

Core Skills Test Purpose and Content

EnglishScore

Produced together with the Centre For Research In
English Language Learning And Assessment (CRELLA) at
the University of Bedfordshire.

Table of contents

I. The purpose and use of EnglishScore.....	2
II. Test design.....	4
a. Test development.....	4
b. Writing assessment material for EnglishScore.....	4
III. The content of EnglishScore.....	8
a. Overview.....	8
b. Information on assessment tasks.....	10
c. Scoring assessment tasks.....	19
d. Reporting scores on EnglishScore as a whole.....	19
e. Assessment results and analysis.....	22
f. Item review and health.....	25
Contact Information.....	27
About the British Council.....	27
Contact EnglishScore.....	27
Bibliography.....	28

I. The purpose and use of EnglishScore

Who should take EnglishScore?

EnglishScore is an international assessment taken by young adult (16 and 17) and adult (18 and over) learners of English worldwide. Test-takers may come from any language background and any region of the world.

The meaning of EnglishScore results

EnglishScore provides evidence of proficiency in understanding and using English in everyday life and the workplace.

The test is primarily concerned with the *occupational*, *public* and *personal* domains¹ with items that are more *personal* at the lowest levels of difficulty, but that focus more on the *public* and then the *occupational* domains as the difficulty level increases.

It relates to a wide range of contexts of language use² with a focus on common workplace and social contexts.

The impact of using EnglishScore

EnglishScore aims to encourage people around the world to unlock the potential of the English language by certifying their scores, helping them to prove their level to potential employers. For universities, employers and other organisations, EnglishScore provides a cost-effective means of large-scale English language testing that is used to inform professional development initiatives, course placements and recruitment efforts.

Ownership of EnglishScore

EnglishScore is owned and administered by the British Council (www.britishcouncil.org), the United Kingdom's international organisation for cultural relations and educational opportunities.

Use of EnglishScore results

EnglishScore can be used by employers, universities and governments to assess a test-taker's general English proficiency. Results can be used by employers to benchmark the proficiency of their workforce or assess a future employee's language ability, by universities and language schools as a placement or progress measure for their students or by governments and other stakeholders as an index of a learner's general English proficiency. In addition, English language

¹ See Council of Europe (2001, pp.10, 14, 42–100) for information on domains.

² See Council of Europe (2001, pp.30, 101–130) for more information on communicative language competences.

learners themselves can use the test to understand their level in relation to the CEFR, set individual language learning goals and select appropriate courses.

Recognition of EnglishScore results

Today, over 1,000 organisations around the world, representing a diverse set of industries, use and recognise EnglishScore certificates. Employers have used EnglishScore as part of the process of recruitment and screening of potential staff and for upskilling their existing workforces. Universities have used EnglishScore as part of their admissions and placement procedures, and also as an exit credential for graduates entering the workforce.

To learn more about EnglishScore, please visit www.englishscore.com.

Test delivery

EnglishScore Core Skills test is an on-demand test and is administered and proctored through a mobile device. Users download an app (available on iOS and Android), register their details and then take the test on their phone. It is free to access, and results are provided instantly, with the option to purchase a certificate on completion of the test. More details on proctoring and other security features are detailed in the [EnglishScore Security Report](#).

II. Test design

a. Test development

Development of EnglishScore

EnglishScore was developed by the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire, UK (www.beds.ac.uk/crella) and is partly modelled on the British Council's *International Language Assessment*. CRELLA is widely recognised as the UK's leading centre for language assessment research.

Content of EnglishScore

EnglishScore Core Skills is informed by the socio-cognitive model of language use originating in Cyril Weir's *Language Testing and Validation* (2005).

This theoretical framework suggests that language use across spoken and written modalities rests on knowledge of vocabulary and grammar. For the Grammar and Vocabulary section, English Profile (2015) Reference Level Descriptions³ for English are used to guide choice of vocabulary and grammatical structures to be targeted by the test.

The Reading and Listening sections target five operations with which skilled readers and listeners engage: *decoding*, *word recognition*, *parsing*, *meaning construction* and *discourse construction*. Test questions represent a hierarchy of difficulty in terms of the cognitive load imposed by different processes or levels of reading and listening. Questions targeting the lower levels of the CEFR tend to be limited to decoding, word recognition and parsing, while those targeting higher levels involve more attention to meaning and discourse construction.

b. Writing assessment material for EnglishScore

Test writer qualifications

EnglishScore writers are teachers of English with a teaching qualification such as a master's degree or Diploma in English Language Teaching and a minimum of five years' experience as teachers of English. They are also familiar with the CEFR and able to write items to the different CEFR levels. Before being accepted for training, writers complete a qualifying item writing task.

Test writer training

All writers are given an induction programme to the test, where they are introduced to the test specifications and practise writing assessment material. Writers regularly participate in review

³ The English Profile helps teachers and educationalists to understand what the CEFR means for English. It describes what aspects of English are typically learned at each CEFR level.

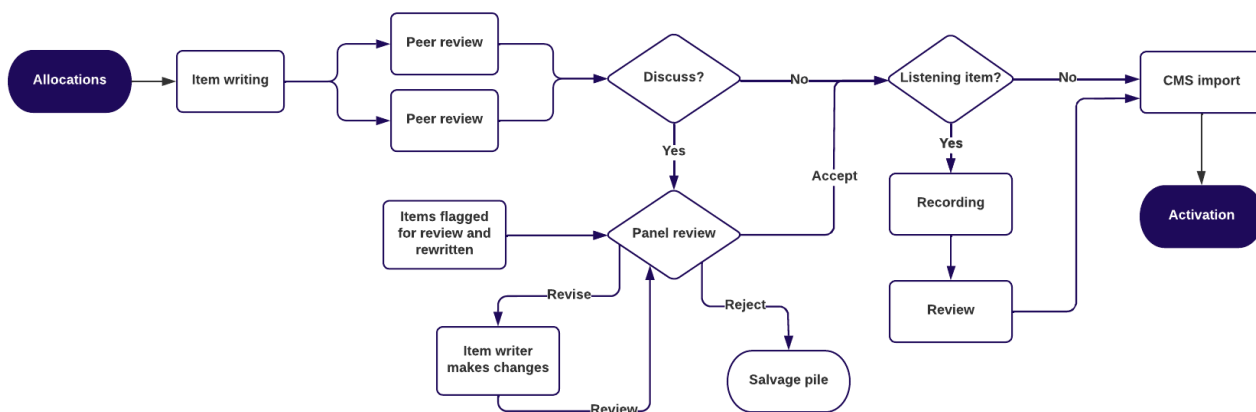
meetings and are required to complete a training course every three years to continue working as contributors to the EnglishScore assessment.

