Oxford Test of English
Test specifications

Test development and validation March 2019

OXFORD
UNIVERSITY PRESS

Contents

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

Most educational institutions need a valid and reliable means of assessing students at key stages of their language development – especially in relation to the widely understood levels of the Council of Europe's Common European Framework of Reference for Languages (CEFR). The Oxford Test of English was developed to meet this need for learners of English studying on courses in a wide range of institutions, such as language schools, colleges and universities or company language training programmes. The test content is designed to be suitable for students aged 16 and above.

The starting point for the development of any new test is the perceived needs of the prospective stakeholders, for example the learners, their teachers, institutions and other involved parties, such as educational bodies and employers. Bachman and Palmer, in *Language Assessment in Practice* (Bachman and Palmer, 2010), stress the need to identify and describe the benefits a test can bring to the learners and other key stakeholders. With this in mind, the Oxford Test of English was designed to meet both institutional and individual needs. Many institutions require information on their students' language proficiency, especially at the end of their courses. They need to know whether students are ready to move on to follow higher-level language courses, or pursue further studies or activities that require a specific level of English proficiency. The test also serves the individual learner's need for external verification of their language proficiency for study or career progression.

The Oxford Test of English has been designed to measure language proficiency at CEFR levels B2, B1 and A2. 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. It is therefore ideally suited for measuring students' general proficiency in English at key points in their learning programmes.

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.

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.
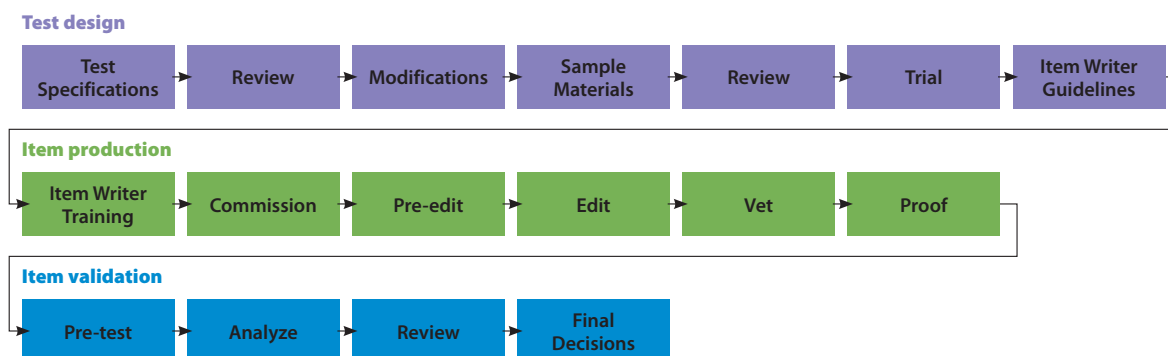
The audits represent a continuous process aimed at maintaining and improving the quality of the Oxford Test of English. They involve scrutiny of the different stages of design, production, and administration. The process continues beyond the launch of the test and includes regular reviews of test administrations 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

Test design

Test Specifications	Review	Modifications	Sample Materials	Review	Trial	Item Writer Guidelines

Item production

Item Writer Training	Commission	Pre-edit	Edit	Vet	Proof

Item validation

Pre-test	Analyze	Review	Final Decisions

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 1 shows an overview of the Oxford Test of English.

Table 1: Oxford Test of English overview

Module	Part	No. tasks	No. items	Structure	Timing
Speaking	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	
Listening	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	
Reading	Part 1	6	6	Six short texts from a variety of sources, each with one question	35 minutes
	Part 2	1	6	Six texts, profiling people, are matched to four descriptions	
	Part 3	1	6	Six extracted sentences are inserted into a longer text	
	Part 4	1	4	A longer text with four questions	
Writing	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 2 shows a summary chart of the tasks in the Speaking module.

Table 2: Overview of the Speaking module

Part	No. tasks	No. items	Structure	Testing focus
Part 1	2	6 (+ 2 unassessed)	**Interview** 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) **Audience:** the audience is the interviewer/assessor **Preparation time:** none **Response time:** Questions 1 and 2: 10 seconds per question Questions 3–8: 20 seconds per question	• responding to questions • giving factual information • expressing personal opinions on everyday topics
Part 2	2	2	**Voicemail message** Leaving two voicemail messages **Voicemail 1:** test taker leaves a voicemail Audio-visual input consisting of a situation with three prompts requiring the test taker to leave a voicemail **Audience:** 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) **Preparation time:** 20 seconds **Response time:** 40 seconds **Voicemail 2:** test taker replies to a voicemail 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 **Audience:** the audience is specified in the task, and the relation to that audience is informal (e.g. friend) **Preparation time:** 20 seconds **Response time:** 40 seconds	• organizing and sustaining extended discourse • sociolinguistic appropriacy • sustaining relationships
Part 3	1	1	**Talk** 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 **Audience:** the audience is specified and is typically the test taker's classmates **Preparation time:** 30 seconds **Response time:** 1 minute	• organizing and sustaining extended discourse • describing • comparing and contrasting • speculating • suggesting

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

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 3 shows a summary chart of the tasks in the Listening module.

