

# **Oxford Test of English**

## Test specifications

*Test development and validation May 2021*

## **Contents**

1	Introduction	3
2	Test description and rationale	3
3	Quality assurance	4
4	The test development process	4
5	Alignment to the CEFR	12
6	Test delivery	14
7	Accessibility	14
8	Test marking and scoring	15
9	Results reporting	17
10	Results reviews and appeals	18
11	Test monitoring, impact and review	18
12	Acknowledgements	19
13	References	19

## **Appendices**

	Appendix 1 – Oxford Test of English Speaking criteria	20
	Appendix 2 – Oxford Test of English Writing criteria	22
	Appendix 3 – Sample responses and marking commentaries	23

## 1 Introduction

This paper provides an overview of the development and validation of the Oxford Test of English. It sets out the rationale behind the need for the test, how it was developed, and the procedures employed to ensure and maintain its quality. The development stages include:

- the rationale behind developing the test
- the test design process
- the development of the test specifications
- the procedures for the production of test material
- the processes involved in aligning the test to the Common European Framework of Reference for Languages (CEFR).

## 2 Test description and rationale

Many English language learners require evidence of their proficiency in relation to the levels of the Council of Europe's Common European Framework of Reference for Languages (CEFR). This can be for educational purposes, such as proving that the learner meets entry requirements to a higher education course, or for professional purposes, such as proving the learner's ability in English to a prospective employer. The Oxford Test of English was developed to meet these needs, providing a valid and reliable measurement of English proficiency at A2, B1, and B2 levels. Performance below level A2 is indicated as 'Below A2' in test results.

The content of the test is independent of any specific course of study and reflects a wide range of English language learning programmes, so no test-specific course needs to be followed. Free downloadable practice materials are available at [oxfordtestofenglish.com](http://oxfordtestofenglish.com). With respect to the number of guided study hours required, this will depend on a variety of factors such as age, motivation, teaching methodology, etc. Table 1 provides general guidance on the approximate number of hours of study.

Table 1: Approximate number of guided study hours per CEFR level

CEFR level	Approximate number of hours
A2	200
B1	375
B2	500

The Oxford Test of English focusses on English language learners' ability to both understand and communicate in English, as measured by four modules:

- Speaking
- Listening
- Reading
- Writing.

All modules are delivered entirely online and can be taken individually, or in any combination, on an on-demand basis.

There are two versions of the Oxford Test of English: the Oxford Test of English (suitable for test takers aged 16 or above), and the Oxford Test of English for Schools (suitable for test takers aged 12 to 16). Both versions of the test have the same format and are certified by the University of Oxford. The Oxford Test of English for Schools treatment of topics is more appropriate for the 12 to 16 age group and Speaking Module Voicemail 1 is informal, while in the Oxford Test of English it is formal. Test takers aged 16 can choose to take the Oxford Test of English or the Oxford Test of English for Schools.

As the format and administration of both versions of the test is the same, any reference to the Oxford Test of English in this document applies to both versions of the test unless specified.

### 3 Quality assurance

The Oxford Test of English is produced by Oxford University Press (OUP), a department of the University of Oxford. As a result of quality audits carried out by the University's Department of Continuing Education on behalf of the University of Oxford Education Committee, the University of Oxford officially certifies the Oxford Test of English.

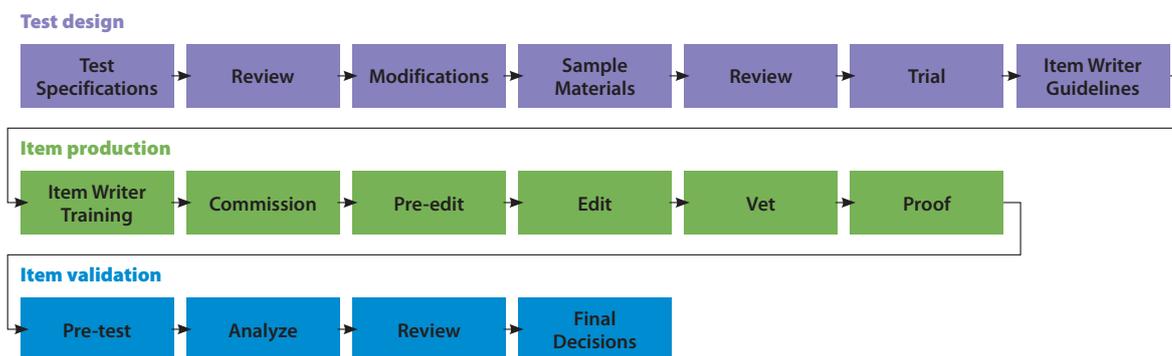
The audits represent a continuous process aimed at maintaining and improving the quality of the Oxford Test of English. They involve scrutiny of the different stages of design, production, and administration. The process continues beyond the launch of the test and includes regular reviews of test administrations to ensure that every test taker receives a fair and valid result.

### 4 The test development process

The test was developed through an iterative design process (see Figure 1), involving:

- initial test design
- drafting of specifications
- production of sample materials
- reviews by internal assessment staff and external assessment consultants
- modification on the basis of the reviews
- trialling with students in teaching centres around the world
- test production
- pretesting
- analysis and review
- item banking.

Figure 1 – The test development process



The Oxford Test of English reflects current language teaching and learning methodology. The test is designed to emulate the kinds of tasks that language learners encounter outside of test and classroom settings so that users of test results can be confident that test takers are able to perform real-world tasks.

## 4.1 Test design

The first phase of test development involves producing comprehensive test specifications. The specifications detail the test format, the content for each of the modules and each of the tasks contained within them. Well-crafted specifications communicate the test designers' vision, underpinning the consistency of measurement (i.e. reliability) across modules, enhancing the quality of the test across administrations and helping to ensure that decisions made based on test scores will be fair and valid.

In creating the specifications for the Oxford Test of English, OUP worked closely with institutions, teachers and learners to ensure that the test met their needs, while making certain that the test was also aligned to OUP's approach to language teaching, learning and assessment.

The specifications for the Oxford Test of English were derived from:

- level and domain descriptions in the CEFR: each task in the test is related to one or more CEFR Can Do descriptors
- communicative teaching practice
- course outlines and content from OUP teaching materials.

The test is designed to cover as wide a range of domains as possible within the confines of a two-hour administration.

Independent language-testing professionals were invited to comment on the draft specifications to help ensure appropriate coverage of domains and levels. These draft specifications were reviewed by an internal OUP panel and revised ahead of the production of sample materials. The specifications were then reviewed a second time, along with these sample materials, and further modifications were made.

Experienced item writers were commissioned to draft item writer guidelines for each module, based on the specifications and sample materials. These guidelines help our item writers to produce comparable, good-quality tasks to ensure consistency across different instances of the test and to ensure that tasks continue to reflect the intentions of the designers.

A team of item writers was trained to write an initial set of test materials. These fed into small-scale trialling in which groups of students were asked to take these tasks and provide feedback on the experience. Another round of minor revisions was then made based on the comments from other item writers and from trial students. Further sets of materials were then commissioned. These were pretested more extensively on representative samples of students in a range of countries worldwide.

## 4.2 Test format

All modules are delivered online so the test format was developed to reflect modern communication methods and includes task types not usually covered in traditional paper-based tests. Examples of this include an email activity in the Writing module and leaving a voicemail message in the Speaking module. Online delivery also meant that aspects of language proficiency that cannot easily be tested in paper-based tests could be incorporated, such as timed reading tasks. By allocating specific times to tasks it is possible to differentiate between speed, or expeditious, reading activities and careful reading exercises, which require more time. Efforts have also been made to tap into inferred or pragmatic meanings, as well as testing more concrete understanding. A key element of the test has also been to ensure that the CEFR is covered, not just in terms of level, but also with regard to the breadth of domains covered in each skill.

The test is broken up into four modules which can be taken together in one sitting or individually. All four modules are timed, and test takers move from task to task either by selecting a 'next' button on completion of a task, or by being automatically moved to the next task at the end of the allotted time. Table 2 shows an overview of the Oxford Test of English.

Table 2: Oxford Test of English overview

Module	Part	No. tasks	No. items	Structure	Timing
<b>Speaking</b>	Part 1	2	6 (+ 2 unassessed)	Interview: eight spoken questions on everyday topics	Approx. 15 minutes
	Part 2	2	2	Two voicemails with spoken and written input	
	Part 3	1	1	A talk on an issue or scenario, with spoken and written input and picture prompts	
	Part 4	1	6	Six spoken questions related to the theme of the Part 3 talk	
<b>Listening</b>	Part 1	5	5	Five discrete short monologues/dialogues with picture options, each with one question	Approx. 30 minutes
	Part 2	1	5	A longer monologue with a note-completion task	
	Part 3	1	5	A longer dialogue with a task focusing on identifying opinions	
	Part 4	5	5	Five discrete short monologues/dialogues with text options, each with one question	
<b>Reading</b>	Part 1	6	6	Six short texts from a variety of sources, each with one question	35 minutes
	Part 2	1	6	Six items to match with three profiles of people OR Six profiles of people to match with four longer text descriptions	
	Part 3	1	6	Six extracted sentences are inserted into a longer text	
	Part 4	1	4	A longer text with four questions	
<b>Writing</b>	Part 1	1	1	Email (80–130 words)	45 minutes
	Part 2	1	1	Essay (100–160 words) OR Magazine article or Review (100–160 words)	

### 4.2.1 Speaking module

There are four parts in the Speaking module.

