Cambridge English Language Assessment: Principles of good practice for test development and quality management

Dr Angeliki Salamoura, Research and Validation Group
Graeme Harrison, Head of Assessment Services, Southern Cone & Andes
Overview

• Cambridge English Language Assessment: who we are & what we offer
• Principles of test development
• Quality management: Question Paper Production process
• Cambridge English: Advanced (CAE)
• Teaching Knowledge Test (TKT)
A not-for-profit university department

University of Cambridge

Cambridge English

Cambridge English Language Assessment

Cambridge University Press
Cambridge English educational value chain

Language Policy & strategy

Programme implementation & evaluation

Analysis & diagnostics

Curriculum reform

Assessment & certification

Teacher Development

Materials & resources
Assessment and Certification

- English language exams since 1913
- Over 50,000 preparation centres worldwide
- One of the largest dedicated research teams
- 2,700 test centres in 130 countries
- 4.5 million candidates per year
- Accepted by 12,500 organisations globally
English Opens Doors: Chile

Cambridge English Language Assessment worked closely with the Chilean Ministry of Education to gather and diagnostically analyse data on language learning levels that helped validate and refine its English Opens Doors programme.

In 2003, as part of its strategy for improving quality and equity in its education system, the Chilean Ministry of Education launched a national English project, ‘English Opens Doors’.

One of the project’s aims was for all state school students to reach a level equivalent to Cambridge English: Key (Level A2) by the end of primary school and a level equivalent to Cambridge English: Preliminary (Level B1) by the end of high school.

Evaluating the existing English competence of school students was an important first step in this strategy, and we were chosen to design a diagnostic test to establish students’ listening and reading comprehension.

The diagnostic test we designed was developed from our own bank of materials, which had already been extensively internationally trialled and validated.

The diagnostic test consisted of two parts which tested Reading and Listening and was calibrated to the Council of Europe’s Common European Framework of Reference for Languages.

A sample of 12,000 students, representing 8th grade (13–14 year olds) and 12th grade (17–18 year olds) school students, took the test in 2004. A similar profile of students was tested again in 2008 to determine what progress had been made.

Cambridge English Language Assessment is delivering bespoke English language exams across the Chilean state education sector, testing up to 240,000 students over three years. SIMCE (Quality of Education Evaluation System) is a Department of the Chilean Ministry of Education (MINEDUC). They required English language testing as part of their biennial evaluation of student performance, undertaken in a number of key subjects. The Instituto Chileno Británico (a major training provider in Chile) used Cambridge English expertise to successfully bid for the contract in June 2012.

As a result of the bid, a bespoke test was created - SIMCE Ingles - which is now used extensively to test reading and listening comprehension in the Chilean education sector, with up to 240,000 students aged 15 assessed over three years, (2012-2014). As part of the contract, we are also responsible for the processing of all test results.

SIMCE Ingles is closely based on a Cambridge English Main Suite exam and was selected by SIMCE as its testing system as it met all SIMCE’s testing objectives, including test alignment with the Common European Framework of Reference (CEFR).

In addition, SIMCE also valued our experience of working with clients in the state sector, and the quality of the management processes underpinning the testing system. This uncompromising approach to quality, accuracy and reliability was a major reason why SIMCE opted to work with the Cambridge team, as it ensured the successful, nationwide delivery of a major testing programme.
1913: The first Certificate of Proficiency in English (CPE)

The 1913 UCLES Regulations:

‘The Certificate of Proficiency in English is designed for Foreign Students who desire a satisfactory proof of their knowledge of the language with a view to teaching it in foreign schools.’
1913: The first CPE

- Candidates: 3
- Exam centres: 2
- Fee: £3 (≈ 2,500 CLP)
- Length (time):
1913: The first CPE

- Candidates: 3
- Exam centres: 2
- Fee: £3 (≈ 2,500 CLP)
- Length (time): 12 hours!
1913: The first CPE

2. The Subjects of Examination will be:
   (i) Written: (a) Translation from English into French or German: 2 hours.
       (b) Translation from French or German into English, and questions on English Grammar: 2½ hours.
   (c) English Essay: 2 hours.
   (d) English Literature (The paper on English Language and Literature [Group A, Subject 1] in the Higher Local Examination): 3 hours.
   (e) English Phonetics: 1½ hours.