Test writer guidance

To ensure that the content follows the developers' intentions and that it is parallel across different versions of EnglishScore, when preparing material, the writers follow detailed writer guidelines for each section of the test. These include examples of good (and poor) tasks, wordlists, lists of grammatical structures and guidance on features such as text and recording length, rates of speech and complexity. There are self-assessment checklists that writers use to confirm that their work conforms to the guidelines before they submit it. Item writers are also instructed to use recognised databases to ensure vocabulary and grammatical structures are at the appropriate level for the item. Additional training and feedback are provided to item writers by the EnglishScore Head of Assessment as needed, and reviews are included in the item development process.

Test material development process

Test items are developed by a team of trained item writers in a series of item commissions throughout the year. To ensure quality of items and the test as a whole, there is a standardised set of review procedures that all items go through. This helps to ensure that test quality is maintained and that the test results are valid, reliable and fair. An overview of the item development process is provided below:



Notes

- At any stage of the item development process, material may be accepted for the next stage, edited, returned to the writer for revision or rejected.
- Items are reviewed by item writers individually and as a panel and evaluated against the test specifications, assessing aspects such as item content, CEFR level, word count and the answer options and key.
- All review decisions and feedback during the process are securely recorded for reference at a later date if needed.
- Final checks include an editorial review for proofreading and style checks, before being ingested into the CMS and individually reviewed and activated by the EnglishScore Head of Assessment.

Once the items are live, regular checks are conducted to ensure the items are performing as expected. Any items that fall outside the quality parameters are flagged, deactivated and removed from the test (see 'Trialling test materials' below).

Item bank and test security

All the test items are stored in a secure item bank, which includes the item content, answer key, item media and metadata (level, skill, etc.). Items from this database are selected to create a large number of unique test forms, following the template outlined in Section C. The item bank is large enough to ensure there is minimal repetition of items across multiple test attempts by the same user. This helps to maintain the security of the item bank, as well as the validity and reliability of the test.

The item bank and associated CMS are maintained by EnglishScore. Access to the item bank is restricted and controlled through a username and password. All changes to item content are logged with date/time/username, with access permissions regularly reviewed.

Additional details on test security are available in the *EnglishScore Security Report*.

Taking account of user needs

EnglishScore takes account of the diversity of the test-taking population by collecting data about their location and by asking users about their motivation for learning English.

The test material is designed for adult learners of English (aged 16 and over) and aims to avoid any bias associated with gender, nationality or ethnic identity. These issues are addressed in the guidelines for item writers and considered as part of the review process. In addition, items are checked to ensure that they do not include controversial topics, do not require specialist knowledge and that they are culturally neutral, i.e. do not require knowledge of a particular culture or country to be answered correctly. This ensures test fairness for all users around the world. The test interface is designed to be accessible to colourblind users.

Each item has a time limit to ensure fair and consistent test delivery for all test takers. These timings allow sufficient time for test takers to answer the question, and the time is visible to the test taker for each question.

Trialling test materials

Once items have been reviewed and approved, and activated by the Head of Assessment, the new test material is added to the EnglishScore item bank. New items are presented to users alongside items from the operational item bank that have known levels of difficulty. The volume of new items added is monitored carefully so as to have a negligible impact on score results used for proof purposes (certification, employment or other significant purposes).

As part of EnglishScore's quality checks, test sitting data is regularly extracted and mathematically modelled using Rasch analysis. This analysis includes all items in the EnglishScore item bank and provides the assessment team with deep insight into the performance of each item across a wide group of test-takers. Key measures reported on a per item basis include difficulty, mean, infit, outfit and PTMA (point measure correlation for all observations).

Items that fall outside of accepted thresholds on any of the above measures are flagged for further review and re-enter the item review process. Flagged items are either revised or fully rejected.

Feedback on item performance is consistently shared with the item writers responsible, ensuring that new items are less likely to be flagged in the future.

III. The content of EnglishScore

a. Overview

Note that this is an adaptive test. Each user is presented with an individually tailored test with questions suited to their level. Material is selected for the user on later parts of the test based on their performance in earlier parts. No user takes all twelve parts. See Figure 1 below for more detail.

EnglishScore includes three sections: Grammar and Vocabulary, Reading, and Listening. Performance in the Grammar and Vocabulary section determines the questions that users will see in the Reading and Listening sections. Users are placed on one of four possible pathways through the test:

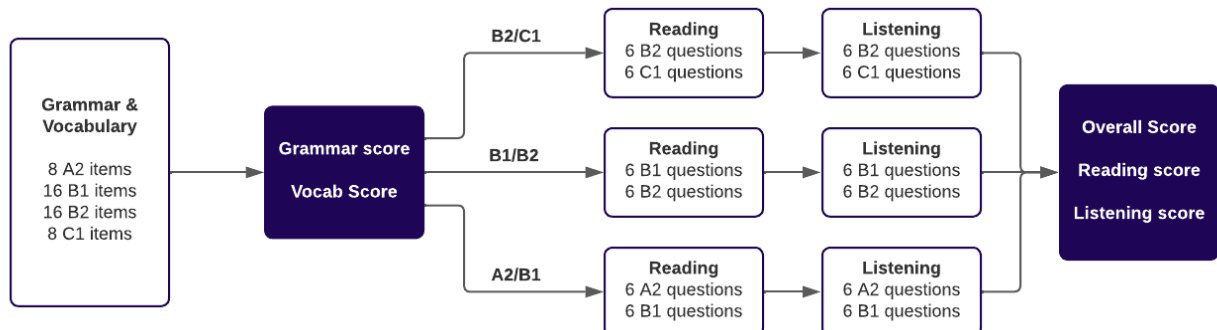
Lowest scorers – no further testing, awarded a score Below A1

Breakthrough+ users – Reading and Listening items targeting A2/B1

Threshold+ users – Reading and Listening items targeting B1/B2

Vantage+ users – Reading and Listening items targeting B2/C1

Figure 1. The EnglishScore Core Skills Test structure.