In Part 1, test takers are asked to respond to eight spoken single-sentence questions on everyday topics. The first two questions are for practice purposes and are not assessed.

In Part 2, test takers are required to leave two voicemail messages.

In Part 3, test takers give a one-minute talk based on visual and audio prompts.

In Part 4, test takers answer six audio questions based on the topic of the talk presented in Part 3.

In the Speaking module, test takers wear a headset and speak into a microphone to answer questions delivered by computer. A clock displayed on the screen shows how much time is available to answer each question. Preparation time is given for the voicemails in Part 2 and for the talk in Part 3.

Input is either audio-only (i.e. the text of the task is heard, but not shown on screen) or audio-written (i.e. the text of the task is heard *and* shown on screen). Where preparation time is given, this is after the task has been presented and before the test taker has to begin speaking. Table 3 shows a summary chart of the tasks in the Speaking module.

Table 3: Overview of the Speaking module

Part	No. tasks	No. items	Structure	Testing focus
Part 1	2	6 (+ 2 unassessed)	<p><b>Interview</b></p> <p>Answering eight spoken single-sentence questions on everyday topics            Questions 1 and 2 are always the same and are given to all test takers            Questions 3–5 are topic related            Questions 6–8 are topic related (on a different topic to questions 3–5)</p> <p><b>Audience:</b> the audience is the interviewer/assessor</p> <p><b>Preparation time:</b> none</p> <p><b>Response time:</b> Questions 1 and 2: 10 seconds per question            Questions 3–8: 20 seconds per question</p>	<ul style="list-style-type: none"> <li>responding to questions</li> <li>giving factual information</li> <li>expressing personal opinions on everyday topics</li> </ul>
Part 2	2	2	<p><b>Voicemail message</b></p> <p>Leaving two voicemail messages</p> <p><b>Voicemail 1:</b> test taker leaves a voicemail</p> <p>Audio-visual input consisting of a situation with three prompts requiring the test taker to leave a voicemail</p> <p><b>Audience:</b> the audience is specified in the task, and the relation to that audience may be informal (e.g. friend) or neutral (e.g. shop manager)</p> <p><b>Preparation time:</b> 20 seconds</p> <p><b>Response time:</b> 40 seconds</p> <p><b>Voicemail 2:</b> test taker replies to a voicemail</p> <p>Audio-visual input consisting of a situation with three prompts, plus audio-only input (in the form of a voicemail which the test taker hears) requiring the test taker to leave a voicemail</p> <p><b>Audience:</b> the audience is specified in the task, and the relation to that audience is informal (e.g. friend)</p> <p><b>Preparation time:</b> 20 seconds</p> <p><b>Response time:</b> 40 seconds</p>	<ul style="list-style-type: none"> <li>organizing and sustaining extended discourse</li> <li>sociolinguistic appropriacy</li> <li>sustaining relationships</li> </ul>
Part 3	1	1	<p><b>Talk</b></p> <p>Audio-visual input in the form of a rubric and four photo prompts on an issue (e.g. what things are important for a happy life) or a scenario (e.g. how a language school can attract more students) on which the test taker gives a talk</p> <p><b>Audience:</b> the audience is specified and is typically the test taker's classmates</p> <p><b>Preparation time:</b> 30 seconds</p> <p><b>Response time:</b> 1 minute</p>	<ul style="list-style-type: none"> <li>organizing and sustaining extended discourse</li> <li>describing</li> <li>comparing and contrasting</li> <li>speculating</li> <li>suggesting</li> </ul>

<b>Part 4</b>	1	6	<b>Follow-up questions</b> Answering six audio-only single sentence questions related to the Part 3 talk <b>Audience:</b> the audience is the interviewer/assessor <b>Preparation time:</b> none <b>Response time:</b> 30 seconds per question	As in Part 3, plus: <ul style="list-style-type: none"> <li>• responding to questions</li> <li>• expressing, justifying and responding to opinions</li> <li>• expressing feelings</li> </ul>
---------------	---	---	--	---

### 4.2.2 Listening module

There are four parts in the Listening module.

In Part 1, test takers listen to five audio recordings and, choosing from a set of options, select one picture to represent the overall meaning or specific detail of each recording.

In Part 2, test takers listen to an informational/descriptive monologue and complete a set of notes consisting of five three-option multiple-choice items.

In Part 3, test takers listen to a longer dialogue and match five statements to the speaker who expresses them.

In Part 4, test takers listen to five recordings and answer one question per recording.

The timing of all parts of the Listening module is predetermined. In each part, test takers hear each recording twice and are given a set time to check their answers before the test automatically progresses to the next recording. Table 4 shows a summary chart of the tasks in the Listening module.

Table 4: Overview of the Listening module

Part	No. tasks	No. items	Structure	Testing focus
<b>Part 1</b>	5	5	<b>Multiple choice – picture options</b> Five short monologues/dialogues each with a three-option multiple choice picture option question Time to check answers: 10 seconds Audioscript length: A2 = 30–65 words; B1 = 55–85 words, B2 = 70–96 words.	Listening to identify: <ul style="list-style-type: none"> <li>• specific information</li> </ul>
<b>Part 2</b>	1	5	<b>Note completion</b> A longer monologue with a note-completion task Five three-option multiple-choice questions Time to check answers: 15 seconds Audioscript length: A2 = 150–250 words; B1 = 250–350 words, B2 = 350–450 words.	Listening to identify: <ul style="list-style-type: none"> <li>• specific information</li> </ul>
<b>Part 3</b>	1	5	<b>Matching opinions with people who say them</b> A longer dialogue with a task focusing on identifying opinions Five three-option multiple-choice questions Time to check answers: 15 seconds Audioscript length: A2 = 200–300 words; B1 = 300–400 words, B2 = 400–525 words.	Listening to identify: <ul style="list-style-type: none"> <li>• stated opinion</li> <li>• implied meaning</li> </ul>
<b>Part 4</b>	5	5	<b>Multiple choice</b> Five short monologues/dialogues each with a three-option multiple choice question Time to check answers: 10 seconds Audioscript length: A2 = 30–65 words; B1 = 55–85 words, B2 = 70–96 words.	Listening to identify: <ul style="list-style-type: none"> <li>• attitude/feeling/opinion</li> <li>• gist</li> <li>• function/reason/purpose</li> <li>• speaker relationship</li> <li>• topic</li> <li>• type/genre</li> </ul>

### 4.2.3 Reading module

There are four parts in the Reading module.

In Part 1, test takers read six short texts from a range of genres and answer one three-option multiple-choice question on each text.

In Part 2, test takers must quickly read six profiles of people with requirements and match each to one of four topic-related factual texts.

In Part 3, test takers read a text from which six sentences have been removed, leaving gaps. Test takers choose missing sentences from a list and insert one into each gap.

In Part 4, test takers read a text and answer four three-option multiple-choice questions about the content.

All texts used in the Reading module are based on authentic material intended to be of relevance or interest to a general readership. Texts may be formal, neutral or informal in register.

The time allowed for completion of each task in the Reading module is predetermined. If the test taker does not complete the task within the allotted time, the system will automatically progress to the next task. Table 5 shows a summary chart of the tasks in the Reading module.

Table 5: Overview of the Reading module

Part	No. tasks	No. items	Structure	Testing focus
<b>Part 1</b>	6	6	<p><b>Multiple-choice questions on short texts</b></p> <p>Six short texts from a variety of sources each with a three-option multiple choice question</p> <p>Texts may be adverts, blogs, emails, notes, notices, and text messages</p> <p>Time to process the texts and complete the tasks: 1 minute 20 seconds per task (8 minutes in total)</p> <p>Text length: A2 = 20–35; B1 = 20–50; B2 = 40–70.</p>	<p>Reading to identify:</p> <ul style="list-style-type: none"> <li>main message</li> <li>purpose</li> <li>detail</li> </ul>
<b>Part 2</b>	1	6	<p><b>Multiple matching</b></p> <p>Six items to match with three profiles of people</p> <p>Texts are factual and may be from magazine articles or blogs</p> <p>Eight minutes to process the texts and complete the tasks</p> <p>Time to for each profiles: A2 = 75–90 word; B1 = 145–160 words; B2 = 190–205 words.</p>	<p>Expeditious search reading at local and global levels to identify</p> <ul style="list-style-type: none"> <li>specific information</li> <li>opinion and attitude</li> </ul>
			<p>OR</p> <p>Six profiles of people to match with four longer text descriptions</p> <p>Texts may be from brochures, advertisements, or magazine articles</p> <p>Eight minutes to match the profiles and texts</p> <p>Time to process the texts and complete the task: 8 minutes</p> <p>Text length for each description: A2 = 45–60 word; B1 = 80–100 words; B2 = 100–125 words.</p>	<p>Expeditious search reading at local and global levels to identify</p> <ul style="list-style-type: none"> <li>specific information</li> <li>opinion and attitude</li> </ul>
<b>Part 3</b>	1	6	<p><b>Gapped text</b></p> <p>Six extracted sentences are inserted into a longer text</p> <p>Texts are from newspaper and magazine articles</p> <p>Six text-completion questions</p> <p>Time to process the text and complete the task: 11 minutes</p> <p>Text length: A2 = 200–220; B1 = 350–375; B2 = 400–425.</p>	<p>Reading to identify:</p> <ul style="list-style-type: none"> <li>text structure</li> <li>organizational features of a text</li> </ul>
<b>Part 4</b>	1	4	<p><b>Multiple-choice questions on longer texts</b></p> <p>Four three-option multiple-choice questions</p> <p>Texts are from newspaper and magazine articles</p> <p>Time to process the text and complete the task: 8 minutes</p> <p>Text length: A2 = approx. 235; B1 = approx. 280; B2 = approx. 350.</p>	<p>Reading to identify:</p> <ul style="list-style-type: none"> <li>attitude/opinion</li> <li>purpose</li> <li>reference</li> <li>the meanings of words in context</li> <li>global meaning</li> </ul>

#### 4.2.4 Writing module

There are two parts in the Writing module.