(ii) Oral: (f) Dictation: ½ hour.
      (g) Reading aloud and Conversation: ½ hour
1913: The first CPE

2. The Subjects of Examination will be:

(i) Written: (a) Translation from English into French or German: 2 hours.
(b) Translation from French or German into English, and questions on English Grammar: 2½ hours.
(c) English Essay: 2 hours.
(d) English Literature (The paper on English Language and Literature [Group A, Subject 1] in the Higher Local Examination): 3 hours.
(e) English Phonetics: 1½ hours.

(ii) Oral: (f) Dictation: ½ hour.
(g) Reading aloud and Conversation: ½ hour
1913: The first CPE

Pass rate: 0%

2. The Subjects of Examination will be:

(i) Written: (a) Translation from English into French or German: 2 hours.
    (b) Translation from French or German into English, and questions on English Grammar: 2½ hours.

(c) English Essay: 2 hours.

(d) English Literature (The paper on English Language and Literature [Group A, Subject 1] in the Higher Local Examination): 3 hours.

(e) English Phonetics: 1½ hours.

(ii) Oral:  
(f) Dictation: ½ hour.

(g) Reading aloud and Conversation: ½ hour
1913: The first CPE

The 1913 examiner report on Phonetics:

‘None of the candidates [had] given the subject sufficient attention. The transcriptions were distinctly poor and suggested that the candidates had had but little practice…’
Cambridge English exams and the CEFR

- **Flyers** (C2), **Movers** (B2), **Starters** (B1), **Beg-A2**
- **KET** (A2) 1988
- **PET** (B1) 1981
- **FCE** (B2) 1939
- **CAE** (C1) 1991
- **CPE** (C2) 1913

- **Breakthrough Level**
- **Council of Europe Threshold level**
- **Effective Operational Proficiency level**
- **Mastery level**
Maintaining standards over time - an item banking approach using IRT

Probabilistic models:
Estimating the probability that a candidate of known ability will succeed on an item of known difficulty

Item Response models and Rasch analysis provide the statistical approach for building a measurement scale and doing item banking
A test is a specific (and necessarily limited) sample of performance, from which users seek to make broader generalisations about the knowledge, skills, or abilities a person has.
Test qualities…

- Validity
- Reliability
- Impact
- Practicality
Balancing test qualities…

Validity

Reliability

Impact

Practicality
The longer the test...

...the more ___________ it is.

...the less ___________ it is.

valid       reliable       practical
A Fine Balance
Multiple Choice Test
Short Response

MARINE WILDLIFE PHOTOGRAPHER

Bruce says that [7] is the most important aspect of his work.

Before going on a trip, Bruce makes [8] of the photographs he hopes to take.

Knowing the type of photographs he wants to take helps Bruce to choose the right [9]

Bruce disagrees with people who say his way of taking photographs is not [10]
Speaking Test
Speaking Test

Describe and explain why…
- Evaluate performance on a scale

Repeat the phrase you hear in the recording
- Accurately repeated = 1
- Not accurately repeated = 0

© UCLES 2013
Balancing test qualities…

Validity

Reliability

Impact

Practicality
How do we ensure reliability?
Speaking and Writing

Recruitment

Training

Certification

Standardisation & Recertification

Monitoring

Examiner Administration
How do we ensure test validity & fitness for purpose?
Socio-cognitive framework for test validation
(Weir 2005)
How are the characteristics of test takers catered for by this exam?
Are the characteristics of the test tasks and the test administration fair to the test takers?
Socio-cognitive framework for test validation
(Weir 2005)

Are the cognitive processes required to complete the task appropriate?
How far can we depend on the scores of the test?
Socio-cognitive framework for test validation
(Weir 2005)

What effects does the test have on its various stakeholders?
What external evidence is there outside of the test scores themselves that the test is doing a good job?
Does the test measure the right things?