Table 3: Overview of the Listening module

Part	No. tasks	No. items	Structure	Testing focus
Part 1	5	5	Multiple choice – picture options	Listening to identify:
			Five discrete short monologues/dialogues with picture options	• specific information
			Five three-option multiple-choice questions	
			Time to check answers: 10 seconds	
			Audioscript length: A2 = 30–65 words; B1 = 55–85 words, B2 = 70–96 words.	
Part 2	1	5	Note completion	Listening to identify:
			A longer monologue with a note-completion task	• specific information
			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.	
Part 3	1	5	Matching opinions with people who say them	Listening to identify:
			A longer dialogue with a task focusing on identifying opinions	• stated opinion
			Five three-option multiple-choice questions	• implied meaning
			Time to check answers: 15 seconds	
			Audioscript length: A2 = 200–300 words; B1 = 300–400 words, B2 = 400–525 words.	
Part 4	5	5	Multiple choice	Listening to identify:
			Five discrete short monologues/dialogues with text options	• attitude/feeling/opinion
			Five three-option multiple-choice questions	• gist
			Time to check answers: 10 seconds	• function/reason/purpose
			Audioscript length: A2 = 30–65 words; B1 = 55–85 words, B2 = 70–96 words.	• speaker relationship
				• topic
				• type/genre

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 4 shows a summary chart of the tasks in the Reading module.

Table 4: Overview of the Reading module

Part	No. tasks	No. items	Structure	Testing focus
Part 1	6	6	Multiple-choice questions on short texts Six short texts from a variety of sources including: adverts, blogs, emails, notes, notices and text messages Six discrete three-option multiple-choice questions Time to process the texts and complete the tasks: 1 minute 20 seconds per task (8 minutes in total) Text length: A2 = 20–35; B1 = 20–50; B2 = 40–70.	Reading to identify: • main message • purpose • detail
Part 2	1	6	Multiple matching Matching six profiles of people with requirements (e.g. requirements for a particular type of holiday) to four descriptions (e.g. of four different kinds of holiday) Texts from brochures, advertisements, magazine articles Six multiple-matching questions Time to process the texts and complete the task: 8 minutes Text length for each description: A2 = 45–60 word; B1 = 80–100 words; B2 = 100–125 words.	Expeditious reading to identify: • specific information • opinion and attitude
Part 3	1	6	Gapped text Six extracted sentences are inserted into a longer text Texts are from newspaper and magazine articles Six text-completion questions Time to process the text and complete the task: 11 minutes Text length: A2 = 200–220; B1 = 350–375; B2 = 400–425.	Reading to identify: • text structure • organizational features of a text
Part 4	1	4	Multiple-choice questions on longer texts Four three-option multiple-choice questions Texts are from newspaper and magazine articles Time to process the text and complete the task: 8 minutes Text length: A2 = approx. 235; B1 = approx. 350; B2 = approx. 350.	Reading to identify: • attitude/opinion • purpose • reference • the meanings of words in context • global meaning

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 5 shows a summary chart of the tasks in the Writing module.

Table 5: Overview of the Writing module

Part	No. tasks	No. items	Structure	Testing focus
Part 1	1	1	**Email** 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	• giving information • expressing and responding to opinions and feelings • transactional functions such as inviting/requesting/suggesting
Part 2	1	1	A choice of writing tasks: an essay or a magazine article/review	
			Essay 100–160 words Writing an essay on a topic typical of classroom discussions Time to process the task and complete the response: 25 minutes	• expressing and responding to opinions • developing an argument
			or **Magazine article/Review** 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	• describing • narrating • expressing feelings and opinions • recommending

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 unique 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 fourteen 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[1]:

- 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 Pronunciation, Fluency, Grammar, and Lexis. The table below summarizes the main elements of the marking criteria. See Appendix 1 for detailed Speaking marking criteria.

[1]For Spain, these accommodations are being rolled out in 2019.