In Part 1, test takers read and respond to an input email. Responses are either informal or neutral and need to include three points from the input.

In Part 2, there is a choice of either writing an essay or a magazine article/review.

In both parts, test takers type their responses. The tasks specify a target audience and a minimum and maximum word count. There is an automatic word-count facility. Test takers will be penalized if their responses are under length.

There is a clock so that test takers always know how much time they have remaining for each part. Table 6 shows a summary chart of the tasks in the Writing module.

Table 6: Overview of the Writing module

Part	No. tasks	No. items	Structure	Testing focus
Part 1	1	1	<p><b>Email</b> 80–130 words Test taker responds to an email There are three points which the test taker must include in their email The response may be informal or neutral in tone Time to process the task and complete the response: 20 minutes</p>	<ul style="list-style-type: none"> <li>giving information</li> <li>expressing and responding to opinions and feelings</li> <li>transactional functions such as inviting/requesting/suggesting</li> </ul>
Part 2	1	1	A choice of writing tasks: an essay or a magazine article/review	
			<p><b>Essay</b> 100–160 words Writing an essay on a topic typical of classroom discussions Time to process the task and complete the response: 25 minutes</p>	<ul style="list-style-type: none"> <li>expressing and responding to opinions</li> <li>developing an argument</li> </ul>
			<p>or</p> <p><b>Magazine article/Review</b> 100–160 words Writing a general article (such as the profile of a famous sports person) or writing a review (such as a review of a website) The target reader is usually an English teacher Time to process the task and complete the response: 25 minutes</p>	<ul style="list-style-type: none"> <li>describing</li> <li>narrating</li> <li>expressing feelings and opinions</li> <li>recommending</li> </ul>

### 4.3 Test production

Before test tasks are accepted for use in the Oxford Test of English, procedures are systematically followed to ensure optimum test item quality. Rigorous adherence to such procedures helps to strengthen test quality and provides evidence that important decisions about learners' language proficiency, based on their test scores, will be valid and fair.

Our quality assurance process involves a number of steps. These include pre-editing, editing, vetting and proofreading before material is pretested (See Figure 1).

Teams of item writers, led by a team leader (an expert item writer), are commissioned to work on each test module. The initial commissioning of materials is followed by a pre-editing meeting. A panel of experts reviews the materials to ensure that they closely adhere to test specifications and item-writing guidelines. The panel asks for amendments and the materials are returned to the writers to make the required changes. Once all changes have been made, the materials are further scrutinized and refined in an editing meeting.

Changes made in editing meetings are then registered on the item database, at which time the materials are vetted by an external content expert. This step provides an independent view of the material and identifies any further improvements to the task. The vetter also helps to detect: (1) whether testing points are biased towards certain language groups or cultures; (2) if items are levelled appropriately across tasks; (3) the degree to which test content is accessible on a global level; and (4) whether the test items include any unwanted taboo topics (for example, alcohol and serious illnesses). This activity safeguards against threats to test fairness.

At this point, additional materials such as audio files and graphics are added. The tasks are then proofread for instances of formatting issues and typographical errors.

The purpose of the next step in the process – pretesting – is to determine the difficulty and effectiveness of the items for use in the official, or 'live', Oxford Test of English. Students who participate in pretesting sessions are representative of the same population of students who are targeted to take the Oxford Test of English.

Data from pretesting sessions is analysed by a team of research and validation experts who employ both quantitative and qualitative methods to determine item levelling, the quality of the item options, and fit statistics for the items across tasks and levels. The statistical output, generated by the analyses, are then used for a substantive review by a panel consisting of specialists from OUP and external experts. Following pretesting and review, materials may be accepted for use in the test, sent back to item writers to be rewritten and re-pretested, or rejected and discarded.

## 5 Alignment to the CEFR

The CEFR is now recognized around the world as a key framework for interpreting language proficiency. Many institutions base materials, teaching programmes and tests on the CEFR levels. In developing the Oxford Test of English, every effort has been made to ensure alignment to the CEFR.

The content of the Oxford Test of English is specifically designed to elicit performances at the following levels of proficiency: CEFR levels B2, B1, and A2. This means that a test taker taking the Oxford Test of English can receive one of four results: B2, B1, A2 or Below A2. Test takers who score below level A2 receive the result 'Below A2'. This grade indicates that they are below the levels reported in the test and that we cannot ascribe a specific CEFR level to their performance. Further information on score reporting can be found in Section 8.

The CEFR has been embedded in the development of the Oxford Test of English through a range of activities. These include:

- (1) employing CEFR **Can Do statements** in the test design
- (2) surveying **OUP course materials** at each of the CEFR levels
- (3) conducting **data analyses** on pretested items
- (4) **aligning the Oxford Test of English scale** to the Oxford Online Placement Test (OOPT). Pollitt (2009) refers to work done to align OOPT to the CEFR
- (5) conducting complementary **standard-setting activities** based on the Council of Europe's *Manual for Relating Language Examinations to the Common European Framework of Reference for Learning, Teaching, Assessment* (2009), henceforth referred to as 'the CEFR Manual,' to align test items to the CEFR across four modules.  
A brief explanation of these activities is presented below.

### (1) Can Do statements

In the design of the Oxford Test of English modules, careful attention was given to embedding links, in the form of descriptors, between the CEFR and the test items. A great effort was made to familiarize OUP item writers, item writer trainers, vetters and assessors with the CEFR with a view to linking the test specifications, item-writing guidelines and ultimately the test items to targeted CEFR Can Do statements.

### (2) OUP course materials

In the development of the test specifications for the Oxford Test of English, OUP surveyed the grammatical features, degree of syntactic complexity and frequency of the lexis typically featured in Oxford University Press ELT coursebooks. On the basis of this analysis, item types and item content were identified at each CEFR level. Such findings fed into the design of the test, which benefitted from both the common understanding of levels provided by the CEFR, and from OUP's long-term practical engagement in producing English language education materials.

### (3) Data analysis of pretested items

The test has been pretested around the world with over 10,000 students across thirty-seven countries from a wide number of first-language backgrounds at each of the targeted CEFR levels. Pretesting provides a good deal of information related to overall item quality (such as the quality of the item options), the performance of the test takers and the extent to which new test items could be scaled to the intended CEFR levels. Using Rasch analysis to evaluate objectively marked test tasks, a difficulty scale was plotted for the Oxford Test of English items based on Oxford Online Placement Test (OOPT) anchors [see (4) below]. Inferences about the CEFR levels can be made from such empirically-derived analyses.

#### **(4) Aligning the Oxford Test of English to the Oxford Online Placement Test**

To provide evidence of how well the Oxford Test of English is scaled to the CEFR levels, an *external-anchor design* was selected. Items from the Oxford Online Placement Test that had previously been related to the CEFR were administered to test takers as ‘anchor’ items alongside new material from the Oxford Test of English. An external-anchor design is often used in equating or scaling studies in which certain items link the performance of test takers across two test instruments which measure closely related knowledge and skills. (Dorans et al., 2010).

Through a series of statistical analyses, it was found that the Oxford Test of English functions on a similar scale to that of the Oxford Online Placement Test, thus providing evidence that if test takers took both tests, their test results could be interpreted on a shared scale. In other words, this provides further evidence that the Oxford Test of English and the Oxford Online Placement Test both map test takers to the CEFR in a similar way.

#### **(5) Standard-setting activities**

To strengthen inferences made from the data-driven analyses (in (3) and (4) above), a number of additional steps have been taken to ensure that the test items are appropriately aligned with the CEFR levels. Standard-setting (or benchmarking) activities were conducted to complement the pretesting-review process. Benchmarking activities, adapted from the CEFR Manual, are conducted with independent expert raters in a multi-step process:

- (a) The independent experts attend a series of webinars which provide a macro- and micro-view into what the learners at each CEFR level ‘can do’ and what the test tasks are designed to measure.
- (b) They are provided with the test specifications and item-writing guidelines.
- (c) They are shown numerous examples of the test tasks from each of the modules, at varying CEFR levels, after which they are polled to determine their level of agreement. This results in the assignment of a CEFR level estimate for that item.
- (d) An arbiter collects the poll results and instigates a discussion when rater disagreement requires additional adjudication. Several rounds of adjudication can occur before benchmark estimates can be established for each item.
- (e) After benchmarking activities are completed, additional analyses are conducted to adjust the calibration of the benchmarked items and reconcile these with the previously pretested items. The alignment of benchmarking results with pretesting difficulties allows us to identify cut points for the CEFR levels on the Oxford Test of English scale at the B2, B1, and A2 levels.

The above procedures all contribute to the alignment of the Oxford Test of English to the CEFR, and provide evidence for the Oxford Test of English score reporting scale (see Section 9).

