1			9
8				8
7				7
6				6
5				5
4				4
3				3
2	2			
1	1			
0	0			

The three-page candidate report (Appendix 1) reveals how the LanguageCert Global Scale is operationalised.

The Global Scale allows ease of interpretation for test users and provides a finely-tuned results service across all language skills. As shown, performance can be separated both by each skill and overall, so that a candidate is not only described as having “B2 ability”, but a more precise level of detail is provided on a candidate’s performance. The Test Report shows an overall score, the overall CEFR level of attainment reached, and the score for each of the skills using both the Global scale and the CEFR level of attainment.

The Global Scale, launched with the LanguageCert Test of English (LTE), measures from pre-A1 to high C2. The LTE has been successfully administered to tens of thousands of candidates worldwide, and the Global Scale has received good customer feedback in terms of its simplicity, clarity, and ease of use.

Items in the LCG Reading and Listening Tests range in difficulty from CEFR level A2 to C1, with the majority of items focusing on the B1 and B2 levels (Intermediate and Upper Intermediate). Item difficulty is established through pre-testing and live test calibration using Rasch and Classical Statistics. All Reading and Listening items are calibrated to the LanguageCert Global Scale which runs from CEFR Pre-A1 to C2 levels. Examples of the ways in which items are calibrated using Rasch and Classical Statistics are described in Falvey and Coniam (2023) and reveal that this method of calibration is demonstrably robust.

Each LCG Reading and Listening Test is designed to cover a wide range of the B1/B2 CEFR 'syllabus' (i.e., those areas covered by the Can-Do statements in the CEFR). A broad range of Reading and Listening sub-skills are tested, as is a range of grammar, vocabulary, and awareness of functional language. Tasks are set in contexts that are appropriate for the nature of the candidature and the desired outcomes of the test.

For the LCG Writing and Speaking Tests, detailed mark schemes are used by examiners. In Writing, candidates complete two writing tasks. Task 1 requires candidates to produce a short letter, email or report of approximately 100 to 150 words covering three required pieces of information in response to a short input text. In Task 2, candidates need to produce a slightly longer piece of informal writing – either an informal email, a narrative or descriptive text or article of 150 to 200 words which addresses an experience, ideas on a topic, future plans or explaining feelings.

In the marking of Writing, candidates are assessed against four criteria. These are:

1. Task Fulfilment
2. Accuracy and Range of Grammar
3. Accuracy and Range of Vocabulary
4. Organisation and Coherence

The use of separate criteria to measure different aspects of Writing performance allows the LCG test to deliver rich feedback to both candidates and receiving organisations and provides indications as to where further development is needed by the candidate. The marking criteria have been adapted from the LanguageCert IESOL B2 examination Writing marking criteria. At the outset, the criteria were based on the descriptors for Writing in the CEFR in conjunction with the nature of the task. These original criteria have been developed over many years, with active consideration of their relevance and applicability. Feedback has been collected from trainers, examiners, and examiner-monitors (senior examiners) to fine-tune the wording of the criteria so that examiners find them easy to use, so that they reflect candidate output, and so that the key features expected from candidates in the examination at each CEFR level are considered.

The criteria have been extended to measure performance across a broader range of ability (from A2 to C1) to report reliably across an extended range of CEFR levels.

Writing scripts are marked by two human examiners. If there is a significant difference in the marks awarded, the script is passed to a third (more senior) examiner whose decision is final. It is intended, that in the medium to longer term, auto-marking by computer will be introduced as part of a hybrid scoring approach.

For Speaking, the test is split into four parts. Part 1 involves responding to transactional questions across a range of topics. In Part 2, candidates take part in role-plays which are set in a range of real-life scenarios relevant to a migrant living, working or studying in an English-speaking context. In Part 3, candidates read aloud a short passage of around 80 words in length on a topical issue and answer follow-up questions selected from a list by the interlocutor. In Part 4, candidates talk about a topic selected by the interlocutor for up to two minutes. The candidate has preparation time, and after giving their talk they then answer follow-up questions selected from a list held by the interlocutor.

In the marking of Speaking, candidates are assessed against five criteria. These are:

1. Task Fulfilment and Communicative Effect
2. Coherence
3. Accuracy and Range of Grammar
4. Accuracy and Range of Vocabulary
5. Pronunciation, Intonation and Fluency

Just as for Writing, the use of separate criteria to measure different aspects of Speaking performance allows the LanguageCert General test to deliver rich feedback to both candidates and receiving organisations and provides indications as to where further development is required on the part of the candidate.

