

# 1. How the INTO *English Language Assessment* has been developed by Cambridge English.

## 1.1 Design

### i) Constructing tests linked to the CEFR

The INTO *English Language Assessment* has been designed to indicate where candidates are on the proficiency continuum for which results are scored and mapped to the CEFR. A guiding principle for the INTO *English Language Assessment* is for the test content and test reporting to be aligned to the Common European Framework for Reference (CEFR). Cambridge English has a long history of involvement with the development and practical application of the CEFR in its task production and test design. The test material that Cambridge English has developed for the INTO *English Language Assessment* reflects the CEFR's action-oriented, functional model of language use, while ensuring relevance for the target population. The socio-cognitive model adopted by Cambridge English is based on the CEFR's model of language use and learning, and identifies two dimensions – the social dimension of language in use, and the cognitive dimension of language as a developing set of competences, skills and knowledge. These are used to define testable abilities across the target range of proficiency levels. In order that the resulting test construct should be implemented comparably across levels, these abilities are mapped to specific task types, drawing chiefly on task types used successfully for many years in Cambridge English's operational exams.

The pre-requisite for any claims about alignment is that assessment standards need to be maintained in the long-term using a range of techniques and professional processes, including:

- item banking to establish common measurement scales and to allow for both item-based and person-based equating to be employed in the selection of test tasks and in the monitoring of standards over time;
- routine test validation processes to quality assure test outcomes;
- iterative cycles of test development and revision.

These stages of test development and test validation are explained in more detail in the following sections looking at the test specifications and test production processes.

### ii) Test Specifications

#### Test content

Test content is approached using the categories proposed by the CEFR. The CEFR identifies *domains* of language use (personal, public, educational and professional), *communication themes* (personal identification, daily life, travel, etc.), *topic-specific notions* (leisure, hobbies and interests, etc.), *general notions* (the meanings and concepts expressed through language whatever the specific situation), and *language functions* (imparting and seeking information, expressing and finding out attitudes, etc.) Together these communication themes, notions and functions provide the basis for categorising and selecting texts for use in the INTO *English Language Assessment*.

#### Unsuitable topics

It is important that test materials do not contain anything that might offend or upset participants, thereby potentially affecting their performance or distracting them during the tests. Thus, certain topics such as war, politics, serious family problems, etc, are considered unsuitable.

#### International language

English is used and learnt in a wide range of international contexts. To reflect this, test materials will feature a range of appropriate texts and accents from a wide variety of sources.

### Task types

The theoretical underpinning of the question and interaction types follows the latest research in this area, supported to a great extent by Cambridge English’s own recognised and renowned Research and Validation department.

Once the process of identifying the test content and the set of testable sub-skills or abilities to be assessed is complete, the next step is to map each ability to a specific profile of task types, and this informs task production. The computer adaptive nature of the Reading and Listening component of the INTO *English Language Assessment* means that each test experience will be different in terms of content, but the range of task types and range of task difficulties will be similar for candidates of similar ability. A rigorous test design is implemented which can be replicated across languages, thus maximising coherence and consistency in the implementation of the construct. The following task types are used:

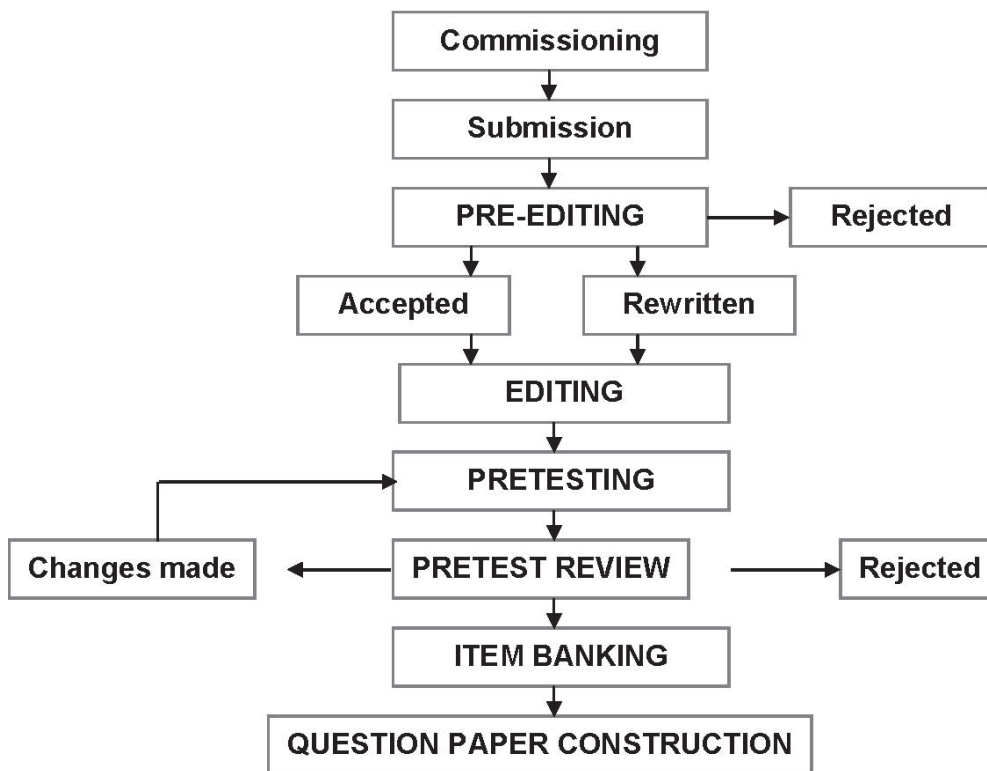
Skill	Task type and focus
<b>Reading and Language Knowledge</b>	Multiple choice. Reading to understand e.g. notices, messages, timetables, adverts, leaflets, graphs.
	Multiple choice. Reading a longer text for understanding, e.g. a newspaper or magazine article, advert, leaflet.
	Multiple choice cloze. Medium length gapped text focusing on lexis and lexico-grammar.
	Multiple choice. Reading a medium or longer text with six sentences removed to decide from which part of the text each sentence has been removed. Focussing on text cohesion, coherence and structure. B2 and above
	Multiple choice. Reading a single text or a series of short texts. Matching a series of prompts to elements of the text. Focus on reading for detail, opinion and implication. B2 and above
	Open cloze. Medium length gapped text focusing on grammar and lexico-grammar. Up to B2.
	Multiple choice. Gapped sentences focusing on grammar and vocabulary, e.g. semantic precision, collocations, fixed phrases, linking words.
<b>Listening</b>	Multiple choice. Understanding short conversations or monologues. Candidates listen and select the correct answer.
	Multiple choice. Understanding short conversations or monologues. Candidates listen and select the correct picture or graphic.
	Multiple choice. Listening to extended speech for detail and inference. One monologue and one dialogue.
<b>Writing</b>	Essay. Write at least 250 words considering various aspects of a given issue.
<b>Speaking</b>	General interaction and social language, giving personal information.
	Organising a larger unit of discourse, describing and comparing two pictures.
	Responding appropriately to a series of questions on a common theme, giving information, expressing and justifying opinions, suggesting, comparing and contrasting.
	Organising a coherent and extended opinion/argument with examples and justification, speculating, describing .

**iii) Test production process**

Comparability and quality requires the adoption of clearly defined :

- test specifications and item writer guidelines
- test production processes
- item authoring tools and item banking systems
- quality control processes

The steps in the test development process are shown below:



**Test specifications and item writer guidelines**

The test specifications described in the test design section above are developed into detailed item writer guidelines for each of the tested skills. These guidelines specify the requirements of each task type at each level in terms of overall testing aim, testing focus, level of distraction in the options, input text length, etc. They also provide explicit guidance on the selection and manipulation of text types and topics, and the production of artwork and recordings. The avoidance of any topics or contexts of language use which might introduce a bias against any group of candidates of a particular background (i.e. on the basis of sex, ethnic origin, etc.) is specified in the guidelines.

Quality criteria relevant to each task type are listed and these criteria provide the basis for the acceptance, rejection and editing of tasks as they proceed through the item production process. In essence, the strict guidelines are followed in order to ensure that all new materials written conform fully to the test specifications.

## Item writers and commissioning

A large number of specialist item writers from our established and experienced teams were commissioned to produce the number of items required to operate computer adaptive tests across all levels of the CEFR in the INTO *English Language Assessment*, and to produce specific academic focussed writing tasks. Some item writers specialise in certain skills, levels or task types. Item writers are organised into teams and managed by team leaders and specialist language testing product managers.