## 6 Test delivery

Unlike more traditional paper-based or linear online tests, the Oxford Test of English does not have fixed test versions in which all test takers encounter the same set of questions. Instead, it operates using an item bank and a series of selection rules. An item bank is a large collection of test questions or items that can be used during the test. The large number of items helps to ensure that different test takers using the test at the same time receive different sets of questions. The Listening and Reading modules are computer adaptive, so the tasks adapt to the ability level of the test taker. The test selection rules determine which items are presented to each test taker, for example, 'choose five Part 1 Listening items'. Each item presented to the test taker is drawn from the bank using the selection rules and an algorithm which calculates the estimated ability of the test taker and the appropriate difficulty of the next task to be presented. A randomness element is also factored into the selection of tasks, so that each test taker receives their own individualized version of the test. The Speaking and Writing modules are not adaptive, but do exploit the randomness element. This approach has several advantages over traditional linear session-based tests. As test takers do not receive the same set of items, test security is improved, allowing the Oxford Test of English to be used on an on-demand basis, rather than limiting delivery to scheduled sessions. And, as the test is delivered wholly online, no materials need to be transported to test centres and stored on site, which also increases security. The item bank is refreshed on a regular basis to ensure that items do not become over-exposed. Finnerty (2015) gives further details about the advantages and workings of computer-adaptive testing (CAT).

The Oxford Test of English can only be administered by approved institutions (test centres), which are subject to ongoing quality-control checks and audits. Test centres have to provide evidence that they meet technical requirements and have the appropriate facilities and suitable staff to administer the test. Requirements for test centres are detailed in the *Oxford Test of English Test Centre Handbook*.

Once approved, a test centre can purchase test licences to run the test. The test centre then selects the date or dates on which they wish to run the test and allocates licences to that session. The Oxford Test of English can be taken on any date, though OUP usually requires seven days' notice of a test session – this ensures that sufficient assessors are allocated for the marking of Speaking and Writing modules. The Oxford Test of English is usually taken as a complete test (all four modules are administered in the course of a session), but test centres may choose to run sessions for single modules or any combination of modules. It is also possible for test takers to choose to resit individual modules, rather than resitting the whole test.

## 7 Accessibility

Oxford University Press is committed to providing accommodations to make the Oxford Test of English accessible to learners with special requirements where possible. Whilst there are some limitations to the range of accommodations that can be provided in an online test, OUP, as part of long term roadmap, will be adding additional functionality to its assessment system over the coming years to accommodate an increasing range of test taker special requirements. In the first phase of test launch, the following accommodations will be available in every test centre:

- additional time for Reading and Writing modules
- a range of colour contrast options
- increased font size.

Wherever possible, test centres will also provide the following to accommodate special requirements:

- building access for wheelchair users
- separate test sessions
- extended breaks between modules
- extra invigilation support.

Applications for special requirements are made by the test centre on behalf of the test taker. The option to adjust colour contrast and font size are applied to the test taker's test profile by OUP and *do not require* supporting medical documentation. Requests for additional time, extended breaks or a separate test session *need to be accompanied* by the appropriate medical certificate.

## 8 Test marking and scoring

### 8.1 Listening and Reading

The Listening and Reading modules employ an adaptive algorithm. Depending on whether correct or incorrect responses are received for each task, the system increases or decreases the difficulty of the following task as the test progresses. Responses for Listening and Reading are marked by computer and the ability of the test taker is estimated according to the responses given in relation to the difficulty of the questions presented. The Oxford Test of English employs the *Weighted Maximum Likelihood Estimation* (Warm, 1989) in its test algorithm. The equation in this formula uses the test taker's responses to items of different Rasch difficulties to estimate their ability at each decision point, i.e. at the end of each item or set of items.

As the test progresses, the estimate of the test taker's ability is refined using additional information from each item or set of items and the statistical error associated with the estimate is reduced. To ensure that each test taker has the same test experience, the Oxford Test of English delivers a standard test format to each test taker. That is, all test takers receive the same task types and the same number of items. The final ability estimate is derived once the complete set of test items in a module has been delivered. Ability estimates are then converted to a standardized score and this is also reported in terms of a CEFR level.

### 8.2 Speaking and Writing

Speaking and Writing tasks are selected at random from the item bank, according to a pre-defined number and order of tasks, and the responses are returned online and sent electronically to trained assessors who mark them according to analytic criteria (also known as 'rubrics') derived from the CEFR level descriptors. Analytic criteria are used as they ensure that assessors focus on a range of marking elements rather than focus too heavily on one area of the test taker's performance, as can be the case in holistic criteria.

Speaking criteria consist of Task fulfilment, Pronunciation and Fluency, Grammar, and Lexis. Table 7 summarizes the main elements of the marking criteria. See Appendix 1 for detailed Speaking marking criteria.

Table 7: Elements of the Speaking criteria

Task fulfilment	Pronunciation and Fluency	Grammar	Lexis
<ul style="list-style-type: none"><li>• Fulfilling task requirements</li><li>• Register</li><li>• Impact on the reader</li></ul>	<ul style="list-style-type: none"><li>• Phonological and word stress precision</li><li>• Phonological linking</li><li>• Intonation, rhythm, and stress</li><li>• Coherence</li><li>• Cohesion</li></ul>	<ul style="list-style-type: none"><li>• Range of structures</li><li>• Accuracy of structures</li></ul>	<ul style="list-style-type: none"><li>• Range of lexis</li><li>• Accuracy of lexis</li></ul>

Writing criteria consist of Task fulfilment, Organization, Grammar, and Lexis. Table 8 summarises the main elements of the marking criteria. See Appendix 2 for detailed Writing marking criteria.

Table 8: Elements of the Writing criteria

Task fulfilment	Organization	Grammar	Lexis
<ul style="list-style-type: none"><li>• Fulfilling task requirements</li><li>• Format</li><li>• Register</li><li>• Length</li></ul>	<ul style="list-style-type: none"><li>• Coherence</li><li>• Cohesion</li></ul>	<ul style="list-style-type: none"><li>• Range of structures</li><li>• Accuracy of structures</li></ul>	<ul style="list-style-type: none"><li>• Range of lexis</li><li>• Accuracy of lexis</li></ul>

The criteria are on an eight-point scale (0-7) ranging from below A2 to C1. There are detailed descriptors for bands 1, 3, 5, and 7, which represent the criteria required to demonstrate performance at the relevant CEFR level. These descriptors apply to performance at both the criterion level and the plus level for the CEFR level, e.g. the descriptors for band 3 relate to performance at both B1.1 and B1.2 (band 4). To decide whether a performance is at the criterion level or plus level, the assessor decides if all the descriptors have been 'minimally' or 'comfortably' met. If all positive descriptors have been comfortably met, a plus level is awarded. Assessors refer to the exemplar responses provided in the self-access standardization materials for examples of criterion and plus level performance at each CEFR level. Whilst the marking criteria cover below A2 to C1 levels, results are only reported up to B2 level. The C1 criteria are used as some B2 learners may demonstrate aspects of C1 criteria in their responses, allowing a greater range of marks to be awarded. The test does not award C-level grades as the tasks presented have not been designed for this purpose.

The length and relevance of the test taker response is taken into account when awarding marks. In the Speaking module, different penalties are applied depending on the extent and relevance of the response. In the Writing module, caps are in place depending on the extent and relevance of the response. See the criteria for further details.

Test taker responses are anonymized and split into two 'scripts', as shown in Table 9, and each script for a module is marked separately. The marks of the two scripts are combined and converted to a standardized score and CEFR level for each module.

*Table 9: Scripts in the Writing and Speaking modules*

	Script 1	Script 2
Speaking module	Speaking Part 1 and Part 2	Speaking Part 3 and Part 4
Writing module	Writing Part 1	Writing Part 2

See Appendix 3 for sample responses and marking commentaries.

### **8.3 Assessors and marking quality assurance**

All Speaking and Writing assessors have significant English language-teaching experience and recognized English language-teaching qualifications.

Assessors follow a standardized training and certification process before being allowed to participate in marking. Their marking is then monitored to ensure consistency.

Automated quality assurance monitoring is carried out using Speaking and Writing responses which have been marked previously by a number of experienced assessors and so have agreed benchmark ratings. These 'seeded' responses are interspersed with test taker responses to check that the assessors continue to be accurate in their marking to within set tolerances. All responses are anonymous, so assessors are unaware whether the responses they are marking are test taker responses or seeded responses. Assessors whose marking falls outside of agreed tolerances are removed from the marking process and asked to complete a re-standardization process, after which they can resume marking. Assessors who do not successfully complete re-standardization are permanently withdrawn from marking.

## 9 Results reporting

Performance on the Oxford Test of English is reported in terms of standardized scores on a scale ranging from 0 to 140. Standardized scores are independent of test sessions and give a standard reference point for students taking the test on different occasions.

Results are also displayed as a bar chart, showing how performance on the test relates to the relevant CEFR levels.

Table 10 shows the relationship between the Oxford Test of English scale and the CEFR.