EnglishScore at a glance

Sections	Part	Timing	No. of questions	Tasks	Input	Response	Points
Grammar and Vocabulary	A2	Each item has a time limit, dependent on the skill, item type and target CEFR level. Details on timings are provided in the following section.	8	Complete a gap in a short text or brief dialogue by choosing one of three options.	Around 15 words	Questions on all sections of EnglishScore involve choosing the correct answer from among three options.	Each response contributes to the estimate of the test-taker's EnglishScore for that section and their overall EnglishScore.
	B1		16	Complete a gap in a short text or brief dialogue by choosing one of three options.	Around 20 words		
	B2		16	Complete a gap in a short text or brief dialogue by choosing one of three options.	Around 25 words		
	C1		8	Complete a gap in a short text or brief dialogue by choosing one of three options.	Around 30 words		
Reading	A2		6	Read a short, incomplete text. Choose one of three options to complete the text.	Around 30 words.		
	B1		6	Read an opening sentence. Order three further sentences to create a coherent text.	Around 70 words		
	B2		6	Read a short text and select a suitable heading from three options.	Around 100 words		
	C1		6	Read a more complex text and answer three questions about the text by selecting the correct answer from among three options for each question.	Around 250 words		
Listening	A2		6	Listen to a monologue or dialogue and identify specific, factual information to answer a question by selecting the correct answer from among three options.	Around 65 words. Approximately 150 words per minute.		
	B1		6	Listen to a monologue or dialogue and identify specific, factual information; or identify main points to answer a question by selecting the correct answer from among three options.	Around 85 words. Approximately 170 words per minute.		
	B2		6	Listen to a monologue or dialogue and identify main points and specific details; infer meanings; or follow discussion to answer a question by selecting the correct answer from among three options.	Around 100 words. Approximately 180 words per minute.		
	C1		6	Listen to a monologue or dialogue and identify main points and specific details; infer meanings; extract relevant information from a complex/abstract speech; follow discussion; or identify the speaker's attitude and line of argument to answer a question by selecting the correct answer from among three options.	Around 120 words. Approximately 190 words per minute.		

b. Information on assessment tasks

EnglishScore Grammar and Vocabulary section

i. Rationale

Overview of the section

At A2 level to C1 level, for both grammar and vocabulary, this section involves choosing a word or phrase to fill a gap in either a short text or a two-turn dialogue.

Connection to language use

The tasks are similar to grammar and vocabulary exercises commonly found in English textbooks around the world.

Instructions

The instructions are given in English. All instructions for the Grammar and Vocabulary section are given in writing.

Timing

The test taker has 45 seconds to answer each question. It is possible for the test taker to answer and move to the next question before the time is complete.

ii. What the user reads (the input)

The input

As the CEFR level targeted by the item increases, the domain of the input changes: at A2 level, the input mainly relates to the *Personal* domain, at B1 level, the *Public* domain and at B2 and C1 levels, the *Occupational* domain. In addition, each item has a short title to establish the context of the item for the test-taker.

Communication themes

Communication themes may include *personal identification/house, home, environment/daily life/free time/entertainment/relations with other people/shopping/food and drink/places and weather* (Council of Europe, 2001, p. 52).

Sources of input

The input is prepared by the item writers specially for EnglishScore. Items are written to reflect the grammatical structures and vocabulary test-takers would encounter in the relevant domains.

Nature of input

At lower levels, the input is mostly very *concrete* and *familiar*, e.g. about *where people live, people they know and things they have, common objects around them*, progressing to less familiar and more abstract topics at the higher levels, e.g. *a wide range of both familiar and unfamiliar topics normally encountered in personal, social or vocational life*.

Each Grammar and Vocabulary item is assigned a 'test bucket'. Each bucket has a CEFR level and focus, e.g. *B1 adjectives and adverbs* or *B2 modality*. This ensures that each test covers a range of vocabulary and grammatical constructions, appropriate to the item's CEFR level.

Extent of input

Level	Length	Grammar	Vocabulary
A2	10–20 words	simple grammatical structures and sentence patterns	sufficient range of vocabulary for routine, everyday transactions involving familiar situations and topics
B1	15–25 words	some complex sentence forms	sufficient range of vocabulary to express most topics pertinent to everyday life such as family, hobbies and interests, work, travel and current events
B2	15–30 words	a limited range of complex structures	sufficient range of vocabulary for most general topics with varied formulation
C1	20–35 words	a wide range of complex grammatical structures	broad range including idiomatic expressions and colloquialisms

Difficulty level of the input

The input is likely to be comprehensible (i.e. reading and understanding most of the details) to a language learner at a CEFR level just below the target level. In other words, an item targeting A2 is intended to be comprehensible at A1, an item targeting C1 is intended to be comprehensible at B2+, etc.

iii. What the user needs to do (the expected response)

Throughout the Grammar and Vocabulary section, the responses involve selecting from a choice of three possible answers.

At A2 level to C1 level, users choose a word or phrase to fill a gap in a short text or two-turn dialogue.

In responding to the EnglishScore Grammar and Vocabulary section, users are expected to draw on their knowledge of *personal/daily life*, *basic communication needs* and *common, general, non-specialised knowledge*.

Examples of Grammar and Vocabulary items

Items at all levels follow a 'gap fill' or 'chat' format.

Target: A2 Gap Fill

13:52 59%

27:53 Grammar and Vocabulary

Looking for an apartment

I think this apartment is good, but I prefer the other one because it's _____

more cheap

cheaper

cheaply

Next

Target: B1 Chat

13:52 59%

27:53 Grammar and Vocabulary

New manager

We've been wondering _____ for some time when the new regional manager is going to come.

I heard she's coming next month.

We've been wondering

We had wondered

We'll be wondering

Next

EnglishScore Reading section

i. Rationale

Overview of the section

Each user will encounter a different combination of questions from among the following types, depending on their level (based on their results on the Grammar and Vocabulary section).

- Targeting A2 level, reading a short, incomplete text. Choosing one of three options to complete the text
- Targeting B1 level, reading an opening sentence and ordering three further sentences to create a coherent text
- Targeting B2 level, reading a short text and selecting a suitable heading from three options
- Targeting C1 level, reading a more complex text and answering three questions about the text by, for each question, selecting the correct answer from among three options.

Connection to language use

At A2 to B1 level, tasks are based on situations commonly encountered in everyday personal and public contexts, e.g. getting out and about, carrying out simple transactions and handling short social exchanges, including, at B1, exchanges of opinion and coping with less routine situations. At B2 to C1 level, tasks are based on situations commonly encountered in the workplace, the media or in public life.

Instructions

The instructions are given in English. All instructions for the Reading section are presented on screen in writing.