Before item writing begins, the number of items required are calculated and apportioned across the item writer teams. It is imperative to plan for adequate coverage of construct, domains and topics for all tasks at each level. Each item writer therefore receives a detailed commissioning brief specifying the task types, levels and topics to ensure adequate and consistent coverage of the CEFR domains as specified above.

## Pretesting

After pre-editing and editing, all items are pretested. Pretesting plays an important role as it allows for questions and materials with known measurement characteristics to be banked so that new banks of test material can be produced as and when required. The pretesting process helps to ensure that all versions conform to test requirements in terms of content and level of difficulty for the appropriate target level of use.

During pretesting new items are taken by many different candidates and the data compared to anchor items that have known measurement characteristics, resulting in the new items being calibrated and linked to a common scale of difficulty.

Pre-testing and calibration are one part of our established tried and tested ongoing review process, carried out by dedicated teams of experts in the field.

## 1.2 Cambridge English Item Banking

All materials for Cambridge English tests are drawn from a Cambridge English Local Item Banking System (LIBS). This in-house development is a secure electronic item bank. LIBS is a unique test item storage and management system which allows for attributes to be assigned to every item in the database, including measurement data that links all test materials to the Common European Framework of Reference (CEFR).

LIBS is fully secure. It is used for all stages of item production, including test construction. Before test items are included in live Cambridge English tests, they will have been through a rigorous test production process, which includes quality assurance at every stage to ensure the quality, validity and reliability of all test items. Test items are fully measurable against the CEFR, free from bias and suitable for use with students from all language and cultural backgrounds.

The tasks for use in the INTO *English Language Assessment* are selected from the item bank in order to construct a test specific selection that are used by the Computer Adaptive Test algorithm to deliver items to candidates according to their ability. Exposure of tasks is monitored by the Research and Validation department to ensure that material security is optimised. This test specific pool of items is monitored, updated and refreshed on a regular basis.

## 1.3 Test reliability

### Reliability

The computer-based INTO *English Language Assessment* is adaptive, presenting candidates with items appropriate to their estimated ability. This means that different candidates will not receive the same items and that split-half methods for calculating reliability, such as Cronbach’s Alpha, cannot be used. An analogous measure, the Rasch reliability, is used instead (Linacre, 1993; Jones, 2000). Rather than raw scores, this measure uses candidates’ ability estimates (measured in logits) to evaluate the test reliability. Table 1 below reports the Rasch reliability of the INTO *English Language Assessment* based on all tests sat between 1 January and 20 June 2015.

**Table 1: Reliability (Rasch) of INTO *English Language Assessment* by component and as a whole.**

Listening reliability	Reading and Language Knowledge reliability	Overall test reliability
0.87	0.85	0.91

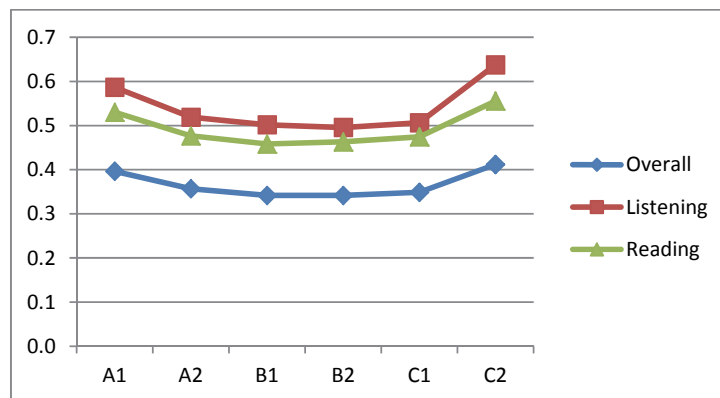
### Standard Error of Measurement

Another way of describing the precision of a test is in terms of the ability estimates of individual candidates, and the likely variation in those estimates from the candidate’s real or true ability. This is what the *Standard Error of Measurement* (SEM) provides (Wright and Stone, 1979). Table 2 below reports the mean SEM for the simulated INTO *English Language Assessment* tests, reported in logits. The figure below it shows the same information broken down by CEFR level.