Table 10: The Oxford Test of English scale and the CEFR

CEFR band	Oxford Test of English score range
B2	111–140
B1	81–110
A2	51–80
Below A2	0–50

The Oxford Test of English reports scores between CEFR levels B2 and A2. This means that although a test taker's responses may indicate performance that is above B2 level, a test taker cannot receive a test score above B2. The rationale for this is that the test taker has received tasks designed for CEFR levels B2, B1, and A2, so we cannot be certain how they would have performed on tasks designed for C1 or C2 test takers. The Oxford Test of English does, however, give an indication of 'Below A2' performance. Below A2-level performance means that a test taker is not at the level the test was designed to measure and that no precise statement of level can be made. For the objectively marked Reading and Listening modules, the final ability estimates obtained through the test algorithm are converted to standardized scores and these are used in determining the CEFR levels. For Speaking and Writing, marks are awarded by assessors, using the analytical marking criteria. These marks are then converted into standardized scores.

Test takers receive a standardized score and CEFR level on a Module Report Card for each module taken. If a test taker completes all four modules, they also receive an overall score and CEFR level on an Oxford Test of English Certificate. The overall score is calculated as an average of the scores obtained in each of the four modules. See Figure 2 for a sample test certificate.

Figure 2: Sample test certificate

**OXFORD TEST OF ENGLISH**

**Certificate of Proficiency**

TEST TAKER NAME	DATE OF BIRTH	TEST TAKER NUMBER		
Fernando García León	03 January 1990	123 456 789		
CERTIFICATE REFERENCE NUMBER				
ABC 456 789 123				
OVERALL CEFR LEVEL		OVERALL SCORE		
<b>B2</b>		<b>120</b>		
MODULE SCORE				
MODULE	SCORE	A2 (51–80)	B1 (81–110)	B2 (111–140)
Speaking Taken 03 September 2016	100			
Listening Taken 25 June 2016	112			
Reading Taken 25 June 2016	128			
Writing Taken 03 September 2016	140			

**Score Guide**  
The Oxford Test of English measures proficiency in English at A2, B1, and B2 levels of the Common European Framework of Reference (CEFR). The Certificate provides a CEFR level and a standardized score from 0–140. For more information on test scores go to [oxfordtestofenglish.com](http://oxfordtestofenglish.com)

**Results Verification**  
Please go to <https://verify.oxfordtestofenglish.com> for details on how to officially verify the results shown on this certificate.

*Peter Marshall*  
Peter Marshall, Managing Director, English Language Teaching, Oxford University Press

*Charles Boyle*  
Dr Charles Boyle, Deputy Director, International Programmes, University of Oxford

**OXFORD UNIVERSITY PRESS**

## 10 Results reviews and appeals

However effective a testing programme may be, test takers or other stakeholders may wish to challenge or appeal their result and transparent procedures must be open to them. There is a two-stage process for challenging a result on the Oxford Test of English: results review and appeal.

For a results review, the test results for one or more modules are checked or re-marked. For Speaking and Writing, a results review involves a re-mark of the responses. This is done by inviting senior assessors to re-mark the module in question. If the re-mark results in a score that improves the module or overall CEFR level, the results enquiry is upheld and the test taker receives a replacement result.

For Listening and Reading, the results review will involve a results check. As Listening and Reading are both marked by computer, there is no scope for re-marking as the re-mark result would be identical to the original result. However, a check is made by OUP on the tasks presented to the test taker to ensure that they received tasks at the appropriate level and that their ability estimate was correctly calculated. If an error is identified with the result, a decision will be made as to whether a revised result can be issued or whether the test taker should be given the opportunity to resit the module.

A test taker can also request an appeal via their test centre. An appeal differs from a results review in that an appeals panel, which is entirely independent of OUP, undertakes the investigation of the test taker's responses and marks to ensure that all appropriate steps have been taken in reviewing the result. The Oxford University Department for Continuing Education (OUDCE) acts as the independent appeals body for the Oxford Test of English. An appeal must be preceded by a results review.

An administrative fee is charged for all results reviews and appeals, but the fee is refunded if the review results in a change of CEFR level for either a module or the whole test, or if the appeal is upheld. All results reviews and appeals are processed on behalf of the test taker by the test centre at which the test was administered.

## 11 Test monitoring, impact and review

The development and administration steps outlined above have been designed to ensure that every administration of the Oxford Test of English provides reliable results that serve as a valid basis for decision-making.

To ensure that the Oxford Test of English continues to fulfil its stated purpose, and to seek opportunities for further improvements in quality, OUP monitors test administrations and carries out analyses of the performance of test materials, test takers and assessors at regular intervals.

Data from test administrations and feedback from assessors and stakeholders will lead to opportunities to review the test and improve its format and content in the light of experience over future years.

## 12 Acknowledgements

Oxford University Press would like to thank the following consultants for their input into and/or reviews of the design and development of the Oxford Test of English:

Professor Charles Alderson  
Dr Nathan T Carr  
Dr John Field  
Professor Anthony Green  
Professor Claudia Harsch  
Dr Alastair Pollitt  
Dr Philida Schellekens  
Dr Norman Verhelst

## 13 References

- Bachman, L. F., and Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A Manual*. Strasbourg, France: Language Policy Division.
- Dorans, N., Pommerich, M., and Holland, P. (eds.) (2010). *Linking and Aligning Scores and Scales (Statistics for Social and Behavioral Sciences)*. New York: Springer Publishing Company.
- Finnerty, C. (2015). 'The CAT is out of the bag'. *Modern English Teacher*, 24 (3): 15–17.
- Pollitt, A. (2009). *The Meaning of OOPT Scores*. <https://www.oxfordenglishtesting.com>  
Oxford: Oxford University Press.
- Warm, T.A. (1989). 'Weighted Likelihood Estimation of Ability in Item Response Theory'. *Psychometrika*, 54(3): 427–450.

## Appendix 1 – Oxford Test of English Speaking criteria

CEFR	Marks	Task fulfilment	Pronunciation and fluency	Grammar	Lexis
C1.1	<b>7</b>	<ul style="list-style-type: none"> <li>well-developed and clearly communicated message</li> <li>register is consistently appropriate to purpose of task and audience</li> </ul>	<ul style="list-style-type: none"> <li>can easily be understood with very little effort from the listener</li> <li>produces natural, intelligible spoken discourse with only occasional lapses which do not affect intelligibility or effectiveness</li> <li>can use a wide range of cohesive devices efficiently to link ideas, generally appropriately</li> </ul>	<ul style="list-style-type: none"> <li>a wide range of structures to express viewpoints clearly and fully</li> <li>consistently maintains a high degree of accuracy; errors are rare and difficult to spot</li> </ul>	<ul style="list-style-type: none"> <li>broad range of lexis to express viewpoints clearly</li> <li>consistently maintains a high degree of accuracy; errors are rare and difficult to spot</li> </ul>
B2.2	<b>6</b>	Comfortably meets the positive descriptors of 5 (B2.1)	<ul style="list-style-type: none"> <li>can generally be understood with minimal effort from the listener</li> <li>can produce stretches of speech, produced with a fairly even tempo with some hesitation, and few noticeable long pauses</li> <li>generally uses appropriate intonation, places word and sentence stress correctly and articulates individual sounds clearly</li> <li>uses a range of cohesive features to link ideas into clear, coherent, discourse, though not always appropriately</li> </ul>	<ul style="list-style-type: none"> <li>a range of structures with some complex sentence forms to express viewpoints clearly</li> <li>errors are non-impeding, and these errors can often be corrected</li> </ul>	<ul style="list-style-type: none"> <li>good range of lexis to express viewpoints clearly</li> <li>few errors which rarely impede communication</li> <li>paraphrases to cover lexical gaps, where necessary</li> </ul>
B2.1	<b>5</b>	<ul style="list-style-type: none"> <li>the test taker effectively meets the requirements of the tasks</li> <li>listener is fully informed within the context of the task</li> <li>register is generally appropriate to purpose of task and audience</li> </ul>			
B1.2	<b>4</b>	Comfortably meets the positive descriptors of 3 (B1.1)	<ul style="list-style-type: none"> <li>can generally be understood with occasional effort from the listener</li> <li>keeps going comprehensibly, though pausing is evident</li> <li>generally controls word and sentence stress, and individual phonemes</li> <li>uses simple cohesive features to link a series of short, discrete simple elements into a connected linear sequence of points</li> </ul>	<ul style="list-style-type: none"> <li>uses, with reasonable accuracy, an adequate range of structures and a repertoire of frequent 'routines' associated with more predictable situations</li> <li>attempts complex structures but errors are likely to occur</li> <li>errors are generally non-impeding</li> </ul>	<ul style="list-style-type: none"> <li>an adequate range of lexis to express viewpoints on most everyday topics</li> <li>some lexical errors/limitations occur but these rarely impede communication</li> </ul>
B1.1	<b>3</b>	<ul style="list-style-type: none"> <li>task requirements are generally fulfilled; points made are not always sufficiently expanded</li> <li>listener is adequately informed within the context of the task</li> </ul>			
A2.2	<b>2</b>	Comfortably meets the positive descriptors of 1 (A2.1)	<ul style="list-style-type: none"> <li>can generally be understood by a listener who is prepared to concentrate</li> <li>can make themselves understood in very short utterances, despite evident lengthy pauses, hesitation, and false starts</li> <li>can produce individual phonemes and word stress to a limited extent</li> <li>can link groups of words with simple, commonly-used connectors, such as and, but and because</li> </ul>	<ul style="list-style-type: none"> <li>some simple structures, often used repetitively, to express viewpoints on familiar topics</li> <li>errors are frequent and sometimes impede understanding</li> </ul>	<ul style="list-style-type: none"> <li>sufficient lexis for routine, everyday transactions involving familiar situations and topics</li> <li>errors may be frequent and impede communication</li> </ul>
A2.1	<b>1</b>	<ul style="list-style-type: none"> <li>task requirements partly covered</li> <li>ideas are not often expanded</li> <li>tasks may be misunderstood or left unresolved</li> </ul>			
N/A	<b>0</b>	Response does not fulfil all the positive descriptors of 1 (A2) or response is not sufficient (see Section 6.3 for more details)			