Timing

The test taker has between 45 seconds and 4.5 minutes to answer each question, depending on the item type. It is possible for the test taker to answer and move to the next question before the time is complete.

ii. What the user reads (the input)

The input

As the CEFR level targeted by the item increases, the domain of the input changes: at A2 level, the input mainly relates to the *Personal* domain, at B1 level, the *Public* domain and at B2 and C1 levels, the *Occupational* domain. Each item has a short title to establish the text context for the test-taker except for B2 items, where the test-taker themselves must select the correct title.

Communication themes

Communication themes may include *personal identification/house, home, environment/daily life/free time/entertainment/relations with other people/shopping/food and drink/places and weather* (Council of Europe, 2001, p. 52).

Skills assessed

The Reading test items assess a range of different reading skills, including:

- understanding of the main idea of the text
- understanding at the paragraph level, including knowledge of discourse and organisational features
- global understanding of a text
- detailed understanding of information, ideas, the writer's opinions and feelings in a text.

Sources of input

The input is prepared by the item writers specially for EnglishScore. Items are written to represent the types of text that test-takers would read in the relevant domains.

Nature of input

At lower levels, the input is mostly very *concrete* and *familiar*, e.g. about *where people live, people they know and things they have, common objects around them*, progressing to less familiar and more abstract topics at the higher levels, e.g. *a wide range of both familiar and unfamiliar topics normally encountered in personal, social or vocational life*.

Extent and linguistic characteristics of the input

Level	Length	Grammar	Vocabulary
A2	Up to 30 words	sufficient range of vocabulary for routine, everyday transactions involving familiar situations and topics	simple grammatical structures and sentence patterns
B1	45–60 words	sufficient range of vocabulary to express most topics pertinent to everyday life such as family, hobbies and interests, work, travel and current events	some complex sentence forms
B2	Around 100 words	sufficient range of vocabulary for most general topics with varied formulation	a limited range of complex structures
C1	Around 250 words	broad range including idiomatic expressions and colloquialisms	a wide range of complex grammatical structures

Difficulty level of the input

The input is likely to be comprehensible (i.e. reading and understanding most of the details) to a language learner at a CEFR level just below the target level. In other words, an item targeting A2 is intended to be comprehensible at A1, an item targeting C1 is intended to be comprehensible at B2+, etc.

iii. What the user needs to do (the expected response)

At all levels, the person being assessed chooses from among three options. In responding to the EnglishScore Reading section, users are expected to draw on their knowledge of *personal/daily life*, *basic communication needs* and *common, general, non-specialised knowledge of the world*.

Examples of Reading items

Target: A2 Long Gap Fill

13:52 59%

27:53 Grammar and Vocabulary

Shopping at the market

John always likes to go to the market on Saturday mornings. It is usually very busy because it sells the best fresh food.

He _____

enjoys having dinner there with his friends

brings fresh food from the farm to his home

brings the food home and cooks lunch for his family

Next

Target: B1 Drag and Drop Chat

13:52 59%

27:53 Reading

Drag the sentences into the correct order

The teddy bear

One good thing about plastic is that it is strong and lasts a long time.

A toy shop owner saw the cartoon and created 'Teddy's bear'. The toy was a great success and remains popular to this day.

However, Roosevelt did not pull the trigger. This story became a newspaper cartoon.

In a hunting trip, Roosevelt's assistants saw a bear, inviting him to shoot it.

Next

Target: B2 Gap Fill Title

13:52 59%

27:53 Reading

The director of an environmental think-tank has pointed to a new study as evidence that social policy has a role to play in the fight against global warming. Richard Wilcott, head of the Institute for Environment Research (IER), said the study found that the rate of global warming dropped in conjunction with historical events that slowed economic activity. Mr Wilcott said the study proved it was necessary to strictly control activities known to be linked to global _____

Choose the best title for the text

Immediate actions against global warming are needed.

The role of historical events in economic activity.

The place of public policy to stop global warming.

Next

Target: C1: Long Text

13:52 59%

27:53 Reading

Challenges in management

Most business executives know that the competitive battles ahead of their company will involve not only ensuring that their company gets better at what it does but is also different from others. To make a difference, they need talented managers who can produce more imaginative strategies for growth and improvement, make faster decisions, be more flexible, be better prepared to anticipate threats and opportunities, and who can _____

Why have companies failed in improving their competitiveness efficiency?

Authority for decision-making is too restricted.

Customer needs are not given sufficient priority.

There are too many layers of management.

Next

EnglishScore Listening section

i. Rationale

Overview of the section

Each user will encounter a different combination of questions from among the following types, depending on their level (based on their results on the Grammar and Vocabulary section).

- Targeting A2 level, listening to a brief monologue (instruction or announcement) or dialogue (conversation with a minimum of four turns) on a concrete, familiar topic and choosing from among three options focussed on specific, usually factual information
- Targeting B1 level, listening to a brief monologue (instruction or announcement) or dialogue (conversation with a minimum of four turns) on a concrete, familiar topic and choosing from among three options focussed on the main points in the recording
- Targeting B2 level, listening to a monologue (instruction or announcement) or dialogue (conversation with a minimum of four turns) relating to the personal, public or occupational domain that might be a conversation, business meeting, news, announcement or instructions and choosing from among three options focussed on the main points in the recording, inferred meanings and conclusions
- Targeting C1 level, listening to a monologue (instruction or announcement) or dialogue (conversation with a minimum of four turns) relating to the public or occupational domain that might be a lecture or talk on work-related issues, current events or news and choosing from among three options focussed on main points and specific details, inferred meanings or the speaker's attitude and line of argument.

Connection to language use

At A2 to B1 level, tasks are based on situations commonly encountered in everyday personal and public contexts, e.g. getting out and about, carrying out simple transactions and handling short social exchanges, including, at B1, exchanges of opinion and coping with less routine situations. At B2 to C1 level, tasks are based on situations commonly encountered in the workplace, the media or in public life. Tasks are written to ensure they do not require any specialist knowledge to answer them.

Instructions

The instructions are given in English. All instructions for the Listening section are presented on screen in writing.

Timing

The test taker has between 105 and 120 seconds to answer each question. It is possible for the test taker to answer and move to the next question before the time is complete.

ii. What the user listens to (the input)

The input

As the CEFR level targeted by the item increases, the domain of the input changes: at A2 level, the input mainly relates to the *Personal* domain, at B1 level, the *Public* domain and at B2 and C1

levels, the *Occupational* domain. Each item has a short title and rubric to establish the listening context for the test-taker.