Table 6: Elements of the Speaking criteria

Pronunciation	Fluency	Grammar	Lexis
• Phonological and word stress precision • Phonological linking • Intonation, rhythm, and stress	• Coherence • Cohesion • Register	• Range of structures • Accuracy of structures	• Range of lexis • Accuracy of lexis

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

Table 7: Elements of the Writing criteria

Task fulfilment	Organization	Grammar	Lexis
• Fulfilling task requirements • Format • Register • Length	• Coherence • Cohesion	• Range of structures • Accuracy of structures	• Range of lexis • Accuracy of lexis

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. These represent the 'criterion level', or standard for that CEFR level. For bands 2, 4, and 6, there are referential descriptors which refer to the bands below and above. Bands 2, 4, and 6 are 'plus levels' (A2.2, B1.2, B2.2) which represent a level of proficiency which is significantly higher than that represented by the criterion level, but which does not achieve the standard for the level above. 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 the table below, 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 8: 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 6 shows the relationship between the Oxford Test of English scale and the CEFR.

Table 9: 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

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

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	Pronunciation	Fluency	Grammar	Lexis
C1.1	7	• can easily be understood with very little effort • phonemes and word stress are clearly produced • phonological linking between words and phrases achieves effective communication • demonstrates natural use of pausing between and within sentences to express meaning clearly • varies intonation, rhythm and sentence stress effectively	• can communicate spontaneously in longer complex stretches of speech to achieve desired purpose • can use reformulation, suitable 'fillers' or hedging when necessary, with the ability to react appropriately with no noticeable long pauses • can use a wide range of cohesive devices efficiently to link ideas, generally appropriately • register is consistently appropriate to the purpose of the task and audience	• a wide range of structures to express viewpoints clearly and fully • consistent grammatical control of a variety of forms • errors are rare, difficult to spot and generally corrected when they do occur	• a wide range of vocabulary used effectively to express viewpoints on a broad range of topics • occasional non-impeding errors when attempting more complex language • able to paraphrase, re-formulate or backtrack where necessary • uses situationally appropriate language to complete required task, with good command of common idiomatic language
B2.2	6	Fulfils all of the positive descriptors of 5 (B2) and some of the descriptors in 7 (C1)			
B2.1	5	• can be understood with a little effort • can produce individual phonemes and word stress; errors are non-impeding • can segment and phonologically link utterances to express meaning clearly; errors are non-impeding • can use intonation, rhythm and sentence stress to express meanings such as emphasis or contrast	• can produce stretches of speech, produced with a fairly even tempo with some hesitation, and few noticeable long pauses • can use a range of cohesive devices to link ideas into clear, coherent discourse, though not always appropriately • register is mostly appropriate to the purpose of the task and audience	• a range of structures with some complex sentence forms to express viewpoints clearly • errors are non-impeding, mainly when attempting more complex language; these errors can often be corrected	• a range of vocabulary to express viewpoints on most general topics • some lexical errors; these are mainly when attempting more complex language and do not impede communication • can use paraphrase to cover lexical gaps • generally uses situationally appropriate language
B1.2	4	Fulfils all of the positive descriptors of 3 (B1) and some of the descriptors in 5 (B2)			
B1.1	3	• can generally be understood with occasional effort from the listener. • can generally produce individual phonemes and word stress, but errors may be intrusive • can generally segment and phonologically link utterances; errors are generally non-impeding • can generally control intonation, rhythm and sentence stress, but errors may be intrusive	• can keep going comprehensibly, though pausing is evident • can use simple cohesive devices to link a series of short, discrete simple elements into a connected linear sequence of points • register is generally appropriate to the purpose of the task	• an adequate range of structures and a repertoire of frequent 'routines' associated with more predictable situations to express basic viewpoints • errors do not generally impede communication; it is clear what he/she is trying to express	• an adequate range of vocabulary to express viewpoints on most everyday topics • some lexical errors/limitations occur but these do not impede communication, but may cause repetition and even difficulty with formulation at times • some attempt at using paraphrase to cover lexical gaps
A2.2	2	Fulfils all of the positive descriptors of 1 (A2) and some of the descriptors in 3 (B1)			
A2.1	1	• can generally be understood by a sympathetic listener who is prepared to concentrate • can produce individual phonemes and word stress to a limited extent; frequent errors may affect intelligibility • can segment and phonologically link to a limited extent; frequent errors may affect intelligibility • can control intonation, rhythm and sentence stress to a limited extent; frequent errors may affect intelligibility	• can make him/herself understood in very short utterances, despite evident lengthy pauses, hesitation, and false starts • can link groups of words with simple, commonly-used connectors, such as *and, but* and *because* • there is some evidence of the ability to adapt speech to audience in terms of register	• some simple structures to express basic viewpoints • basic systematic errors may be frequent and may affect intelligibility	• sufficient vocabulary, with memorized phrases, groups of a few words and formulae, to conduct routine, everyday transactions involving familiar situations and topics • can adapt well-rehearsed memorized simple phrases to particular circumstances through limited lexical substitution • errors may be frequent and impede communication • limited ability to paraphrase
N/A	0	Response does not fulfil all the positive descriptors of 1 (A2) or tasks not attempted or 50% or more of the response is irrelevant. See also table below.			