**Table 2: Mean SEM (in logits) of INTO *English Language Assessment*, by component and as a whole.**

Listening SEM	Reading and Language Knowledge SEM	Overall test SEM
0.52	0.47	0.35

**Figure: Mean SEM (in logits) of INTO *English Language Assessment* by CEFR level**



**Note:** The mapping between logits, reported here, and standard scores, reported on the test itself, is non-linear—equal bands of ability measured in logits do not always map onto equal bands measured in scores. We use logits here as it more clearly illustrates that the test’s precision is greatest in the middle CEFR levels.

## 2. Description of test reporting in relation to the CEFR

### 2.1 What are the test specifications?

The INTO *English Language Assessment* is currently available as a Reading and Listening test, combined with a Writing test and a Speaking test.

The Reading and Listening test is adaptive, which means that the test responds directly to the candidate's responses: if the candidate gets questions right, the test gets more difficult; if they get questions wrong, it gets easier until the correct level for the candidate is found. This also means that each INTO *English Language Assessment* is unique. The Reading and Listening test is marked automatically and results are available immediately.

The Writing and Speaking tests are marked by examiners, so results are available once the tests have been marked.

### 2.2 How are results reported and what do they mean?

Components of the INTO *English Language Assessment* are reported using the Cambridge English Scale. For each component the highest score is 185. Scores between 180 and 185 receive a CEFR level of 'C1', although scores of 185 should be interpreted as 'C1 or above'.

Score	CEFR level
180-185	C1 (Effective proficiency)
160-179	B2 (Vantage)
140-159	B1 (Threshold)
120-139	A2 (Waystage)
100-119	A1 (Breakthrough)
82-99	Below A1

Candidates who sit all four components will also receive an Overall Score and CEFR level. Candidates scoring below 82 do not receive a score or a level. It will be displayed on the Test Report Form as below:

<b>Overall Score</b>	150	<b>CEFR Level</b>	B1
----------------------	-----	-------------------	----

Mode	Score	CEFR Level
Listening	126	A2
Reading	115	A1
Writing	173	B2
Speaking	185	C1

#### Reporting results for Reading and Listening

The INTO *English Language Assessment* Reading and Listening component is a fully computer adaptive test (CAT). Results are reported instantly as an individual score for each component on the Cambridge English Scale. Scores are reported from 82-185, and indicate a CEFR level from Below-A1 – C1 or above.

**Marking the Writing component**

The Writing component is automarked with results available soon after test completion. Results are reported as an individual score on the Cambridge English Scale. Scores are reported from 82-185, and indicate a CEFR level from Below-A1 – C1 or above.

The automarker engine marks a range of elements relating to Vocabulary, Grammar and Content at text, sentence and word levels, and produces a mark based on a weighted combination of the features identified.

**Marking the Speaking component**

The Speaking component is examiner marked. The examiner marks are pegged to fixed points in the Cambridge English Scale to represent the CEFR grades and interim grades that examiners award.

These examiner marked components are reported within 48 hours of the test being taken.

The Speaking component is marked by examiners using the mark scheme below. A candidate must meet all the descriptors in a band to be awarded that band. The mid-level bands, e.g. A2+, B1+ etc. allow an examiner to award a suitable mark when a performance displays aspects of a higher and lower level descriptor in a single response.