Communication themes

Communication themes may include *personal identification/house, home, environment/daily life/free time/entertainment/relations with other people/shopping/food and drink/places and weather* (Council of Europe, 2001, p. 52).

Sources of input

The input is prepared by the item writers specially for EnglishScore. Spoken input is read by actors, working from a script, and represents the types of listening texts test-takers would hear in the relevant domain.

Nature of input

At lower levels, the input is mostly very *concrete* and *familiar*, e.g. about *where people live, people they know and things they have, common objects around them*, progressing to less familiar and more abstract topics at the higher levels, e.g. *a wide range of both familiar and unfamiliar topics normally encountered in personal, social or vocational life*.

Extent and linguistic characteristics of the input

Level	Word count and pace	Grammar	Vocabulary
A2	65 words, slow and clearly articulated (approx 150 wpm ⁴)	sufficient range of vocabulary for routine, everyday transactions involving familiar situations and topics	simple grammatical structures and sentence patterns
B1		sufficient range of vocabulary to express most topics pertinent to everyday life such as family, hobbies and interests, work, travel and current events	some complex sentence forms
B2		sufficient range of vocabulary for most general topics with varied formulation	a limited range of complex structures
C1	120 words, fast natural rate (approx 190 wpm)	broad range including idiomatic expressions and colloquialisms	a wide range of complex grammatical structures

⁴ Words per minute.

Spoken delivery

The delivery ranges from a slow and clearly articulated rate at A2 level to a fast natural rate at C1 level. Recordings are delivered with a range of accents.

There is minimal background noise or distortion on the recordings. The recordings all involve one or two speakers. The users may play each recording twice.

Difficulty level of the input

The input is likely to be comprehensible (i.e. listening/reading/viewing and understanding most of the details) to a language learner at a CEFR level below the level of the item.

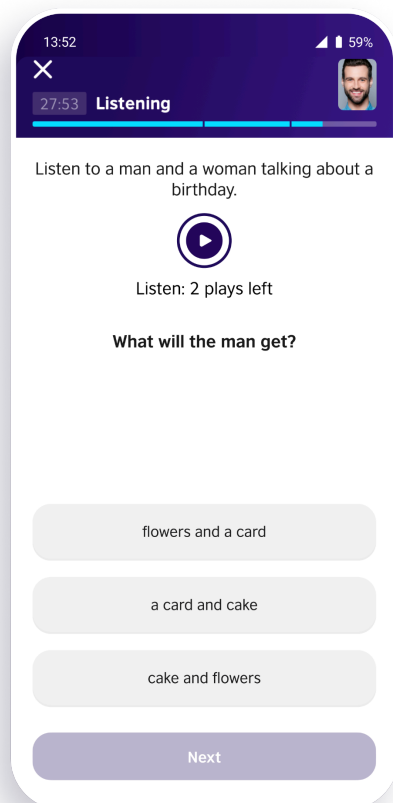
iii. What the user needs to do (the expected response)

At all levels, the person being assessed chooses from among three options. In responding to questions in the EnglishScore Listening section, users are expected to draw on their knowledge of *personal/daily life, basic communication needs* and *common, general, non-specialised knowledge of the world*.

Examples of Listening items

Listening items at all levels have a similar format.

Target: A2



The user hears:

[Man]: *I'm just going to the shops. Can I get you anything?*

[Woman]: *Have you got something for mum yet?*

[Man]: *Why?*

[Woman]: *You know, it's her birthday today.*

[Man]: *Oh no! I completely forgot.*

[Woman]: *Uh, I thought so. I've ordered a cake from the bakery. Can you get it on your way home?*

[Man]: *OK. Shall I buy some flowers and a card?*

[Woman]: *I've already got a card but flowers would be nice.*

[Man]: *OK.*

c. Scoring assessment tasks

The elements presented above focused on developing and assembling tests. The phases of marking, grading and reporting test results have different aims, such as:

- to provide an accurate and reliable mark for each person taking the assessment (marking)
- to put the performances of each person taking the assessment into a meaningful category so the test result can be understood (grading)
- to provide people taking the assessment with the results needed for an appropriate interpretation (reporting results).

Marking (marking and rating) includes presentation of different rating scales, the rating process and the rater training (Council of Europe, 2011, pp. 38–45).

Scoring procedures

All sections of EnglishScore are scored automatically according to a programmed answer key.

Scoring accuracy

The accuracy of the scoring is the responsibility of the Head of Assessment who conducts regular checks to confirm that automated answer keys are being applied correctly (Council of Europe, 2011, pp. 79–80).

d. Reporting scores on EnglishScore as a whole

Reporting scores

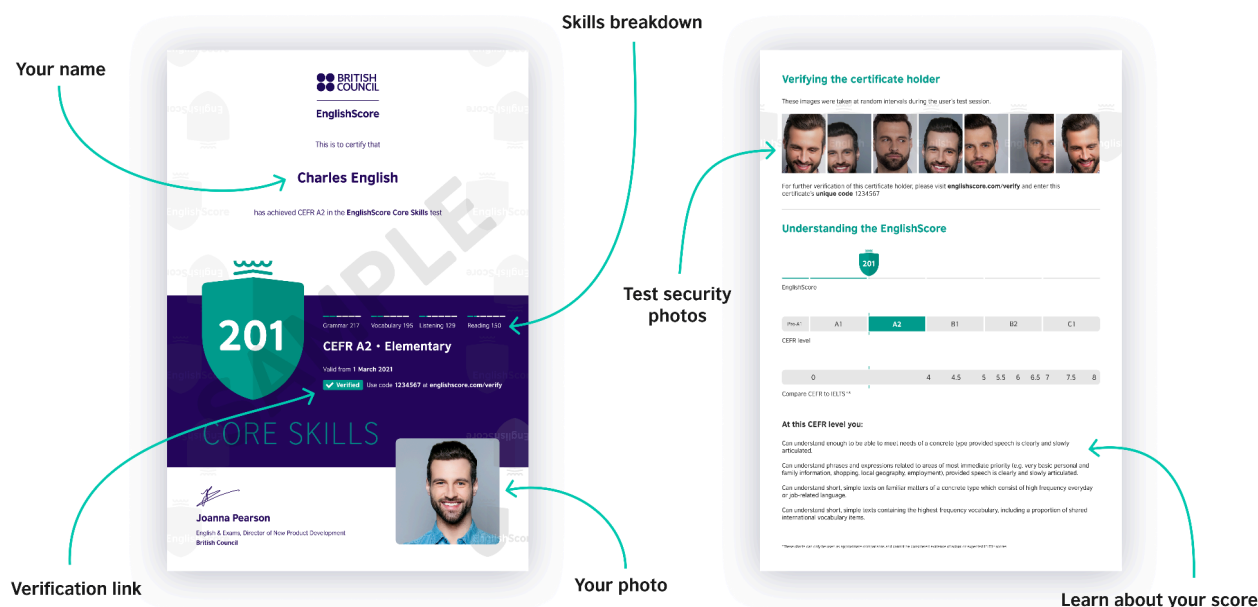
On completing the EnglishScore Core Skills test, the user is given an online report stating their overall Core Skills result with separate scores for Grammar, Vocabulary, Reading and Listening.