Speaking irrelevant/non-response caps

	No penalty for irrelevant/non-responses to:	**Mark down one band** across all four criteria for irrelevant/non-responses to:	**Give band 0** across all four criteria for irrelevant/non responses to:
Script 1	• up to three of the Part 1 questions • **OR** one of the Part 2 voicemail messages	• four or five of the Part 1 questions • **OR** up to three of the Part 1 questions and one Part 2 voicemail message	• all of Part 1 • **OR** all of Part 2 • **OR** more than three Part 1 questions and one Part 2 voicemail message
Script 2	• up to three of the Part 4 follow-up questions	• four or five of the Part 4 follow-up questions	• all of Part 3 • **OR** all of Part 4

Appendix 2 – Oxford Test of English Writing criteria

CEFR	Marks	Task fulfilment	Organization	Grammar	Lexis
C1.1	7	• all task requirements are fulfilled, content is relevant, and fully expanded where appropriate • format is consistently appropriate • register is consistently appropriate to purpose of task and audience	• organization of ideas is consistently coherent and well-structured • uses a wide range of cohesive devices appropriately with very rare instances of misuse or overuse	• uses a wide range of language including an appropriate proportion of complex structures • uses structures which are appropriate and natural • consistently maintains a high degree of grammatical accuracy and any errors are rare and non-impeding	• wide range of vocabulary • good command of idiomatic expressions, situationally appropriate lexis and phrases, and colloquialisms • occasional minor slips, but any errors are rare and non-impeding
B2.2	6	Fulfils all of the positive descriptors of 5 (B2) and some of the descriptors in 7 (C1)			
B2.1	5	• all task requirements are fulfilled and reasonably expanded where appropriate • format is almost always appropriate • register is almost always appropriate to purpose of task and audience	• organization of ideas is generally good • uses a reasonable range of appropriate cohesive devices with only occasional misuse or overuse	• uses a good range of language with a reasonable number of attempts at more complex structures • few grammatical errors which rarely impede communication, and occur mainly when attempting more complex structures	• good range of vocabulary • reasonable command of idiomatic expressions and collocations which are generally appropriate • few errors which rarely impede communication
B1.2	4	Fulfils all of the positive descriptors of 3 (B1) and some of the descriptors in 5 (B2)			
B1.1	3	• task requirements are generally fulfilled; points made are not always sufficiently expanded • format is generally appropriate • register is generally appropriate to purpose of task and audience	• organization of ideas is adequate • uses simple cohesive devices	• adequate range of structures • some grammatical errors, but these do not generally impede communication	• adequate range of vocabulary • paraphrases and uses a range of lexis to avoid repetition though this may be limited • errors do not generally impede communication and usually occur when trying to express more complex ideas
A2.2	2	Fulfils all of the positive descriptors of 1 (A2) and some of the descriptors in 3 (B1)			
A2.1	1	• task requirements partly fulfilled • format may at times be appropriate • there may be some evidence of ability to adapt text to audience in terms of register.	• poorly executed organization sometimes impedes communication of message • sometimes links groups of words with very simple connectors, such as *but* and *because*	• evidence of a basic range and control of simple structures and sentence patterns • errors impede communication at times	• sufficient vocabulary to deal with simple concrete everyday needs with some repetition • errors impede communication at times
N/A	0	Response does not fulfil all the positive descriptors of 1 (A2) or task not attempted or 50% or more of the response is irrelevant			

Task fulfilment caps

Underlength caps for Task fulfilment

Task 1 (email)	Task 2 (essay, article, or review)	Cap
51–70 words	71–90 words	Band 3
50 words or under	70 words or under	Band 1

Prompt caps for Task fulfilment

Issue	Task 1 Cap (email)	Task 2 Cap (essay, article, or review)
Incorrectly responds to one or more prompts	Band 4	Band 1
Does not address one or more prompts	Band 2	Band 1

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 1

This is an example of a Speaking script 2 (Speaking Parts 3 and 4) response that was marked at A2.1 level.