**INTO English Language Assessment Speaking Mark Scheme – Released Version v1.0**

<b>BAND</b>	<b>GLOBAL DESCRIPTORS</b>
<p><b>C2</b></p> <ul style="list-style-type: none"> <li>coherence / discourse management</li> <li>language resource</li> <li>pronunciation</li> <li>hesitation / extent</li> </ul>	<p><b>Fully operational command of the spoken language</b></p> <ul style="list-style-type: none"> <li>Able to express both simple and complex ideas with ease; coherent extended discourse.</li> <li>Consistently, displays wide range and accurate use of grammar and vocabulary.</li> <li>Pronunciation is easy to understand; stress, rhythm and intonation are used to express meaning effectively.</li> <li>Responds promptly with only natural hesitation; makes effective use of the allowed response time.</li> </ul>
<p><b>C1+</b></p>	<p>Some features of 5 and some features of 6 in approximately equal measure</p>
<p><b>C1</b></p> <ul style="list-style-type: none"> <li>coherence / discourse management</li> <li>language resource</li> <li>pronunciation</li> <li>hesitation / extent</li> </ul>	<p><b>Good operational command of the spoken language</b></p> <ul style="list-style-type: none"> <li>Able to express simple and complex ideas; generally extends discourse coherently.</li> <li>Generally, displays wide range and accurate use of grammar and vocabulary.</li> <li>Pronunciation is easy to understand; stress, rhythm and intonation are used to express meaning well.</li> <li>Generally responds promptly, with only natural hesitation; generally makes good use of the allowed response time.</li> </ul>
<p><b>B2+</b></p>	<p>Some features of 4 and some features of 5 in approximately equal measure</p>
<p><b>B2</b></p> <ul style="list-style-type: none"> <li>coherence / discourse management</li> <li>language resource</li> <li>pronunciation</li> <li>hesitation / extent</li> </ul>	<p><b>Generally effective command of the spoken language</b></p> <ul style="list-style-type: none"> <li>Able to express simple ideas and makes some attempt to express complex ideas; mostly coherent, with some extended discourse.</li> <li>There is an adequate range of grammar and vocabulary which is sufficiently accurate to deal with the tasks.</li> <li>Pronunciation can generally be understood; stress, rhythm and intonation are used to express meaning adequately.</li> <li>May be some hesitation while searching for language; generally makes adequate use of the allowed response time.</li> </ul>
<p><b>B1+</b></p>	<p>Some features of 3 and some features of 4 in approximately equal measure</p>
<p><b>B1</b></p> <ul style="list-style-type: none"> <li>coherence / discourse management</li> <li>language resource</li> <li>pronunciation</li> <li>hesitation / extent</li> </ul>	<p><b>Limited but effective command of the spoken language</b></p> <ul style="list-style-type: none"> <li>Able to express simple ideas; little extended discourse; some incoherence.</li> <li>The range of grammar and vocabulary used is sufficient to complete tasks in a limited way. Some language in simple utterances is accurate but basic inaccuracies may impede communication of ideas and achievement of the tasks.</li> <li>Pronunciation can generally be understood but L1 features may cause strain; an attempt is made to use aspects of stress, rhythm and intonation to express meaning.</li> <li>Hesitation may demand patience of the listener; use of the allowed response time may not always be adequate.</li> </ul>
<p><b>A2+</b></p>	<p>Some features of 2 and some features of 3 in approximately equal measure</p>
<p><b>A2</b></p> <ul style="list-style-type: none"> <li>coherence / discourse management</li> <li>language resource</li> <li>pronunciation</li> <li>hesitation / extent</li> </ul>	<p><b>Basic command of the spoken language</b></p> <ul style="list-style-type: none"> <li>No extended discourse</li> <li>The range of language is sufficient to respond to simple prompts but not to complete complex tasks. Some utterances (single words or short phrases) may be accurate but inaccuracies in grammar and vocabulary limit achievement of the tasks and restrict coherence and communication of ideas.</li> <li>Pronunciation of single words may be intelligible but L1 features may make understanding difficult; little attempt is made to use aspects of stress, rhythm and intonation to express meaning.</li> <li>Hesitation is excessive; use of the allowed response time is adequate on only a few occasions.</li> </ul>
<p><b>A1+</b></p>	<p>Some features of 1 and some features of 2 in approximately equal measure</p>
<p><b>A1</b></p> <ul style="list-style-type: none"> <li>coherence / discourse management</li> <li>language resource</li> <li>pronunciation</li> <li>hesitation / extent</li> </ul>	<p><b>Minimal command of the spoken language</b></p> <ul style="list-style-type: none"> <li>Utterances may be limited to single words.</li> <li>The range of language is limited and inadequate to complete the tasks. Some accurate language but frequent inaccuracies may mean the message is not communicated.</li> <li>Pronunciation of single words may be intelligible but L1 features may cause excessive strain to a listener; no attempt is made to use aspects of stress, rhythm and intonation to express meaning.</li> <li>Hesitation is excessive; use of the allowed response time is generally inadequate.</li> </ul>
<p><b>0</b></p>	<p>Throughout the task, responses are not attempted, OR consistently no meaning is conveyed, OR responses are consistently unrelated to the rubric.</p>



### 3. How the test is delivered?

#### **Where can the INTO *English Language Assessment* be taken and who administers the test**

The INTO *English Language Assessment* is taken online (via a web browser), using the Metrica delivery system. Metrica is an intuitive, flexible and secure test delivery system that delivers tests in a wide range of situations. It allows tests to be run anywhere, even with a wireless network, which removes the need for dedicated computer labs. Its leading-edge technology ensures quick and easy test session preparation, and includes a back-end system with marking and reporting tools.