The criteria have been adapted from the IESOL B2 Speaking Test marking criteria. At the outset, the criteria were based on the descriptors for Speaking in the CEFR, in conjunction with the nature of the tasks. These original criteria have been developed over many years, with active consideration of their relevance and applicability. Feedback has been taken from trainers, examiners, and examiner-monitors (senior examiners) to fine-tune the wording of the criteria so that examiners find them easy to use, so that they reflect candidate output, and so that the key features expected from candidates at each CEFR level are considered.

The criteria have been extended to measure performance across a broader range of ability (from A2 to C1).

Currently, candidate output in the Speaking Test is marked by two human examiners; by the interlocutor immediately after the test and by a second examiner who awards marks subsequently by accessing the video recording. The first criterion *Task Fulfilment and Communicative Effect* is marked by the interlocutor and provides a general impression score that contains elements of the more analytical criteria used by the second examiner. The second examiner marks the other analytical criteria. The interlocutor general impression mark is then double-weighted. If there is a significant difference in marks awarded by the two examiners, then the recording goes to a third (more senior) examiner whose marks are final.

In the medium to longer-term, auto-marking by computer is being planned to be introduced as part of a hybrid scoring approach. A hybrid assessment model will garner the proven benefits of both human and machine marking (see e.g., Babitha et al., 2022).

Test Development Process and Quality Assurance

LanguageCert's Assessment Development department contains academics as well as professional linguists and assessors, who publish research on all aspects of the language qualifications. An Advisory Council supports this team and helps it to meet regulatory obligations to bodies such as Ofqual.

All tests and test items are constructed and assured by high-calibre test developers operating to clear guidelines, workflows, and quality assurance protocols which include layers of reviews, editing, statistical analyses, and vetting. The LanguageCert proprietary item bank is used to manage all LanguageCert tests, with strict access protocols, and robust workflows for process compliance. LanguageCert's team of examiners includes expert Chief Examiners as well as Examiners and their Team Leaders. All undergo stringent training before marking live papers. A defined marking process operates within the PeopleCert proprietary marking application, which standardises, and quality assures the process and its outputs. All candidate digital, audio and video interactions during tests are recorded and securely stored so that there is a verifiable evidence base for all results. In addition, robust quality assurance protocols are applied to secure integrity and fairness for the test and the candidate.

To explore whether any subgroup of candidates sitting a test is being unfairly disadvantaged, LanguageCert addresses the challenge at a number of levels. The process starts with comprehensive item writer guidelines and item writer training. This is then supplemented by the detailed vetting and editing of test materials with a focus, amongst other things, on whether there is a risk of candidates of specific backgrounds being disadvantaged. In addition, differential item functioning (DIF) analyses – the key to investigating and dealing with test bias – are conducted. Coniam and Lee (2021) describe DIF analysis conducted on IESOL examinations delivered from 2018 to 2021 (with some of the populations involving IESOL examinations delivered for the UKVI scheme). With gender, typically a key variable in the exploration of DIF, there was a very low incidence of DIF. An examination of Reading or Listening items indicated that there was no significant DIF in either skill. With the findings confirming that the LanguageCert tests analysed exhibit low levels of gender bias, a methodology is in place for the ongoing monitoring of DIF on all LanguageCert examinations.

As an international organisation, LanguageCert strives to ensure its tests are valid, reliable and have a positive impact on learners. An important part of ensuring fairness to candidates is to minimise any bias in the test materials. The process of eliminating bias begins with the formation of the test specifications. These are written with direct reference to the nature of the intended or anticipated candidature to ensure the tests are fully fit-for-purpose. This detail is checked at annual reviews and when the test formats are revised. LanguageCert makes sure writers understand who the target domain test users are, and that they consider aspects such as the level of cognitive processing of typical candidates, and their cultural contexts.

Both LanguageCert's Item Writer Guidelines and the training process stress bias awareness, and the requirement to produce materials which will not favour or discriminate against certain candidates. This entails ensuring test materials are as free from specific regional or national cultures as possible, and that topics are universal. Writers have a list of taboo topics to aid in this. These taboo topics include areas which may cause distress or distraction to candidates or relate to unfortunate experiences they may have suffered (e.g., war or drugs), through to specific aspects of local cultures (e.g., milkmen in Britain) which may be alien to the local culture of the candidate or beyond their life experience. The LanguageCert team also take care to avoid introducing test material which may test general knowledge or specific technical knowledge, rather than language ability.