## Speaking irrelevant/non-response caps

	<b>No penalty</b> for irrelevant/non-responses to:	<b>Mark down one band</b> across all four criteria for irrelevant/non-responses to:	<b>Give band 0</b> across all four criteria for irrelevant/non responses to:
Script 1	<ul style="list-style-type: none"> <li>up to three of the Part 1 questions</li> <li><b>OR</b> one of the Part 2 voicemail messages</li> </ul>	<ul style="list-style-type: none"> <li>four or five of the Part 1 questions</li> <li><b>OR</b> up to three of the Part 1 questions and one Part 2 voicemail message</li> </ul>	<ul style="list-style-type: none"> <li>all of Part 1</li> <li><b>OR</b> all of Part 2</li> <li><b>OR</b> more than three Part 1 questions and one Part 2 voicemail message</li> </ul>
Script 2	<ul style="list-style-type: none"> <li>up to three of the Part 4 follow-up questions</li> </ul>	<ul style="list-style-type: none"> <li>four or five of the Part 4 follow-up questions</li> </ul>	<ul style="list-style-type: none"> <li>all of Part 3</li> <li><b>OR</b> all of Part 4</li> </ul>

## Appendix 2 – Oxford Test of English Writing criteria

CEFR	Marks	Task fulfilment	Organization	Grammar	Lexis	
C1.1	<b>7</b>	<ul style="list-style-type: none"> <li>content is well developed and clearly communicated</li> <li>register is consistently appropriate to purpose of task and audience</li> <li>overall impact on the reader is consistently positive</li> </ul>	<ul style="list-style-type: none"> <li>ideas are well organized and the text is consistently coherent and well structured</li> <li>uses a wide range of cohesive features appropriately with very rare instances of misuse or overuse</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of structures, including an appropriate proportion of complex structures</li> <li>uses structures which are appropriate and natural</li> <li>consistently maintains a high degree of grammatical accuracy and any errors are rare and non-impeding</li> </ul>	<ul style="list-style-type: none"> <li>wide range of lexis relevant to the task</li> <li>lexis is chosen to convey precise meaning; any errors are rare and non-impeding</li> </ul>	
B2.2	<b>6</b>	Comfortably meets the positive descriptors of 5 (B2.1)	<ul style="list-style-type: none"> <li>organization of ideas is generally good</li> <li>uses a reasonable range of appropriate cohesive features with only occasional misuse or overuse</li> <li>the text is generally cohesive, with only occasional breaks in communication</li> </ul>	<ul style="list-style-type: none"> <li>has a good range and command of simple language structures</li> <li>uses some complex grammatical structures, although these may be used unnaturally or inaccurately</li> </ul>	<ul style="list-style-type: none"> <li>good range of lexis relevant to the task</li> <li>choice of lexis is generally appropriate</li> <li>few errors which rarely impede communication</li> </ul>	
B2.1	<b>5</b>	<ul style="list-style-type: none"> <li>all task requirements are fulfilled and reasonably expanded where appropriate</li> <li>register is generally appropriate to purpose of task and audience</li> <li>the majority of the response is understood with minimal effort on the part of the reader</li> </ul>	<ul style="list-style-type: none"> <li>organization of ideas is adequate</li> <li>uses a variety of simple cohesive features</li> </ul>	<ul style="list-style-type: none"> <li>uses a reasonable range of simple language structures</li> <li>some grammatical errors, but these do not generally impede communication</li> </ul>	<ul style="list-style-type: none"> <li>adequate range of lexis relevant to the task</li> <li>errors do not generally impede communication and usually occur when trying to express more complex ideas</li> </ul>	
B1.2	<b>4</b>	Comfortably meets the positive descriptors of 3 (B1.1)	<ul style="list-style-type: none"> <li>task requirements are generally fulfilled but points made are not always sufficiently expanded</li> <li>the response is understood with occasional effort required on the part of the reader</li> </ul>	<ul style="list-style-type: none"> <li>evidence of a range of simple structures</li> <li>some control of simple structures</li> <li>errors impede communication at times</li> </ul>	<ul style="list-style-type: none"> <li>uses basic everyday vocabulary and expressions</li> <li>some vocabulary used inappropriately to cover gaps in repertoire</li> <li>errors impede communication at times</li> </ul>	
B1.1	<b>3</b>	<ul style="list-style-type: none"> <li>task requirements are generally fulfilled</li> <li>points made are not always sufficiently expanded</li> <li>the response is understood with occasional effort required on the part of the reader</li> </ul>	<ul style="list-style-type: none"> <li>organization of ideas is adequate</li> <li>uses a variety of simple cohesive features</li> </ul>	<ul style="list-style-type: none"> <li>uses a range of simple structures</li> <li>some control of simple structures</li> <li>errors impede communication at times</li> </ul>	<ul style="list-style-type: none"> <li>uses basic everyday vocabulary and expressions</li> <li>some vocabulary used inappropriately to cover gaps in repertoire</li> <li>errors impede communication at times</li> </ul>	
A2.2	<b>2</b>	Comfortably meets the positive descriptors of 1 (A2.1)	<ul style="list-style-type: none"> <li>sometimes links groups of words with very simple connectors, such as 'and', 'but' and 'because'</li> </ul>	<ul style="list-style-type: none"> <li>evidence of a range of simple structures</li> <li>some control of simple structures</li> <li>errors impede communication at times</li> </ul>	<ul style="list-style-type: none"> <li>uses basic everyday vocabulary and expressions</li> <li>some vocabulary used inappropriately to cover gaps in repertoire</li> <li>errors impede communication at times</li> </ul>	
A2.1	<b>1</b>	<ul style="list-style-type: none"> <li>task requirements partly covered</li> <li>the response is understood with effort required on the part of the reader</li> </ul>	<ul style="list-style-type: none"> <li>sometimes links groups of words with very simple connectors, such as 'and', 'but' and 'because'</li> </ul>	<ul style="list-style-type: none"> <li>evidence of a range of simple structures</li> <li>some control of simple structures</li> <li>errors impede communication at times</li> </ul>	<ul style="list-style-type: none"> <li>uses basic everyday vocabulary and expressions</li> <li>some vocabulary used inappropriately to cover gaps in repertoire</li> <li>errors impede communication at times</li> </ul>	
N/A	<b>0</b>	Response does not fulfil all the positive descriptors of 1 (A2) or task not attempted or 50% or more of the response is irrelevant				
<b>Task fulfilment caps</b>						
<b>Task fulfilment caps – under length</b>						
<b>Task 1 (email)</b>		<b>Task 2 (essay, article, or review)</b>	<b>Cap</b>	<b>Task 1 Cap</b>	<b>Task 2 Cap</b>	
51–70 words		71–90 words	Band 3	Band 4	Band 4	
50 words or under		70 words or under	Band 2	Incorrectly responds to or does not address one or more prompts		

## Appendix 3 – Sample responses and marking commentaries

### 1 Introduction

Below are examples of test taker responses at different CEFR levels for Speaking and Writing scripts, followed by an explanation of the marks awarded. See Section 8.2 for further information.

### 2 Speaking responses

See Appendix 1 for Speaking marking criteria.

#### 2.1 Speaking: Example

This is an example of a Speaking script 1 (Speaking Parts 1 and 2) response that was marked at A2 level.

Example 1

### Interview

**You are going to answer eight questions.**

1. I'm going to ask you some questions about music. Tell me about the type of music you enjoy most.
2. When and where do you listen to music?
3. Which musical instrument would you like to learn to play, and why?
4. Thinking about things you like, tell me about a book or film you like.
5. Can you describe a person you like?
6. What kind of work would you like to do in the future?

### Voicemail 1

**You are going to leave a voicemail message.**

You want to apply for a part-time job working in a restaurant. You would like some information about the job. Leave a voicemail message for the restaurant and:

- say why you want this job
- ask some questions about the job
- explain which days and times you can work.

### Voicemail 2

**You are going to leave a voicemail message.**

Listen to the message from your friend Mike about a shopping trip. Then, leave a voicemail message for your friend. In your message, you should:

- say how you feel about the situation
- ask some questions about his weekend
- make another suggestion for meeting.

*Hi, it's Mike here. You know we planned to go shopping together next weekend? Well, I have to cancel – sorry. I'm going away with my parents from Friday till Sunday. Really sorry – another time maybe?*

Question 1: I'm going to ask you some questions about music. Tell me about the type of music you enjoy most.

Response: *[no response]*

Question 2: When and where do you listen to music?

Response: *I don't listen to music.*

Question 3: Which musical instrument would you like to learn to play, and why?

Response: *I don't like to use any music instruments.*

Question 4: Thinking about things you like, tell me about a book or film you like.

Response: *Uh ... I like Arabic books. I read ... uh ... some Arabic books. Their name how to be ... uh ... in the future. I saw ... uh ...*

Question 5: Can you describe a person you like?

Response: *I like my ... uh ... King. He is ... uh ... very, very, very popular, and friendly and lovely.*

Question 6: What kind of work would you like to do in the future?