Research Notes: quarterly online publication

Studies in Language Testing:
Examing Writing (2007)
Examing Reading (2009)
Examing Speaking (2011)
Examing Listening (2013)
Cambridge English Language Assessment

- Assessment Managers
- Assessment Administrators
- Operations Administrators
- Validation Officers
- Clerical Markers

External Specialists
- Chair (writing team leader)
- Item Writing Team
- Content Vetters
- Proofreaders
- Examiners

Stakeholders
- Pretest centres
- Pretest candidates
The Recipe for Question Paper Production

Commissioning of material

Pre-editing

Editing

Rejection

Revision

© UCLES 2013
Pre-editing and Editing – the cake mix

Participants
Assessment Manager
Chair (writing team leader)
Writers

Aims
Topic
Topicality
Level of language
Suitability for the task
Length
Focus
Style
Accuracy
The Recipe for Question Paper Production

Commissioning of material

Pre-editing

Editing

Rejection

Pretesting

Revision
Aims of Pretesting – the taste test

• Generate statistical data
  • Difficulty
  • Discrimination
• Generate qualitative feedback
• Check the quality and fairness of items and tasks
The Recipe for Question Paper Production

Commissioning of material
  ↓
Pre-editing
  ↓
Editing
  ↓
Pretesting & Trialling
  ↓
Pretest Review
  ↓
Revision
  ↓
Pretest Review
  ↓
Pretesting & Trialling
  ↓
Editing
  ↓
Pre-editing
  ↓
Commissioning of material
  ↓
Rejection
Pretest Review – the proof of the pudding…

Participants
Assessment Manager
Chair (writing team leader)
Experienced writer

Aims
evaluate results
finalise key
The Recipe for Question Paper Production

Commissioning of material

Pre-editing

Editing

Pretesting & Trialling

Pretest Review

Test Construction
Test Construction

Participants:
Assessment Manager
Chair (writing team leader)
Experienced writer

Aims:
Construct question papers
Ensure correct level of difficulty
Coverage of skills and appropriate content
The Recipe for Question Paper Production

- Commissioning of material
  - Pre-editing
  - Editing
  - Pretesting & Trialling
  - Pretest Review
  - Test Construction
  - Paper checking before sign-off
Cambridge English
A range of exams to meet different needs
What is Cambridge English: Advanced (CAE)?

- Established and high-quality test of English
- Taken by hundreds of thousands of students globally
- Focused on the level needed for university study (CEFR C1)
- Proven validity and reliability
Accepted by 3,000 organisations worldwide

- Employers: Accenture, Bayer, Ernst & Young, Dell
- Higher education: UK, Australia, North America
- Immigration: UK Border Agency, DIAC - Australia
- Teachers: Mexico, Finland
- State schools: Romania, Germany
Why is CAE different?

More than just a snapshot of English skills

Thermometer test

C2
C1
B2
B1
A2
A1

Level of English needed for most academic purposes
Helps students improve their English

- Positive impact
- Students take a language improvement course
- Focus on communicative language skills
- Focus on grammar and vocabulary
Test content

<table>
<thead>
<tr>
<th></th>
<th>Activity</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reading</td>
<td>75 min</td>
</tr>
<tr>
<td>2</td>
<td>Writing</td>
<td>90 min</td>
</tr>
<tr>
<td>3</td>
<td>Use of English</td>
<td>60 min</td>
</tr>
<tr>
<td>4</td>
<td>Listening</td>
<td>40 min</td>
</tr>
<tr>
<td>5</td>
<td>Speaking</td>
<td>15 min</td>
</tr>
</tbody>
</table>
# Test content

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Reading and Use of English - 90 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Writing – 90 min</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Listening - 40 min</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Speaking - 15 min</td>
</tr>
</tbody>
</table>
Enhanced security

- Test day photos
- Statistical analysis of results before release
- Online results verification service
Results and Certificate

UNIVERSITY of CAMBRIDGE
ESOL Examinations
English for Speakers of Other Languages

Cambridge ESOL Level 2 Certificate in ESOL International

This is to certify that
A.N. EXAMPLE
has been awarded
Grade B
in the
Certificate in Advanced English
Council of Europe Level C1

Score
73/100

Grade C

Candidate Profile

© UCLES 2013
Why take Cambridge English: Advanced (CAE)?

- **Established** – recognised worldwide
- **Positive impact** – focus on language skills
- **High quality** – test content; administration; results
Teaching Knowledge Test (TKT)

- For both experienced and new teachers
- Flexible, modular structure
TKT - Overview

Module 1: Background to language teaching

Module 2: Planning for language teaching

Module 3: Classroom management

TKT: Practical

TKT: Young Learners

TKT: CLIL

TKT: Knowledge about Language
Summary

- Quality Assurance through the QPP process
- Cambridge English: Advanced for C1 certification
- Teaching Knowledge Test for teacher development
Further information

- Learn more about Cambridge English teaching qualifications
- www.cambridgeenglish.org/sector/teaching

- For more information on Cambridge English: Advanced

- Harrison.G@CambridgeEnglish.org
Thank you!