Ongoing Development, Monitoring and Evaluation

Ongoing stakeholder engagement is crucial in the continuous development of LCG. Feedback is provided by way of regular webinars, presented by development staff to stakeholders such as institutional administrators, admissions tutors and other key personnel involved in the admission, tutoring and mentoring of successful candidates coming to the UK for education purposes. LanguageCert disseminate findings of their research and invite comment and participation via a quarterly update from the assessment research and validation team, *Research Insights*. This publication also has a role in communicating and inviting dialogue with our stakeholders and LanguageCert General and Language Cert Academic research will become a regular feature in this publication as the qualification roll-out is widened.

Conclusion

This paper has described how an examination evolves to ensure the target language use domain is covered and provides a valid, fair, inclusive and reliable assessment tool.

The paper has provided the rationale for the evolution of the LCG test, its purpose and the needs it meets, the curricular factors in play, the development of the examination, and its pretesting, piloting and eventual offering to the public. LCG is closely based on the LanguageCert IESOL B2. Its development, and the guiding body of research, has informed the ongoing review and evolution of that examination.

It has been outlined how the LanguageCert General test focuses on general language requirements for use in the migrant employment target language domain. The test has been developed by LanguageCert personnel and pre-tested and piloted internationally – at LanguageCert-approved test centres under secure test-taking conditions, with pretesting populations which are representative of each test's intended candidature. All these factors underscore the care taken to employ the best research findings, methodology, and statistical procedures in order to develop and improve the quality the test.

References

- ALTE (Association of Language Testers for Europe). (2016). *Language tests for access, integration and citizenship: an outline for policy makers*. Strasbourg: Council of Europe.
- ALTE (Association of Language Testers for Europe). (2020). *ALTE Linguistic integration of adult migrants: requirements and learning opportunities*. Strasbourg: Council of Europe.
- Babitha, M. M., Sushma, C., & Gudivada, V. K. (2022). Trends of artificial intelligence for online exams in education. *International Journal of Early Childhood Special Education*, 14(01), 2457-2463.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A.S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Brindley, G., & Burrows, C. (2000). *Studies in immigrant English language assessment*. Sydney: NCELTR, Macquarie University.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Cheng, L., & Sultana, N. (2021). Washback: Looking backward and forward. In *The Routledge Handbook of Language Testing* (pp. 136-152). Routledge.
- Coniam, D., & Lee, T. (2021). Potential bias in LanguageCert IESOL items: A Differential Item Functioning analysis. London, UK: LanguageCert.
- Coniam, D., Lee, T., Milanovic, M. & Pike, N. (2021a). Validating the LanguageCert Test of English scale: The adaptive test. London, UK: LanguageCert.
- Coniam, D., Lee, T., Milanovic, M. & Pike, N. (2021b). Validating the LanguageCert Test of English scale: The paper-based tests. London, UK: LanguageCert.
- Council of Europe (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. Strasbourg: Council of Europe.
- Falvey, P., & Coniam, D. (eds.) (2023). *Certifying quality in assessment & learning: Research and validation at LanguageCert. Volume 2*. London, UK: LanguageCert.
- Jones, C. (2023). *Orienting the LanguageCert IESOL C1 examination to an academic context*. London, UK: LanguageCert.
- Knoch, U. (2021). *A guide to English language policy making in higher education*. International Education Association of Australia. (IEAA). Retrieved from www.ieaa.org.au.
- Lee, T., Papargyris, Y., Milanovic, M., Pike, N., & Coniam, D. (2023). *Aligning LanguageCert SELT tests to the LanguageCert Item Difficulty (LID) scale*. London, UK: LanguageCert.
- Milanovic, M., Pike, N., Papargyris, Y., Lee, T., & Coniam, D. (2023). *The LanguageCert Global Scale*. London, UK: LanguageCert.
- Saville, N. (2009). Language assessment in the management of international migration: A framework for considering the issues. *Language Assessment Quarterly*, 6(1), 17-29.
- van Ek, J. A., & Trim, J. L. M. (1990a/1998a). *Threshold 1990*. Cambridge: Cambridge University Press.
- van Ek, J. A., & Trim, J. L. M. (1990b/1998b). *Waystage 1990*. Cambridge: Cambridge University Press.
- van Ek, J. A., & Trim, J. L. M. (2001). *Vantage*. Cambridge: Cambridge University Press.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Houndmills, UK: Palgrave Macmillan.