Response: *I would like to be a doctor.*

**Voicemail 1 question: You are going to leave a voicemail message.**

You want to apply for a part-time job working in a restaurant. You would like some information about the job. Leave a voicemail message for the restaurant and:

- say why you want this job
- ask some questions about the job
- explain which days and times you can work.

Voicemail 1 response: *Uh ... this job for... uh ... improve my language. I want know where is the job place of the restaurant. I want ... uh ... job ... I want work in ... weekend, Saturday and Sunday.*

**Voicemail 2 question: You are going to leave a voicemail message.**

Listen to the message from your friend Mike about a shopping trip. Then, leave a voicemail message for your friend. In your message, you should:

- say how you feel about the situation
- ask some questions about his weekend
- make another suggestion for meeting.

*Hi, it's Mike here. You know we planned to go shopping together next weekend? Well, I have to cancel – sorry. I'm going away with my parents from Friday till Sunday. Really sorry – another time maybe?*

Voicemail 2 response: *I am very sad because you have uh... some uh... you are going to ... with your ... I hope to you ... get a nice weekend ... maybe in the next weekend we are going to the... shopping trip.*

## Marks and commentary

**Task fulfilment | 1**

The test taker partly covers the task requirements and meets the Band 1 criteria at an entry level. Although ideas are not expanded, the test taker generally understands the questions.

**Pronunciation & Fluency | 1**

It is possible to generally understand the speaker. There is evidence of some linking with connectors such as *as*, *and* and *because*. The test taker's general control of word stress and phonemes is minimally met for Band 1 and there are impeding errors.

**Grammar | 1**

The test taker has a reasonable control of simple structures (present simple, going to), although errors sometimes impede understanding as he attempts to produce longer utterances, for example, *this job for improve my language*, which restrict him to a Band 1.

**Lexis | 1**

The test taker has sufficient lexis for talking about everyday activities and activities, such as *shopping trip*, *instrument* and *Arabic books*.

## 2.2 Speaking: Example

This is an example of a Speaking script 1 (Speaking Parts 1 and 2) response that was marked at B1 level.

Example 2

### Interview

**You are going to answer eight questions.**

1. I'm going to ask you some questions about music. Tell me about the type of music you enjoy most.
2. When and where do you listen to music?
3. Which musical instrument would you like to learn to play, and why?
4. Thinking about things you like, tell me about a book or film you like.
5. Can you describe a person you like?
6. What kind of work would you like to do in the future?

### Voicemail 1

**You are going to leave a voicemail message.**

You want to apply for a part-time job working in a restaurant. You would like some information about the job. Leave a voicemail message for the restaurant and:

- say why you want this job
- ask some questions about the job
- explain which days and times you can work.

### Voicemail 2

**You are going to leave a voicemail message.**

Listen to the message from your friend Mike about a shopping trip. Then, leave a voicemail message for your friend. In your message, you should:

- say how you feel about the situation
- ask some questions about his weekend
- make another suggestion for meeting.

*Hi, it's Mike here. You know we planned to go shopping together next weekend? Well, I have to cancel – sorry. I'm going away with my parents from Friday till Sunday. Really sorry – another time maybe?*

Question 1: I'm going to ask you some questions about music. Tell me about the type of music you enjoy most.

Response: *Uh... really I enjoy in hip-hop music. Uh... this music... uh... make me excited and I like it so much.*

Question 2: When and where do you listen to music?

Response: *Uh... I listen... uh... to music... uh... if I have free times or... eh I like to listen to music sometimes in afternoon, after school.*

Question 3: Which musical instrument would you like to learn to play, and why?

Response: *Uh... I like I want to... uh... learn the piano, uh... because... uh... my uncle, he is know to... uh... playing piano, it is very good.*

Question 4: Thinking about things you like, tell me about a book or film you like.

Response: *I like... uh... football, and... uh... some exercise like... uh... exercise, and like this.*

Question 5: Can you describe a person you like?

Response: *Um OK I describe my... uh... president, the King of Saudi Arabia. I think he's... uh... honest man and... uh... a very good person.*

Question 6: What kind of work would you like to do in the future?

Response: *Uh... I want to be a lawyer, ... uh... same as my bigger brother, he's lawyer, and he is have own office... uh... to lawyer.*

### Voicemail 1 question: You are going to leave a voicemail message.

You want to apply for a part-time job working in a restaurant. You would like some information about the job. Leave a voicemail message for the restaurant and:

- say why you want this job
- ask some questions about the job
- explain which days and times you can work.

Voicemail 1 response: *This... uh... job... uh... to... uh... study in university and... uh... I am really need this job. Uh... I want to ask you about the salary and... uh... what's [unintelligible] what's the times should I... uh... work in... uh... in this restaurant, and... uh... what will be my position in the restaurant. Uh... and I am got a few... or like this.*

### **Voicemail 2 question: You are going to leave a voicemail message.**

Listen to the message from your friend Mike about a shopping trip. Then, leave a voicemail message for your friend. In your message, you should:

- say how you feel about the situation
- ask some questions about his weekend
- make another suggestion for meeting.

*Hi, it's Mike here. You know we planned to go shopping together next weekend? Well, I have to cancel – sorry. I'm going away with my parents from Friday till Sunday. Really sorry – another time maybe?*

*Voicemail 2 response: Um ... really I'm ... uh... very disappointed because ... uh ... I was ... uh... make ... uh... some ideas for this weekend. And ... uh ... where you will go to with your parents this weekend? Uh... I ... uh ... this ... uh ... I just free in this weekend, the next weekend I can't because I will ... uh ... go ... uh ... with my ... uh ... family to another ... uh ... city and I'm really disappointed, OK maybe in the next few days I will tell you.*

## **Marks and commentary**

### **Task fulfilment | 3**

The test taker generally answers the questions and attempts all parts of the task. He doesn't expand on his answers at every opportunity, but does use what language he has to answer the tasks as effectively as possible.

### **Pronunciation & Fluency | 3**

The test taker can be understood, although some effort is required. Sentence stress is a little strained at times and he does make occasional impeding errors, but their control of word stress and range of cohesive features (e.g. *and*, *because* and *this music ...*) is enough to warrant a Band 3.

### **Grammar | 3**

The test taker controls simple tenses effectively to convey his thoughts on routines and everyday situations. The test taker does not use any complex structures throughout the script but uses basic tenses with good control (e.g. *I want to be a lawyer ... , ... in the next few days I will tell you.*).

### **Lexis | 3**

The test taker has an adequate range of vocabulary to answer all tasks. They include vocabulary such as *lawyer* and *salary*, but some lexical errors do occur, such as, *I am really disappointing*, which impedes communication.

## 2.3 Speaking: Example

This is an example of a Speaking script 2 (Speaking Parts 3 and 4) response that was marked at B2 level.

**Example 3**

### Talk

**You are going to give a talk.**

You are studying at a language school in England. You are going to give a talk to your English class about different ways of making friends with English people. Choose **two** photographs. Tell your class about the advantages and disadvantages of these two ways of making friends.



Using the internet



Going to cafes



Joining sports clubs



Dancing

### Follow-up questions

**You are going to answer six questions.**

1. Your talk was about making friends. How did you meet your best friend?
2. How important is it for friends to have the same interests?
3. Some people say you don't need to have a lot of friends. Do you agree?
4. How has technology made it easier to stay in contact with friends?
5. If you had a problem, would you prefer to talk to a friend or to your family?
6. When studying, is it better to live with your friends or family? Which do you think is better?

Question: You are going to give a talk. You are studying at a language school in England. You are going to give a talk to your English class about different ways of making friends with English people. Choose two photographs. Tell your class about the advantages and disadvantages of these two ways of making friends.

Response: *Good morning everyone, I would like to talk to you about making friends in England. Ah ... mm ... It is useful to take into account pros and cons of doing friends in England, especially if you are a foreigner. Ah, so the first thing is using the Internet is a very common way of making friends, or trying at least of making friends, but is not really direct or highly recommended because you don't have face to face contact. The other way is joining sports clubs ... mm ... from my own experience I can tell you that this is the best way to make friends because you have the opportunity to meet a lot of people to ah ... without taking into account ah ... your worries or your fears because you are already having direct contact, so I ... I will say this is the best way.*

Question 1: Your talk was about making friends. How did you meet your best friend?

Response: *I meet ... I met my first friend ah ... in my elementary school. I was ... yeah ... well, actually it was in my kindergarten, and I was like five years old, maybe six ... ah ... his name was Emilio, ... ah ... he was also from Colombia, but ... ah ... his parents were also from Italy for to ...*

Question 2: How important is it for friends to have the same interests?

Response: *It is important to have the same interests, because ... uh ... does ... you can ... yeah, like planned weekends together and to have fun together without being afraid of disappointing tell the other person because ... they don't have the opportunity to enjoy the time with you. So ... although that is important you can have fun with other people...*

- Question 3: Some people say you can have too many friends. Do you agree?  
 Response: *No, I strongly disagree with this statement, because I think that ... ah ... we can only have maybe two or three very good friends, and then the rest could be known people ... very well-known people, but no really friends ... mm ... I meant, you only opened your heart ... uh ... to a few people.*
- Question 4: How has technology made it easier to stay in contact with friends?  
 Response: *Yeah, that's right. I mean, nowadays is easier to keep contact. Even face-to-face through FaceTime, for example ... ah ... thanks to new technologies. But ... uh ... I don't think this will be a very ... very nice way of keeping contact with your friends, I mean if you are truly friends you will prefer to have direct contact and ...*
- Question 5: If you had a problem, would you prefer to talk to a friend or to your family?  
 Response: *I am used to speaking to my parents when I have a problem, but sometimes ... er ... once I have told them about my worries or problems, ah ... yes, I can speak to one of my friends, my very close friends.*
- Question 6: When studying, is it better to live with your friends or family? Which do you think is better?  
 Response: *I have had both experiences, I mean I have living ... I had been living with my parents during my first degree, and then I moved to Germany to study another first degree, and I was living in a ... residence with other students. I think that is the best ... is live with your parents, but that's my own opinion.*

## Marks and commentary

### Task fulfilment | 5

The test taker effectively meets the task requirements and generally adapts her register when necessary. The test taker minimally meets the Band 5 criteria.