Part 3 – Talk

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

Response: *This picture of using the Internet is very… uh… development style, so it is very useful, but it is not communication style. And this photo of going to cafe is very easy communication style so very fun… I think very fun so… uh… it is the advantages style.*

Part 4 – Follow-up questions

Follow-up questions

You are going to answer six questions about your talk.

Start speaking when you hear the tone.

The clock shows how much time you have to answer each question.

Listen

Speak

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

Response: *I meet best friend at high school, and she is very… uh… listen to my opinion… so very nice person.*

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

Response: *I think… uh… my friend interested in same hobbies and… so same community… so important.*

Question 3: Some people say you don't need to have a lot of friends. Do you agree?

Response: *I agree, because I'm… I can many talking with my friends so… uh… many lots of vocabularies and lots of grammars. I learn it.*

Question 4: How has technology made it easier to stay in contact with friends?

Response: *I think communication by mobile phone and email and letters so very… it is so very fast and easy, so useful communication tool.*

Question 5: If you had a problem, would you prefer to talk to a friend or to your family?

Response: *I… I think… consult my… uh… if I had the problem, I consult my family. My family is very know me.*

Question 6: Some students live with friends when they are at university, others live with family. Which do you think is better?

Response: *I think it's different situation because I think… uh… to my friends… uh… I think… to my family is… my family consult.*

Marks and commentary

Pronunciation | band 2

The test taker can generally be understood with occasional effort from the listener. Her individual sounds are fairly clear, but there are some intrusive errors, such as the /p/ and /b/ sounds in *problem* sounding more like /frɒvləm/; the /r/ sounds are also difficult to understand. Her ability to link sounds is limited, with most words produced separately. Her intonation is fairly flat, which requires more concentration on the part of the listener.

Fluency | band 1

Most utterances are fairly short, and there are lengthy pauses, hesitation and false starts. She uses some simple linking devices, but the frequency of false starts and re-phrasing means the linking is not always successful.

Grammar | band 1

The test taker uses some simple structures, but overall there is a lack of control with frequent, basic systematic errors. For example, words are frequently omitted: *I meet (my) best friend | she is (a) very nice person | my friend (is) interested in | so (it's) important*.

Lexis | band 1

Although there is an attempt to use a range of vocabulary (e.g. *development style | communication style | my opinion | consult | different situation*), there are frequent errors and she has difficulty in formulating phrases, both of which impede communication. She demonstrates limited ability to use lexical substitution and to paraphrase: e.g. *She is very listen to my opinion | My family is very know me | I think very fun so it is the advantages style*.

2.2 Speaking: Example 2

This is an example of a Speaking script 2 (Speaking Part 3 and 4) response that was marked at B1.1 level.

Part 3 – Talk
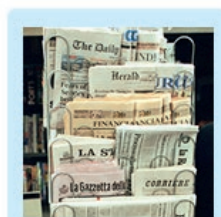
Talk

You are going to give a talk.

You are going to give a talk to your class about different ways of finding out about the news. Choose **two** photographs. Tell your class about the advantages and disadvantages of these two ways of finding out about the news.

Radio

Internet

Newspaper

TV

Response: *The first of I choose this is TV. The TV advantage that there is so much TV channel in the world, and if you want to, uh, want to know what's happen in the world, you watch TV and look that. I think the TV is not have disadvantage. The second who I choose, this is radio. I (unintelligible) the… uh, I will say that radio has disadvantages because the radio have, uh, hasn't picture. Radio has only sound. But this is no enough when you hear radio to know…*

Part 4 – Follow-up questions

Question 1:   Your talk was about the news. Tell me about what type of news stories you are interested in.

Response: *I interest about sport story because I very very like sport. When I was younger, as, I'm used to artistic gymnastic. I used to swim and now I go to ski every week in the winter when I have a free time. That, the sport news I very like.*

Question 2:   Why do you think it's important to find out about the news?

Response: *It's very important about the news because the world is very interesting. I interest … of geograph… geography, of history that I so … have interest of the news. So I like other things. I have interest of weather so …*

Question 3:   The news today has a lot of information about famous people. Is this a good thing or a bad thing?

Response: *Yes, there is about information of famous people, but I think is this is the badly because in the TV, in the news, only speak about famous people, they not speak about normal, ordinary people who go to work every day, so I think that is it's the badly.*

Question 4:   What do you think you could learn from watching television programmes from another country?