Appendix 1: Candidate Report

Language
Cert

LanguageCert General (Listening, Reading, Writing, Speaking)

Test Report

Candidate Information

Last Name:	Candidate's Last Name	 TEST DAY PHOTO	
First Name:	Candidate's First Name		
Date of birth:	xx Month xxxx		
Candidate Number:	99800...		
Candidate URN:	PPC/...		
ID Type:	xxxxx		
ID Number:	xxxxx	Nationality:	xxxxx

Test Centre Information

Date of Test:	xx Month xxxx	Date Test Results issued:	xx Month xxxx
Test Centre number:	xxxxx	Test Centre country:	xxxxx
Mode of Delivery:	Test Centre		

Candidate Results (out of 100 on the LanguageCert Global Scale)

Listening	73	Writing	66
Reading	68	Speaking	71
Total Score	69		
CEFR Level	B2		

Signature LanguageCert
Chairman

Name of LanguageCert
Chairman

THIS IS NOT A CERTIFICATE

LanguageCert is a business name of PeopleCert Qualifications Ltd, UK company number 09620926
LanguageCert reserves the right to amend the information given before issuing certificates to successful candidates.
info@languagecert.org

Candidate Performance Feedback (Writing Part 1)

Task Fulfilment	You addressed all three points of the task and expanded some of your points further. You were able to communicate clearly. Your genre and tone were mostly appropriate. The reader was, on the whole, informed.
Accuracy and Range of Grammar	You used a range of simple grammatical forms accurately and you were able to attempt and have some success with a few complex forms. Some errors occurred which, at times, prevented understanding. Re-reading was sometimes needed.
Accuracy and Range of Vocabulary	Your use of everyday vocabulary was accurate and of an adequate range for the task. There was some misuse of less common words and phrases. A few spelling errors occurred. Any errors did not prevent comprehension, but some re-reading was needed.
Organisation and Coherence	Your text was often well-structured and clear. You made use of a variety of linking words and phrases, usually successfully. You used paragraphs mostly appropriately and your organisation was appropriate to the text type. There were some punctuation errors, but these did not prevent comprehension.

Candidate Performance Feedback (Writing Part 2)

Task Fulfilment	You satisfied the demands of the task to a large extent and fully addressed both points. You could have expanded a few of your points further, but your communication was largely successful. Your genre and tone were highly appropriate. The reader was highly informed.
Accuracy and Range of Grammar	You used a range of simple grammatical forms accurately and you were able to attempt and have some success with more complex forms. A few errors occurred but these did not prevent comprehension. Some re-reading may have been needed.
Accuracy and Range of Vocabulary	You used a range of vocabulary, including less common words and phrases, effectively. A few errors in usage, spelling and word formation occurred but did not prevent comprehension. Minor re-reading may have been needed. There were very few spelling errors.
Organisation and Coherence	Your text was often well-structured and clear. You made use of a variety of linking words and phrases, usually successfully. You used paragraphs mostly appropriately and your organisation was appropriate to the text type. There were some punctuation errors, but these did not prevent understanding.

Candidate Performance Feedback (Speaking)

Task Fulfilment and Communicative Effect	You completed all tasks with ease and, at times, confidence. You communicated what you wanted to say mostly clearly and almost always in a natural manner. Misunderstandings were very rare. Your contributions were always relevant and often fully detailed.
Coherence	Your use of language was clear and often well-structured. You were able to speak at length; some hesitations may have caused coherence to breakdown. You used linking words and phrases almost always effectively.
Accuracy and Range of Grammar	You used a range of grammatical structures. Some errors occurred but were repaired or did not impact in a major way on meaning. You had very good control of basic structures. You attempted complex structures, but errors occurred, and restarts were necessary. Your meaning was almost always clear despite errors.
Accuracy and Range of Vocabulary	You used a very good range of vocabulary. Your use of vocabulary was clear. You were able to use some less common words and phrases and idiomatic language. Occasional errors in usage occurred but did not prevent understanding.
Pronunciation, Intonation and Fluency	You generally maintained a spontaneous flow of language; any hesitations did not strain the listener. During the read-aloud task, you spoke with a high level of naturalness and fluency. Your pronunciation was clear and easily understood. You used stress and intonation appropriately to convey meaning.