#### **What are the exam protocols and procedures?**

The following procedures are to be followed when administering the INTO *English Language Assessment*:

- After receiving the URL/link and log on details from their institution, the candidate will login to the system and are taken to the Welcome Screen. The candidate will complete a short Candidate Information Screen where they will be asked to enter relevant information. The screen will have a link to advise them of appropriate data protection information.
- Each candidate has a home screen containing links to the test components which they have been registered for.
- The candidate will be advised how long the test will take (approx.45 minutes for the Reading and Listening component, and 40 minutes for the Writing component), and be prompted to take a test. Before the Listening items, the candidate will be prompted to check the audio level.
- When the test begins, the candidate will be presented with a series of questions designed to quickly and accurately assess their level.
- After the test has established a rough estimation of the candidate's ability, it will present them with more finely targeted tasks to confirm this evaluation.
- The questions will provide an interesting and engaging variety of interaction types presented in a simple, immediately accessible and intuitive interface.
- A progress bar will show how far the candidate is through the test, encouraging them to finish. The Reading and Listening test will typically take approximately 45 minutes to complete, though as is a Computer Adaptive Test (CAT), this depends on the responses provided by the candidate and the number of questions it needs to confirm the candidate's ability level.

#### **How we will ensure that the person taking the test is the genuine candidate?**

The INTO delivery partners will be responsible for checking candidate ID. The test should be invigilated. A Quick Start Guide and Instructions to Supervisors document is provided with instructions for supervision of the test

As an adaptive test, candidates each get a 'unique' test, seeing different items and in a different order. The item bank is encrypted for security and new items are being added to the bank on a regular basis.