Estimated correspondences to CEFR level are also provided:

EnglishScore range	CEFR level
0–99	Below A1
100–199	A1
200–299	A2
300–399	B1
400–499	B2
500–599	C1

Test-takers have the option of purchasing a certificate as a record of their score. Each certificate includes the test-taker's name, a photograph of the test-taker taken during the administration, a verification ID for use by employers or other score users, and an overall score as well as skill scores. This certificate is available online or printed on paper.

This is an example Core Skills certificate:



Pass marks

There are no pass marks for EnglishScore. Scores are reported in relation to the Common European Framework of Reference for Languages (CEFR) from A1 to C1. Estimates of a user's CEFR level are based on their success in responding to material targeting each level. Further work will be undertaken to set standards in relation to the CEFR and to performance in other tests.

EnglishScore scale

The EnglishScore is a numeric, granular scale which measures English language proficiency from 0 to 599. It builds on the Common European Framework of Reference (CEFR) by showing finer gradations within a learner's CEFR level and can therefore help to measure gradual improvements in a test-taker's English level across the different skills. As well as providing useful and motivating feedback to test-takers, it also gives teachers and other decision makers a more detailed understanding of test-takers' strengths and weaknesses.

Time for results

Core Skill test score results are reported instantly, as soon as a test-taker has completed the test.

Reporting

At the end of the test, the test-taker's Core Skills result is reported as a score from 0 to 599 on the EnglishScore scale, as well as the corresponding CEFR level. In addition, a breakdown of skills and competencies (grammar, vocabulary, reading, listening) is provided, as well as a set of can-do statements to provide additional context to the reported test score.

e. Assessment results and analysis

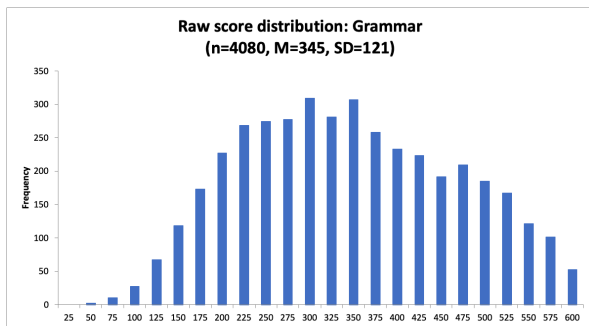
Data

Test response data is routinely collected and analysed for EnglishScore. Further data about users is additionally collected with consent.

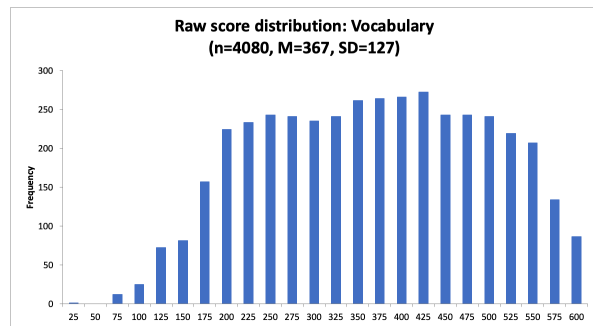
Test means and standard deviations

The following represent results for a sample of just over 4,000 users for Grammar and Vocabulary and nearly 9,000 users for Reading and Listening who took the EnglishScore test in July 2022. In all cases, scores are reported on a scale that ranges from 0 to 599.

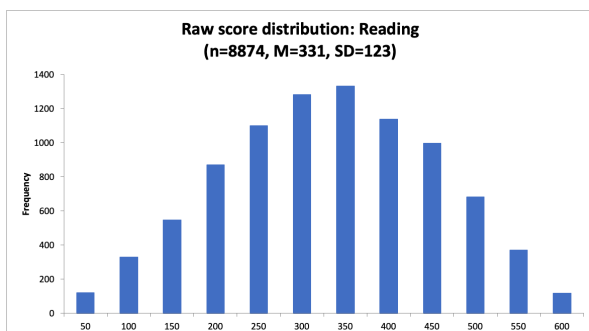
Grammar EnglishScore: Mean 345.0, Standard deviation 121.4



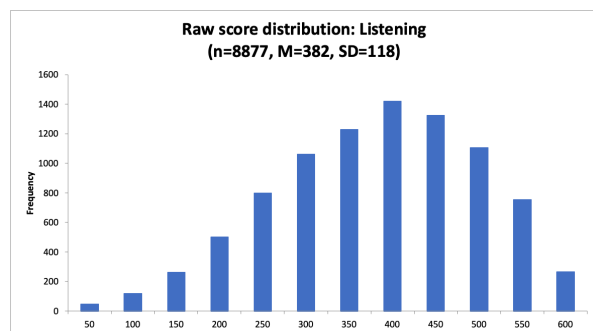
Vocabulary EnglishScore: Mean 366.6, Standard deviation 126.7



Reading EnglishScore: Mean 330.7, Standard deviation 123.0



Listening EnglishScore: Mean 381.8, Standard deviation 117.8



Correlations between sections

	Grammar and Vocabulary	Reading
Grammar and Vocabulary	--	--
Reading	0.938	--
Listening	0.941	0.942

The correlations between Grammar and Vocabulary, Reading and Listening range from 0.938 to 0.942, indicating a strong positive relationship between the sections. This suggests that users who perform well in one section also tend to perform well in the others.