Any questions?
A range of fit for purpose exams
1913: The first CPE

- **items about grammar & lexis:**
  ‘Give the past tense and past participle of… dividing them into strong and weak; add explanations:
  *tell, wake, buy*

- **items on grammar & lexical usage:**
  ‘Embody each of the following words into a sentence in such a way as to show that you clearly apprehend its meaning:
  *commence, comment, commend…*’
What can students do with CAE?

Examples CEFR Level C1 abilities

<table>
<thead>
<tr>
<th>Listening/Speaking</th>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAN follow abstract argumentation, for example the balancing of alternatives and the drawing of a conclusion.</td>
<td>CAN read quickly enough to cope with the demands of an academic course.</td>
<td>CAN write an essay which shows ability to communicate, giving few difficulties for the reader.</td>
</tr>
</tbody>
</table>
### Typical abilities at CEFR Level C1

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Follow an academic course at university level</td>
</tr>
<tr>
<td>2</td>
<td>Communicate effectively at managerial and professional level</td>
</tr>
<tr>
<td>3</td>
<td>Deal confidently with living in an English-speaking environment</td>
</tr>
<tr>
<td>4</td>
<td>Participate effectively in meetings or academic tutorials and seminars</td>
</tr>
<tr>
<td>5</td>
<td>Carry out complex and challenging research</td>
</tr>
<tr>
<td>6</td>
<td>Express yourself with a high level of fluency</td>
</tr>
<tr>
<td>7</td>
<td>React appropriately in different cultural and social situations</td>
</tr>
</tbody>
</table>
What is the impact if students are below this level?

- Increased requirement for support
- Students fail to reach academic potential
- Dissatisfaction amongst home students
<table>
<thead>
<tr>
<th></th>
<th>Test content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reading - 1 hour 15 minutes</td>
</tr>
<tr>
<td>2</td>
<td>Writing - 1 hour 30 minutes</td>
</tr>
<tr>
<td>3</td>
<td>Use of English - 1 hour</td>
</tr>
<tr>
<td>4</td>
<td>Listening - 40 minutes</td>
</tr>
<tr>
<td>5</td>
<td>Speaking - 15 minutes</td>
</tr>
</tbody>
</table>
# Reading

<table>
<thead>
<tr>
<th>Part 1</th>
<th>Part 2</th>
<th>Part 3</th>
<th>Part 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple choice</td>
<td>Gapped text</td>
<td>Multiple choice</td>
<td>Multiple matching</td>
</tr>
<tr>
<td>6 questions</td>
<td>6 questions</td>
<td>7 questions</td>
<td>15 questions</td>
</tr>
</tbody>
</table>
THE THEORY OF EVERYTHING

Time was when physicists dreamed of a final theory of fundamental physics, a perfect set of equations that would describe every force and particle in nature. Today that dream is being overtaken by the suspicion that there is no such thing. Some even fear that all attempts at a deeper understanding of nature are dead ends. This will lend support to those who have long claimed that research into fundamental physics is a waste of time and money; that at best it provides answers to obscure questions which few people understand or care about.

So do these reservations undermine pure physics as a scientific pursuit? Surely, it makes no difference if the truths that physicists seek turn out to be more complex and messy than they once hoped. It could even make the search more intriguing. There are as many profound questions out there as there have ever been, and to answer them physicists need the kind of hard experimental evidence that can only come from pure research.