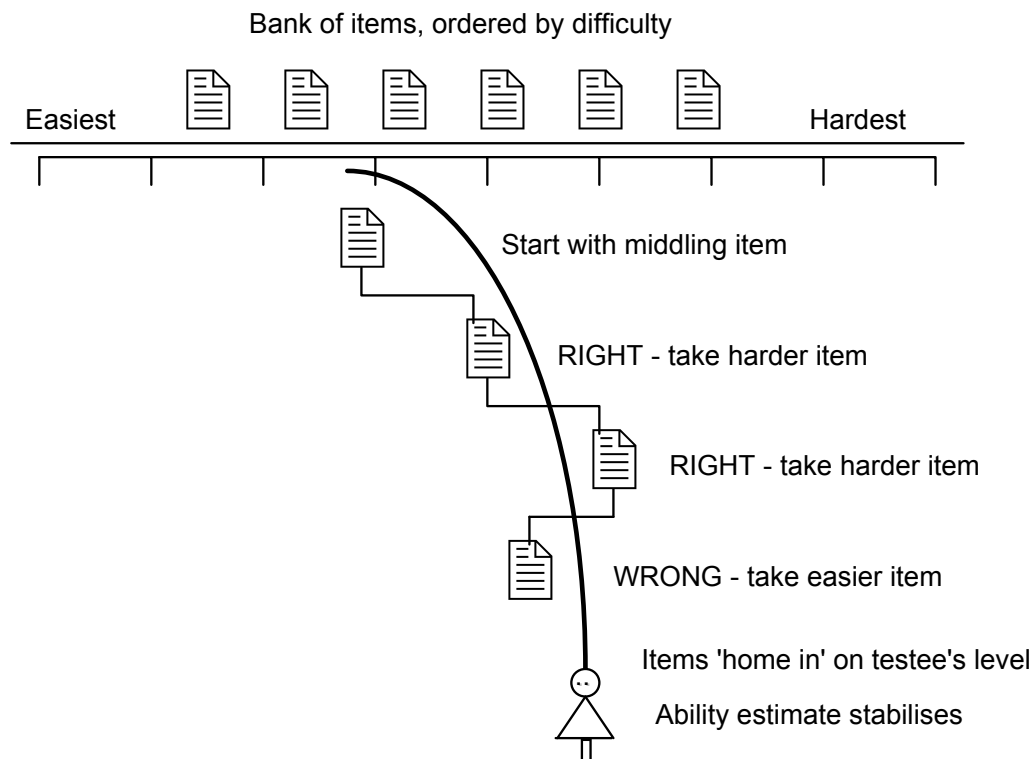
## 4. Quality control

Quality control procedures are included in each stage of the test production process. The strict item writer guidelines are designed to ensure that the materials conform to the test specifications. Items are rejected, rewritten or accepted on the basis of these guidelines. The multi-stage, detailed test production process illustrated above ensures that tasks are reviewed, edited and trialled several times before they are considered fit for purpose. Each individual task is thoroughly and repeatedly checked and proofread by external professional proof readers and signed off by internal team leaders and test production managers before being used in the test bank. Quality control is further enhanced by the use of the test delivery tool described above with its state of the art web-based authoring, banking and test assembly functionality which controls the whole process from task specification, through item writing and editing, to pretesting and live use.

## 5. How does Computer Adaptive Testing work in the INTO reading and listening test?

In a Cambridge English Computerised Adaptive Test (CAT) candidates are exposed to items close to their ability, based on their responses to each item in the test. The computer presents the candidate with an item, assesses their response as right or wrong, and then selects the next item at an appropriate level of difficulty. This means that if the candidate gets an item wrong, the next item will be easier; if they get an item right the next will be more difficult. At each stage the candidate's ability is estimated, and the difficulty of the test adjusted until it has homed in at their true ability level (see Figure 1 below).

**Figure 1 Schematic Diagram Illustrating the Basic Principle of Computerised Adaptive Testing**



The primary advantage of Cambridge English Computer Adaptive Tests is that they are more efficient than traditional linear tests. Items well matched to a candidate provide more information about the candidate's ability than items which are too easy or too difficult. By targeting items specifically at the ability of the candidate, each response in a CAT test contributes a maximum possible amount of information. Thus, the same level of precision can be achieved by a CAT test in a fewer number of items and shorter amount of time than a linear test. From a candidate perspective, the targeting of a CAT test also ensures that the difficulty of the test appears challenging, but not too high or low; stimulating the candidate rather than causing boredom or frustration. To enable this kind of targeting, the item bank underlying each Cambridge English Computer Adaptive Test consists of several hundreds of items, distributed across a range of ability levels and item types. Consequently, unlike a linear test, each candidate sees a different group of purposely selected items.

**References:**

- Jones, Neil (2000) Reliability, SEM and Classification Error. Issues in Controlling and Reporting Reliability. Cambridge ESOL Internal EFL Validation Report No 188.
- Linacre, John Michael (1993) Rasch-Based Generalizability Theory, in Rasch Measurement Transactions 7:1. Rasch Measurement SIG of the AERA.
- Wright, Benjamin D and Stone, Mark (1979) Best Test Design. Chicago: MESA Press.

## Appendix 1: Technical Setup

These are the minimum technical requirements that each computer will need in order to run The INTO *English Language Assessment* correctly:

Minimum Windows specs.	
Processor	Pentium III 800 MHz or equivalent
RAM	512 MB
Free hard drive space	4 GB
Screen resolution	1280 x 1024
Operating system	Windows XP, Vista, 7 and recent versions of Apple OSX
Internet connection	1 Mbps
Web browser	Internet Explorer version 8 (or later) or recent version of Safari, Firefox or Chrome
Adobe Acrobat	version 7 (or later)
Adobe Flash Player	version 10.1 (or later)
Other	Sound card and headphones (or headset for Speaking test)

Please note that the INTO *English Language Assessment* can also be taken on Apple Macs or Linux computers with similar minimum specifications. iOS (iPad or iPhone) or Android devices are **not** currently supported.

The hardware requirements are not very high for Metrica tests; it doesn't take up a lot of memory or disk space. However, all computers will need a sound card (or built-in sound).

For the Reading and Listening test you will need a reliable set of headphones. You should check that these work before candidates start the test.