Response: *I very like to watch travel channel because there is… uh, there are look for many countries, for many difference site of the wars, for bird, and for the sea, ocean, coast, mountain, so I very like TV…*

Question 5:   There is a lot of advertising on television. How do you feel about this?

Response: *I feel… I feel well, because I look it. It's very interesting story who I made… made to watch in the TV, so I… like adventures too, so I want to say that I… I very like this emotional…*

Question 6:   Some people don't watch television at all. Why do you think this is?

Response: *I don't know. I every day watch the TV, but some people don't watch TV because they prefer read the books, so they prefer read the newspaper. I… personally I prefer to watch TV. So many people have no free time to watch TV because they work hardly. This is my opinion.*

Marks and commentary

Pronunciation | band 3

The test taker can be understood in most responses, with some effort from the listener. At word level, she can produce individual phonemes, although there are some persistent errors, e.g. *radio* /rɑːdɪəʊ/, and some that affect intelligibility, e.g. *mountain* /maʊntaɪn/.

She shows some ability to link speech, e.g. *when I was* /wenaɪwəz/ *younger* | *people who go to work* /gəʊtəwɜːk/ *every day*. There is also some control of sentence stress and intonation, e.g. *… but some people don't watch TV because they prefer read the books*. However, in several responses, a lack of sentence stress and intonation makes her meaning difficult to understand.

Fluency | band 3

The test taker produces stretches of speech with a few pauses and false starts. She uses simple cohesive devices, e.g. *and*, *but*, *because*, and attempts others e.g. *first*, *second*. However, her ideas are not always linked coherently. Register is generally appropriate.

Grammar | band 3

The test taker uses basic structures to express herself. She also attempts some higher-level language, e.g. *If you want to know what's happening in the world, you watch the TV* | *I used to swim and now I go to ski every winter*.

While there are some persistent errors, e.g. *I very like*, use of *not* instead of don't, these errors don't generally affect intelligibility.

Lexis | band 3

The test taker uses an adequate range of vocabulary to express her ideas, e.g. *… for the sea, ocean, coast …* | *I used to gymnastics, I used to swim and now I go to ski*.

There are some errors, e.g. *interested of, interested about, I made to watch in the TV, work hardly*, but these don't generally impede communication.

2.3 Speaking: Example 3

This is an example of a Speaking script 2 (Speaking Part 3 and 4) response that was marked at B2.1 level.
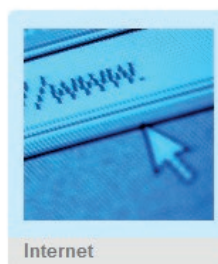
Part 3 – Talk

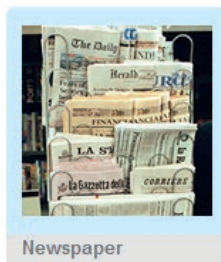### Talk

You are going to give a talk.

You are going to give a talk to your class about different ways of finding out about the news. Choose **two** photographs. Tell your class about the advantages and disadvantages of these two ways of finding out about the news.
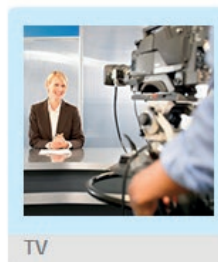
Radio

Internet

Newspaper

TV

Response: *I'm going to talk about newspaper and TV. To watch the news on TV has several advantages like you will be always on… on time and informed about the latest news in the world and in your country and you will have a picture from the events, but there the news are only at the specific times. If you have a newspaper, you will inform… inform yourself about news every time when you want. This is the advantage, and the disadvantage is that the information there is… can't be so completely… uh… full as the TV, the information…*

Part 4 – Follow-up questions

Question 1: Your talk was about the news. Tell me about what type of news stories you are interested in.

Response: *I'm interested in all types of news story. I'm very interested what happened in the world, I want all the time to be informed, that's why all the time I try to look at the news on TV or in… on… Internet, and to inform myself… That's all.*

Question 2: Why do you think it's important to find out about the news?

Response: *We need to be informed about the news. Sometimes they are good, sometimes not too good. For example, the latest news are always for the natural disaster or some kinds of accidents, and if we are informed, we can try to help to the people, like the disaster with floods in…*

Question 3: The news today has a lot of information about famous people. Is this a good thing or a bad thing?

Response: *I think that it is a good thing, because sometimes famous peoples are leaders and other people can copy their behaviour and it will be a good news if the news is really good and is a example of behaviour.*

Question 4: What do you think you could learn from watching television programmes from another country?