CEFR Level	Scaled Score	Performance Descriptors (Listening, Reading, Speaking, Writing)
C2	90 - 100	<ul style="list-style-type: none"> • Can understand with ease any kind of spoken language, provided there is familiarity with the accent. • Can read with ease virtually all forms of the written language, including abstract or linguistically complex texts. • Can produce clear, smoothly flowing well-structured speech with an effective logical structure which helps the recipient to notice significant points. • Can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points.
C1	75 - 89	<ul style="list-style-type: none"> • Can understand an extended speech even when it is not clearly structured and when relationships are only implied. • Can read and understand long and complex texts, appreciating distinctions of style. • Can give clear, detailed presentations on complex subjects, integrating sub themes, developing points and rounding off with an appropriate conclusion. • Can write clear, well-structured texts on complex subjects, underlining relevant issues, expanding and supporting points of view with subsidiary points, reasons and examples, and rounding off with an appropriate conclusion.
B2	60 - 74	<ul style="list-style-type: none"> • Can understand extended speech and lectures and follow complex lines of argument provided the topic is reasonably familiar. • Can read and understand articles and reports in which the writers adopt particular attitudes or viewpoints. • Can give clear, detailed presentations on a range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples. • 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.
B1	40 - 59	<ul style="list-style-type: none"> • Can understand the main points of clear standard speech on familiar matters. • Can read and understand texts that mainly consist of high frequency everyday language. • Can reasonably fluently give a straightforward description of subjects within his/her field of interest, presenting it as a linear sequence of points. • Can write a text on a subject of personal interest, using simple language to list advantages and disadvantages and give his/her opinion.
A2	20 - 39	<ul style="list-style-type: none"> • Can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance. • Can read and understand very short, simple texts such as personal letters. • Can give a simple description of people, daily routines, likes/dislikes etc. as a short series of simple phrases and sentences linked into a list. • Can write a series of simple phrases and sentences linked with simple connectors like 'and,' 'but' and 'because'.
A1	10 - 19	<ul style="list-style-type: none"> • Can recognise very familiar words and phrases when people speak slowly. • Can read and understand very simple sentences on familiar topics. • Can produce simple mainly isolated phrases about people and places. • Can write simple isolated phrases and sentences.

The above descriptors are adapted from the Common European Framework of Reference for Languages (2018). Text from these is reproduced by kind permission of the Council of Europe.

LanguageCert Global scale	CEFR	LanguageCert General	LanguageCert Academic	LanguageCert Global scale
100				100
99				99
98				98
97				97
96				96
95				95
94				94
93				93
92				92
91				91
90			90	90
89				89
88				88
87				87
86				86
85				85
84				84
83				83
82				82
81				81
80				80
79				79
78				78
77				77
76				76
75		75	75	75
74				74
73				73
72				72
71				71
70				70
69				69
68				68
67				67
66				66
65				65
64				64
63				63
62				62
61				61
60		60	60	60
59				59
58				58
57				57
56				56
55				55
54				54
53				53
52				52
51				51
50				50
49				49
48				48
47				47
46				46
45				45
44				44
43				43
42				42
41				41
40		40	40	40
39				39
38				38
37				37
36				36
35				35
34				34
33				33
32				32
31				31
30				30
29				29
28				28
27				27
26				26
25				25
24				24
23				23
22				22
21				21
20		20		20
19				19
18				18
17				17
16				16
15				15
14				14
13				13
12				12
11				11
10				10
9				9
8				8
7				7
6				6
5				5
4				4
3				3
2				2
1				1
0				0

THIS IS NOT A CERTIFICATE

LanguageCert is a business name of PeopleCert Qualifications Ltd, UK company number 09620926
 LanguageCert reserves the right to amend the information given before issuing certificates to successful candidates.
info@languagecert.org

LanguageCert is a business name of PeopleCert Qualifications Ltd, UK company number 09620926.

Copyright © 2023 LanguageCert

All rights reserved. No part of this publication may be reproduced or transmitted in any form and by any means (electronic, photocopying, recording or otherwise) except as permitted in writing by LanguageCert. Enquiries for permission to reproduce, transmit or use for any purpose this material should be directed to LanguageCert.

DISCLAIMER

This publication is designed to provide helpful information to the reader. Although care has been taken by LanguageCert in the preparation of this publication, no representation or warranty (express or implied) is given by LanguageCert with respect as to the completeness, accuracy, reliability, suitability or availability of the information contained within it and neither shall LanguageCert be responsible or liable for any loss or damage whatsoever (including but not limited to, special, indirect, consequential) arising or resulting from information, instructions or advice contained within this publication.



Language
Cert

[languagecert.org](https://www.languagecert.org)