Reliability

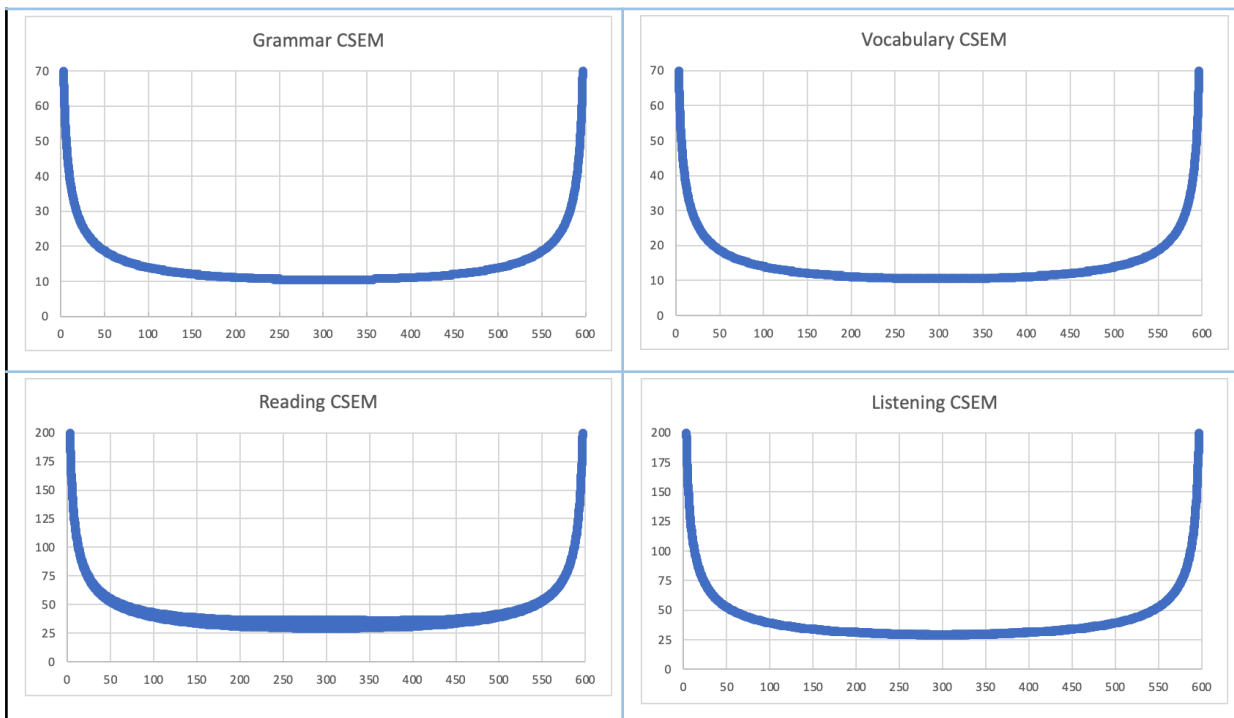
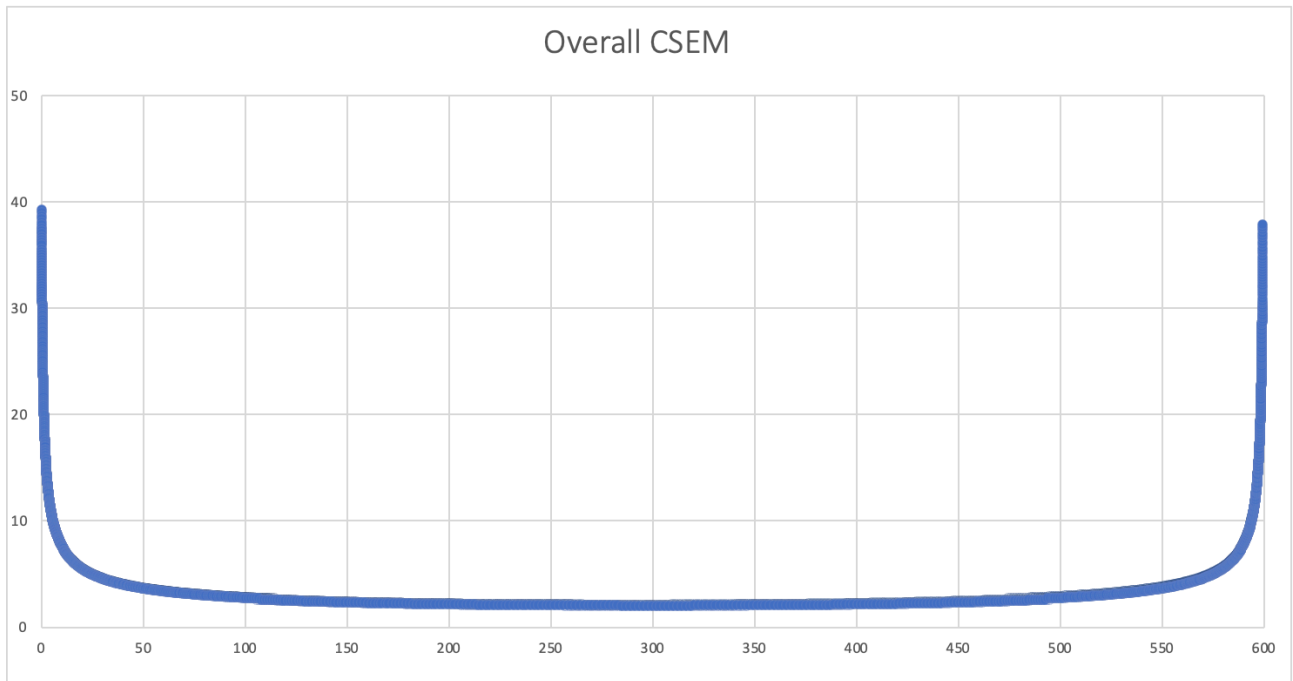
The reliability of a test is an estimate of the extent to which scores are likely to be consistent when a test is administered on different occasions or when it involves different test material. It assumes that a test score is not simply a reflection of a user's 'true score' or level of ability, but that it also involves 'error' resulting from factors other than ability (such as the user's mental or physical state, knowledge of the content or luck in guessing some of the answers). The user's true score is not a single, precise number but is estimated (with a specified degree of confidence) to lie somewhere between two score points. The range between the two scores will be relatively wide when the reliability of the test is low (as there is greater error and less certainty about the true score) and relatively narrow when the test has high reliability.

The reliability of EnglishScore is estimated using Rasch conditional standard errors of measurement (CSEM).

The line graphs below display the CSEM for EnglishScore overall scores and for each of the reported scores (Grammar, Vocabulary, Reading and Listening). The CSEM indicates the degree of measurement error in EnglishScore units but varies in relation to a user's score in each section.