Response: *I very often watch television news on the other country, especially the world news because there is a different point of view on… on event. For example, the latest news for what happen in the world is… was very interesting and it's good to see… to know another point…*

Question 5: There is a lot of advertising on television. How do you feel about this?

Response: *I don't think that adver… adversment are good because they manipulate people, especially young people and children, which think that everything what is advertised is very good and they want to have all the things which are advertised without thinking of money, and there are too many adversment on TV.*

Question 6: Some people don't watch television at all. Why do you think this is?

Response: *Maybe because they inform theirselves from other… uh… other kinds of devices like radio, or they read newspapers or read the news on Internet. But there is a people which are not interested… who are not interested in reading…*

Marks and commentary

Pronunciation | band 5

The test taker can be understood fairly easily. She produces individual phonemes with a few non-impeding errors, e.g. pronouncing the /ð/ in *their* with a /d/ sound, and the /θ/ in *think* with an /f/ sound. She also sometimes extends a final /ɡ/ sound onto words beginning with vowel sounds. Her intonation is rather monotone with occasional variation to convey meaning.

Fluency | band 5

The test taker can produce stretches of speech with a fairly even tempo, and few noticeable long pauses. She can link her ideas with a range of cohesive devices, including referencing, connectors and clauses, although not always appropriately: e.g. *They manipulate people, especially young people, which think that everything what is advertised is* …. Register is mostly appropriate.

Grammar | band 5

The test taker uses a range of structures, including some complex sentence forms.

There are fairly frequent errors but these are non-impeding, e.g. *you will be always on time | the news… Sometimes they are… | Try to help to the people | a good news | I want all the time to be informed*. She is able to correct some errors, e.g. *in… on the Internet*.

Lexis | band 5

The test taker uses a range of vocabulary to express her viewpoints, e.g. *to be informed about | a different point of view | they manipulate people | devices*. There are some lexical errors which do not impede communication, e.g. *you will be on time | informed theirselves*.

Generally she uses situationally-appropriate language.

3 Writing responses

See Appendix 2 for Writing marking criteria.

3.1 Writing: Example 1

This is an example of a Writing script 1 (Writing Part 1) response that was marked at A2.2 level.

You have 20 minutes to write an email. Write 80–130 words.

There is going to be a party at your language school because it is the end of your English course. First, read the email below from the school principal, Mr Lester. Then write to the principal, including the three notes you have made.

Subject: School Party
From: Mr Lester

Dear Student

I am writing about the party we are organizing for the end of the course. Which day do you think would be best?

Explain when

We would like some students to cook some food from their country and bring it to the party. Are you able to cook something?

No because ...

We want everyone to remember this party. How can we make the party special?

Suggest how ...

Thank you

Mr Lester

School Principal

To: Mr Lester
Subject: School Party

Dear Mr Lester,
Tank you very much for your email. I am very glead about this party! I am sure that it will be a great fun. I think that Saturday would be best.
It is a good idea that everyone to can intraduce some food from their country. I will be able to cook something special to.
We sould wear aspesial party hats. It will be very funny. Also everyone wil be able to wear some spesial clothes. I am going to invite a DJ who is awsome.
Iam looking forward to hearing from you!
Best wisis,
Damyanka

Words: 94

Marks and commentary

Task fulfilment | band 2

The test taker has only partly covered the task requirements, and the response needs greater expansion. However, the format and register is generally appropriate to the purpose of the task, giving a band 2 for task fulfilment.

Organization | band 2

Organization of ideas is adequate. However, the response lacks the use of simple cohesive devices, restricting their score to a band 2 for organization.

Grammar | band 3

The response displays the test taker's ability to use an adequate range of structures. There are attempts at some more complex structures, e.g. the use of relative clauses: *I am going to invite a DJ who is awsome*. Errors do not generally impede communication.

Lexis | band 2

The test taker shows that they have sufficient vocabulary to deal with simple concrete everyday needs. There are attempts at using a variety of more descriptive language such as *fun* and *funny*, and *awesome* and *great* to avoid repetition, but there is not a sufficient range for this to be a band 3.

3.2 Writing: Example 2

This is an example of a Writing script 1 (Writing Part 1) response that was marked at B1.2 level.

You have 20 minutes to write an email. Write 80–130 words.

There is going to be a party at your language school because it is the end of your English course. First, read the email below from the school principal, Mr Lester. Then write to the principal, including the three notes you have made.