### Pronunciation & Fluency | 6

Very little effort is required to comprehend what the test taker is saying. The test taker uses a range of cohesive features and responds with stretches of speech effortlessly. They occasionally make minor slips with individual phonemes.

### Grammar | 6

The test taker effectively uses a range of structures which are occasionally complex. These include relative clauses and a good control of adverbial phrases (e.g. *actually, it was in my kindergarten* and *even face-to-face through Face Time*).

### Lexis | 6

The test taker uses a good range of lexis to respond to the tasks, such as, *foreigner, face to face contact, direct contact* and the odd colloquial phrase, such as, *take into account*.

### 3 Writing responses

See Appendix 1 for Writing marking criteria.

#### 3.1 Writing: Example

This is an example of a Writing script 1 (Writing Part 1) response that was marked at A2 level.

The screenshot shows a writing task interface. On the left, a prompt box contains the following text: "You have 20 minutes to write an email. Write 80–130 words. You have an Australian friend called Sam, who is going to visit your country. First, read the email below from Sam about the visit. Then write an email to Sam, including the three notes you have made." Below the prompt is a text area containing the student's response. On the right, a preview box shows the email being written, with the subject line "My visit" and the body text: "Hi Sam, I write an your email. I've really enjoy to visit my country for a two -week next month with your friend Alex. It is a good idea. Congratulation! Tell me about Alex. When your job? You do like go to the restorants and museums? Nice to meet you Alex." At the bottom of the preview box, it says "Words: 50".

#### Marks and commentary

##### Task fulfilment | 1

The test taker has only partly covered the task requirements, and the response needs quite a lot of effort from the reader to understand what the test taker is trying to say.

##### Organization | 1

The organization of ideas is sufficient and the test taker uses an appropriate salutation for an informal email. However, the response lacks the use of very simple cohesive features, restricting their score to a Band 1.

##### Grammar | 1

The response shows the test taker's ability to use simple structures. Errors do impede communication at times, due to a lack of grammatical control, such as *I write an your email* and *you do like to go to the restorants and museums?*

##### Lexis | 1

The test taker demonstrates that they have enough everyday vocabulary to deal with topic of the task. The test taker attempts to use appropriate phrases for communicating in an email, but lacks the control and awareness to communicate appropriately, such as *Nice to meet you Alex.*, at the end of a written email.

## 3.2 Writing: Example

This is an example of a Writing script 1 (Writing Part 1) response that was marked at B1 level.

The screenshot shows a writing task interface. On the left, a prompt box contains the following text: "You have 20 minutes to write an email. Write 80–130 words. You have an Australian friend called Sam, who is going to visit your country. First, read the email below from Sam about the visit. Then write an email to Sam, including the three notes you have made." Below the prompt is a text area containing the student's response. On the right, a preview window shows the email being written, with a "Words: 101" counter at the bottom.

**You have 20 minutes to write an email. Write 80–130 words.**  
You have an Australian friend called Sam, who is going to visit your country. First, read the email below from Sam about the visit. Then write an email to Sam, including the three notes you have made.

**Subject:** My visit  
**From:** Sam  
Hi,  
I've finally decided to come to your country for a two-week holiday next month.  
*Great!*  
I'm coming with my friend Alex. Could you suggest some interesting things for us to do and see while we're there?  
*Yes, give details...*  
Also, it would be nice to meet up with you. *Will that be possible?*  
*No, because...*  
Write back soon.  
Sam

To: Sam  
Subject: My visit

Hi Sam, How nice you decided to come to my country at least. It's great. I'm very happy that you can see my beautiful country. I think you have enough time to do what you want. You should go to see our Black sea because it's exciting. There are a lot of nice hotels where you stay. Also, there are good restaurants which suggest traditional bulgarian meals. When you want to have fun you can go to disko clubs. They work all night. I'm sorry but I can't meet with you I hope this informacin is useful for you. Write back soon.

Words: 101

## Marks and commentary

### Task fulfilment | 3

Task requirements are generally fulfilled. The test taker has expanded well on prompt 1 (*Great!*), expressing excitement over Sam's decision. Prompt 2 (*Yes, give details ...*) has been addressed reasonably well by suggesting things to do (*there are lots of nice hotels*). Prompt 3 (*No, because ...*) has been addressed, although not expanded on.

### Organization | 3

The organization of ideas is generally good. Sentences are quite short though and the response only minimally uses simple cohesive features, therefore, not obtaining a Band 4.

### Grammar | 3

The test taker doesn't make any attempt at using more complex structures, although there is generally a good level of control of simple structures. Most grammatical errors do not impede communication.

### Lexis | 3

The test taker has an adequate range of lexis. Errors are present, but do not generally impede communication, e.g. *informacin* (for information) and *disko* (for disco).

### 3.3 Writing: Example

This is an example of a Writing script 2 (Writing Part 2) response that was marked at B2 level.

The screenshot shows a writing task interface. On the left, there are instructions: "You have 25 minutes to write an essay. Write 100–160 words. You have had a class discussion on clothes. Your teacher now wants you to write an essay. The title of the essay is: Do young people spend too much money on clothes? Write your essay." On the right, there is a text box containing the student's response. At the bottom right of the text box, it says "Words: 155".

**You have 25 minutes to write an essay. Write 100–160 words.**  
You have had a class discussion on clothes. Your teacher now wants you to write an essay.  
The title of the essay is:  
*Do young people spend too much money on clothes?*  
Write your **essay**.

**Do Young People Spend too much Money on Clothes?** Nowadays, no matter it is on the street or on the internet, people can see many clothes shops in variety. However, do people really need that many clothes? Lots of the young people tend to make themselves look better to catch up with the fashion, so they would like to spend money on clothes. Especially for girls, they have more desire to fill up their clothing case. Take my roommates for example, if they have nothing to do, they may go online shopping. Although they already have enough clothes, they still want to buy new clothes because the process makes them feel satisfied. Do young people spend too much money on clothes? Yes, I think, because I have seen a lot of people feel regretful immediately after they buy the clothes, or even wear the clothes only one time. Maybe young people should consider more carefully on using money.

Words: 155

### Marks and commentary

#### Task fulfilment | 5

The test taker's response is fulfilling the task with reasonable detail and minimal effort is required from the reader to understand the text. The register and format are generally appropriate.

#### Organization | 5

The organization of ideas is consistently coherent and well structured. The test taker uses a reasonable range of appropriate cohesive features, e.g. *however, nowadays, although* and pronoun referencing.

#### Grammar | 5

The test taker uses a good range of language, with attempts at more complex sentences. They occasionally make slips when using more complex structures, but overall there is a good attempt at varying grammatical structures, by using a range of tense and aspect, i.e. *I have seen a lot of people, ... if they have nothing to do, they may ... , lots of young people make themselves look better ...*

#### Lexis | 5

There is a good range of vocabulary, including a range of phrasal verbs, such as *catch up* and *fill up*. Their lexis is generally error-free and is used appropriately.

**OXFORD**

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP, United Kingdom

Oxford University Press is a department of the University of Oxford.  
It furthers the University's objective of excellence in research, scholarship,  
and education by publishing worldwide. Oxford is a registered trade  
mark of Oxford University Press in the UK and in certain other countries

© Oxford University Press 2020

The moral rights of the author have been asserted

First published in 2013

2024 2023 2022 2021 2020

10 9 8 7 6 5 4 3 2 1

All rights reserved. No part of this publication may be reproduced, stored  
in a retrieval system, or transmitted, in any form or by any means, without  
the prior permission in writing of Oxford University Press, or as expressly  
permitted by law, by licence or under terms agreed with the appropriate  
reprographics rights organization. Enquiries concerning reproduction outside  
the scope of the above should be sent to the ELT Rights Department, Oxford  
University Press, at the address above

You must not circulate this work in any other form and you must impose  
this same condition on any acquirer

Links to third party websites are provided by Oxford in good faith and for  
information only. Oxford disclaims any responsibility for the materials  
contained in any third party website referenced in this work

### **Photocopying**

The Publisher grants permission for the photocopying of this document  
Under no circumstances may any part of this document be photocopied  
for resale

### **ACKNOWLEDGEMENTS**

Oxford University Press (Using the internet/Chris King/OUP), (Going to  
cafes/Dasha Petrenko/Shutterstock), (Joining sports clubs/bikeriderlondon/  
Shutterstock), (Dancing/Iakov Filimonov/Shutterstock), (Radio/Dorling  
Kindersley Ltd/Alamy Stock Photo), (Internet/Andrew Paterson/Alamy Stock  
Photo), (Newspaper/Chris Pancewicz/Alamy Stock Photo), (TV/Mark  
Richardson/Alamy Stock Photo).