Can we, therefore, justify spending the huge sums of money that such research demands? What it boils down to is whether we think the search for fundamental truths is important. This quest for knowledge is a defining human quality, but it's hard to quantify how our lives have been 'improved' by it. There have been plenty of technological spin-offs from the space race and other experiments. But the spin-offs are not the point. In showing us how the universe works, fundamental physics could also tell us something profound about ourselves. And for that, a few billion dollars would be a small price to pay.
# Writing

<table>
<thead>
<tr>
<th>Part 1</th>
<th>Part 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compulsory question</td>
<td>1 task from choice of 5</td>
</tr>
<tr>
<td>Write one of the following: article, report, proposal or letter.</td>
<td>One of the following: article, competition entry, contribution to a longer piece, essay, information sheet, a letter, a proposal, a report, a review</td>
</tr>
<tr>
<td>180-220 words</td>
<td>220–260 words</td>
</tr>
</tbody>
</table>
The main character in The Pelican Brief is Darby Shaw. At the beginning of the novel she is studying law and having a relation with Thomas Callahan, one of her professors. Together they become interested in the suspicious deaths of two Supreme Court judges. Darby writes a report about it and this document leads to Callahan's violent death. Darby knows she is in great danger but this does not prevent her to continue the investigation.

As this brief description suggests, Darby is a very brave person. She was very lucky that she did not die with Callahan. She knows that her actions are making dangerous men become angry, she tries to hide from them but she does not stop her investigation. She is afraid but she still wants to find the truth.

Darby also possesses many other strengths. She is a very intelligent person, for example she is able to work out the connection between the two judges. In addition, she has very deep principles. She realises that very important people are involved in the situation but she is still determined to make the truth public. She knows that the environment is in threat and this is very important for her.
# Use of English

<table>
<thead>
<tr>
<th>Part 1</th>
<th>Part 2</th>
<th>Part 3</th>
<th>Part 4</th>
<th>Part 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple-choice cloze</td>
<td>Open cloze</td>
<td>Word formation</td>
<td>Gapped sentences</td>
<td>Key word transformations</td>
</tr>
<tr>
<td>12 questions</td>
<td>15 questions</td>
<td>10 questions</td>
<td>5 questions</td>
<td>8 questions</td>
</tr>
</tbody>
</table>
Use of English - example

Part 2

For questions 13 – 27, read the text below and think of the word which best fits each gap. Use only one word in each gap. There is an example at the beginning (0).

Write your answers IN CAPITAL LETTERS on the separate answer sheet.

Example: 0 BY

Mosquitoes

According to the World Health Organisation, malaria, a disease spread (0) ....... mosquitoes, affects millions of people every year. Everyone knows how irritating the noise made by a mosquito, (13) ....... by a painful reaction to its bite, can be. It is astonishing that so (14) ....... is known about why mosquitoes are drawn to or driven away from people, given (15) ....... level of distress and disease caused by these insects. We know that the most effective chemical (16) ....... protecting people against mosquitoes is diethyltoluamide, commonly shortened (17) ....... deet. (18) ....... deet works well, it has some serious drawbacks: it can damage clothes and some people are allergic to it.
## Listening

<table>
<thead>
<tr>
<th>Part 1</th>
<th>Part 2</th>
<th>Part 3</th>
<th>Part 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple choice</td>
<td>Sentence completion</td>
<td>Multiple choice</td>
<td>Multiple Matching</td>
</tr>
<tr>
<td>6 questions</td>
<td>8 questions</td>
<td>6 questions</td>
<td>10 questions</td>
</tr>
</tbody>
</table>

© UCLES 2013
Listening

MARINE WILDLIFE PHOTOGRAPHER

Bruce says that _______ is the most important aspect of his work.

Before going on a trip, Bruce makes _______ of the photographs he hopes to take.

Knowing the type of photographs he wants to take helps Bruce to choose the right _______.

Bruce disagrees with people who say his way of taking photographs is not _______.

© UCLES 2013
# Speaking

<table>
<thead>
<tr>
<th>Part 1</th>
<th>Part 2</th>
<th>Part 3</th>
<th>Part 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversation between the interlocutor and each candidate (spoken questions)</td>
<td>Individual ‘long turn’ for each candidate with brief response from the second candidate. In turn, the candidates are given three pictures to talk about</td>
<td>A two-way conversation between the candidates. The candidates are given spoken instructions with written and visual stimuli, which are used in a decision-making task</td>
<td>A discussion on topics related to the collaborative task (spoken questions)</td>
</tr>
<tr>
<td>3 minutes</td>
<td>A 1-minute ‘long turn’ for each candidate, plus a 30-second response from the second candidate</td>
<td>4 Minutes</td>
<td>4 minutes</td>
</tr>
</tbody>
</table>
Speaking

http://www.cambridgeesol.org/assets/wmv/univ/c1-speaking.wmv