Subject: School Party
From: Mr Lester

Dear Student

I am writing about the party we are organizing for the end of the course. Which day do you think would be best?

Explain when

We would like some students to cook some food from their country and bring it to the party. Are you able to cook something?

No because ...

We want everyone to remember this party. How can we make the party special?

Suggest how ...

Thank you

Mr Lester

School Principal

To: Mr Lester
Subject: School Party

Dear Mr Lester
Thank you for email. I readed it carefully i think the last day at school will be better for the party. Because after it many students are going to go to their hometown.
Yes, I m very good cooker I m able to to cook some bulgarian dishes. This is not problem for me to prepare something for last night at our school. It is good idea to eat something different from all over the world.
I think that national music is a great idea to make our party special. I can play guitar I can play bulgarian music too. Others of my coleaques also are able to dansing nad palying well their national music. It will be great party.
Yours cincerely.

Words: 125

Marks and commentary

Task fulfilment | band 4

Task requirements are generally fulfilled. The test taker has expanded well on prompt 1 ("Explain when"), giving a clear rationale for their choice of date. Prompt 2 ("No because …") has been addressed incorrectly by giving a positive response (i.e. that the test taker can cook something), rather than giving a negative response (i.e. that they can't cook something). This means that the score for task fulfilment is capped at a maximum of band 4. Prompt 3 ("Suggest how …") has been well addressed, with appropriate examples given. Format and register are almost always appropriate.

Organization | band 4

Organization of ideas is generally good. The test taker uses simple cohesive devices such as *because* and *too*. Sentences are quite short and some linkers/punctuation marks are missing, which can place a strain on the reader, e.g. *Yes, I m very good cooker I m able to to cook some bulgarian dishes*.

Grammar | band 4

The test taker makes some attempts at more complex structures, e.g. *This is not problem for me to prepare something for last night at our school*. However, non-impeding errors do occur in the more complex structures, e.g. *Others of my coleaques also are able to dansing nad palying well their national music*.

Lexis | band 4

The test taker has a good range of vocabulary that allows them to express ideas with a certain amount of flexibility. Errors are present, but do not generally impede communication, e.g. *coleaques* (for colleagues) and *cooker* (for cook).

3.3 Writing: Example 3

This is an example of a Writing script 1 (Writing Part 1) response that was marked at B2.2 level.

You have 20 minutes to write an email. Write 80–130 words.

You recently applied to work in a hotel during the summer. First, read the email below from Mrs Wilson, the hotel manager. Then write an email to Mrs Wilson, including the three notes you have made.

Subject: Job application
From: Mrs Wilson

Thank you for your application for a summer job.

Can you come for an interview on Tuesday next week at 11.00 a.m.?

Yes - accept

As you know, there are jobs available at the hotel for receptionists and cooks. Please tell us which job you would prefer and why.

Say which job and why

If you have any questions about the job, please get in touch.

Ask some questions

Sincerely

Mrs Wilson

City Hotel Manager

To: Mrs Wilson
Subject: Job application

Dear Mrs Wilson,

Thank you for your letter. I am available on Tuesday next week so I will be in your office at 11.00 a.m. for the interview.

Regarding the jobs offered, I would definitely prefer the one as a receptionist. I studied Spanish at school and I spent last summer working as a waitress in Alicante. I also spent last summer with a family in France. Apart from this, I have done a course for attending clients. Moreover, I have no problems with night shifts or working at weekends.

I would appreciate any further information about the salaries and vacations.

I am looking forward to seeing you.

Sincerely,

Words: 111

Marks and commentary

Task fulfilment | band 6

Prompts 1 and 2 are fully expanded with good detail. Prompt 3 is minimally covered and needs further expansion, restricting this response to band 6. Register and format are consistently appropriate.

Organization | band 6

The organization of ideas is consistently coherent and well structured. The test taker uses a reasonable range of appropriate cohesive devices, e.g. *moreover*, *regarding* and *so*.

Longer sentences and more complex cohesives would be needed to award this a band 7.

Grammar | band 6

The test taker uses a good range of language, with attempts at more complex sentences, e.g. *Regarding the jobs offered, I would definitely prefer the one as a receptionist*. The test taker could have demonstrated greater range by including some questions in the third paragraph. The test taker displays a high degree of accuracy, although something to bear in mind is the number of simple sentences.

Lexis | band 5

There is a good range of vocabulary, including situationally appropriate lexis, e.g. *I would appreciate any further information* and *Regarding the jobs offered*. Lexis is generally error-free and used appropriately.

OXFORD
UNIVERSITY PRESS