Most importantly, you will need to have a reliable network connection with sufficient bandwidth to enable all users to access the test content at the same time. We recommend that as a first step of technical set up a speed test is run on at least one of the computers to be used for running the tests. It is important to run the speed test at the same time/day as that of the intended test. This is to give the most realistic results based on the volume of internet connection traffic on the institution's systems at that time. Ideally you are looking for upload and download speeds above 6 Mbps.

## Appendix 2: Quality Assurance Processes for IELA Writing and Speaking tests.

### Quality Assurance of IELA Examiners

#### Recruitment

Writing and Speaking Examiners for the IELA test are selected from the existing Cambridge English Writing and Speaking examiner cadres. IELA Examiners must have completed training and annual certification for B1 (Preliminary) and B2 (First) level exams before they are eligible to be trained for IELA marking. This means they will already have experience of marking scripts from A2 to C1 levels.

#### Quality Assurance

Once selected, examiners must successfully complete IELA specific training and certification before they are invited to mark. Examiners train for Writing and/or Speaking separately.

The on-going quality assurance of Examiners is managed by Cambridge English, and is supervised by a Principal Examiner (PE). Cambridge English ensures that all examiners successfully complete examiner training and regular certification of procedure and assessment before they examine.

Annual examiner certification involves the marking of sample tests in an online environment.

Examiners must complete standardisation of assessment for Writing and/or Speaking each year and are regularly monitored during live testing sessions.

Monitoring is managed by Cambridge English and is conducted by PEs. Examiners mark candidate responses in a secure online marking environment. The system randomly allocates candidate responses to ensure that individual examiners do not receive a concentration of good or weak responses, or a concentration from any one centre if multiple centres are testing concurrently. The system allows for examiner marking to be monitored for quality and consistency. During marking, the PEs can view an examiner's progress and offer support and advice, as required. PEs double mark a random sample of examiner scripts before the marks are released. Any divergence is checked and adjustments made. Any examiner thought to be marking off standard is directed back to the certification stage. This must be successfully completed before they can return to marking.

#### Escalated Responses

During marking the automarker can escalate a response to a PE before a mark is released. A Writing response may be escalated if it:

- is vastly off topic.
- Is identified as untypical in some other way.

A writing response will not be awarded a mark if it:

- is very short or blank.
- does not meet the minimum descriptor for a mark.

An examiner may escalate a Speaking response if it:

- is very short or blank.
- is mostly inaudible because it is quiet or because of background noise.
- does not meet the minimum descriptor for a mark.

PEs are provided with guidance on how to approach each of these scenarios. The main aim being to mark a performance wherever possible, but to ensure it is a genuine attempt at the given task.

Responses that are marked as zero are routinely checked by a PE to confirm that mark.

**Writing and Speaking Enquiries on Results**

In situations where an INTO centre has an enquiry about a Writing or Speaking Exit test mark, that forms part of a formal assessment stage Cambridge English will endeavour to check that mark within a 48 hour turnaround.

Enquiries should only be raised when:

- the mark forms part of an overall formal assessment
- the mark forms a jagged profile with marks for other components, i.e. more than one whole CEFR level difference

In these cases the centre would need to provide:

Centre name

Session

Date test taken

User Id

Password

Candidate first name

Candidate surname

The candidate response will be reviewed by a PE, and a response sent to the centre regarding the outcome of the enquiry. If a grade change is required this will be implemented and released into the reporting system within the 48 hours turnaround.

## Appendix 3: Access to IELA Writing Responses

Access to writing responses is for Academic Management and Teachers only. Scripts must not be passed to candidates or any other party. A writing response can be downloaded for any of the following reasons:

- i) As a learning and progress tool to use as a basis for defining learning needs at the start of a course and to compare with performance in the exit test. This may also be useful for showing external adjudicators that the courses you run and the tests you use are effective.
- ii) As a check on jagged profiles to see if there is a justification to raise an enquiry on results.

To address the data protection concerns

- scripts are only to be accessed by Academic Management/Teachers.
- scripts must not be passed to the students.

If a response is being accessed for purpose ii) above, an Enquiry on results can only be made if it meets the other criteria of that process, i.e.

Enquiries should only be raised when

- the mark forms part of an overall formal assessment.
- the mark forms a jagged profile with marks for other components, i.e. more than one whole CEFR level difference.