The CSEM is more helpful than a single test reliability figure because it reflects differences in the precision of measurement at different points on the scales. Measurement is more precise for users with scores associated with small CSEMs and scores are at their most reliable towards the midpoint of each scale.

CSEMs can be used to construct a confidence interval around each point on a scale. Approximately, adding and subtracting $2 \times \text{CSEM}$ to a score on the scale indicates a user's true level of ability on the tested material with 95% confidence. For an overall EnglishScore of 350, with CSEM of 2, a user's true ability is likely to fall between 346 and 354 on the overall EnglishScore scale.



Test–retest reliability

An additional measure of test reliability is the test–retest coefficient. This measures the correlation between scores when a test-taker attempts the test more than once. A valid and reliable test should produce scores which are consistent and only vary slightly from one test attempt to the next (assuming no learning has taken place between attempts). The test–retest reliability coefficient for the Core Skills Test was 0.83 for tests taken between June 2023 and December 2023, where the same test-taker had taken the test twice in 28 days or fewer. This compares well with similar mid-to high-stakes tests. The total number of test repetitions used to calculate the coefficient was over 220,000.

f. Item review and health

The item review process helps to ensure that the items are valid, reliable and appropriate for the intended population. By identifying and removing poorly constructed or ambiguous items, the review process helps to reduce measurement error and increase the precision of language measurements.

In the item review process, the items flagged according to the Rasch parameters are identified and reviewed by our team of language assessment experts and item-writing specialists.

There are three main stages in the review process:

Pre-review stage

At this stage, the flagged items are identified based on defined Rasch parameters such as item difficulty and item fit. The flagged items are then listed and prepared for review. The preparation may include gathering additional information such as distractor analysis.

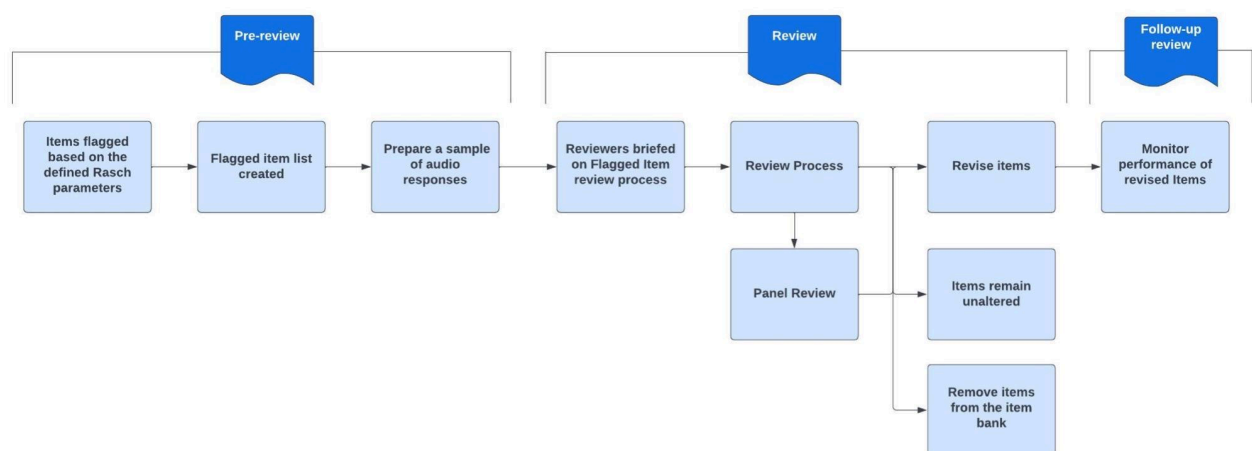
Review stage

At this stage, the flagged items are reviewed by experts who evaluate the items for quality and appropriateness. The review process may involve revising the items, leaving them as they are or removing them from the item bank. Additionally, a panel review is conducted where multiple experts review the flagged items and provide feedback on their quality and appropriateness. This helps to ensure that the items are of high quality and meet the necessary criteria for use in the language assessment.

Follow-up review stage

After the initial review, the updated items are monitored to ensure that they meet the necessary criteria and are functioning as intended. This stage may involve further review or modifications to the items as needed.

Figure 2. Analysing flagged responses in Core Skills Test: item review workflow.



The implementation of a rigorous item review process results in more precise evaluations of language proficiency and better-informed judgements about the language abilities of individuals taking the EnglishScore test. In addition, the findings from the reviews inform the item development process, enabling the refinement of future test items and enhancing the overall quality of the EnglishScore Core Skills Test.

Contact Information

About the British Council

The British Council builds connections, understanding and trust between people in the UK and other countries through arts and culture, education and the English language.

We work in two ways – directly with individuals to transform their lives, and with governments and partners to make a bigger difference for the longer term, creating benefit for millions of people all over the world.

We help young people to gain the skills, confidence and connections they are looking for to realise their potential and to participate in strong and inclusive communities. We support them to learn English, to get a high-quality education and to gain internationally recognised qualifications. Our work in arts and culture stimulates creative expression and exchange and nurtures creative enterprise.

We connect the best of the UK with the world and the best of the world with the UK. These connections lead to an understanding of each other's strengths and of the challenges and values that we share. This builds trust between people in the UK and other nations which endures even when official relations may be strained.

We work on the ground in more than 100 countries. In 2019–20, we connected with 80 million people directly and with 791 million overall, including online and through our broadcasts and publications.

Contact EnglishScore

For questions about the test, including content development, test scoring, security or certification, please contact us at contact@englishscore.com

Bibliography

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Strasbourg: Council of Europe. Available from: rm.coe.int/1680459f97 [Accessed 18 June 2022].

Council of Europe (2011). *Manual for Language Test Development and Testing*. Strasbourg: Council of Europe. Available from: <https://rm.coe.int/manual-for-language-test-development-and-examining-for-use-with-the-ce/1680667a2b> [Accessed 18 June 2022].

Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion Volume*. Strasbourg: Council of Europe. Available from: <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4> [Accessed 18 June 2022].

English Profile (2015). *Reference Level Descriptions* [Online]. Cambridge: Cambridge University Press. Available from: <http://englishprofile.org/the-cefr/reference-level-descriptions> [Accessed 18 June 2022].

North, B., Figueras, N., Takala, S., Van Avermaet, P., and Verhelst, N. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A Manual*. Strasbourg: Council of Europe. Available from: <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680667a2d> [Accessed 18 June 2022].