IELTS Equivalence Research Project (GMC 133)

Appendices

Dr Lynda Taylor and Dr Sathena Chan

5/13/2015

CONFIDENTIAL

Table of Contents

Appendix 1 – Phase 1 analysis of English Language Proficiency (ELP) tests (updated Feb 2015) 3

Appendix 2 – Example letter of enquiry to test provider (a) 9

Appendix 3 – Example letter of enquiry to test provider (b) 11

Appendix 4 - A summary of outcomes from Phase 1 follow-up with test providers (July 2014) 13

Appendix 5 – Detailed analyses of the selected reading tests (IELTS, CPE, CAE, PTE-A, TOEFL iBT, OET, GEPT Advanced, ECPE, IESOL & ISESOL Level 2 and IESOL & ISESOL Level 3) 16

Appendix 6 – Detailed analyses of the selected writing tests (IELTS, CPE, CAE, PTE-A, TOEFL iBT, OET, GEPT Advanced, ECPE, IESOL & ISESOL Level 2 and IESOL & ISESOL Level 3) 40

Appendix 7 – Detailed analyses of the selected speaking tests (IELTS, CPE, CAE, PTE-A, TOEFL iBT, OET, GEPT Advanced, ECPE, IESOL & ISESOL Level 2 and IESOL & ISESOL Level 3) 68

Appendix 8 – Detailed analyses of the selected listening tests (IELTS, CPE, CAE, PTE-A, TOEFL iBT, OET, GEPT Advanced, ECPE, IESOL & ISESOL Level 2 and IESOL & ISESOL Level 3) 96

Appendix 9 - Survey of English Language Proficiency (ELP) tests used for professional registration purposes 117

Appendix 10 – List of Approved English Language Test Providers Criteria For Inclusion (2010) 129

CONFIDENTIAL

Appendix 1 – Phase 1 analysis of English Language Proficiency (ELP) tests (updated Feb 2015)

Name	Full name of test	Provider	CEFR target level	1. High-level security procedures for test production and delivery	2. A verification system that allows score users to check and confirm test results	3. Assessment of all 4 communicative skills (reading, writing, listening, speaking)	4. Parallel forms of the test available for all test components	5. Evidence of the reliability of the test and the dependability of its scores	6. An approved test centre (or centre network) routinely monitored and inspected for QA	7. Demonstrable evidence that the test measures at the CEFR C1 level and/or above	8. Appropriate test domain (e.g. Academic /professional)	Does CRELLA need to send a request to the test provider asking them to supply missing information regarding specific criteria?	Comments
CPE	Cambridge English: Proficiency (CPE)	Cambridge Assessment	C2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	According to the information available on their websites, these 4 tests met all 8 criteria of the Phase 1 screening process. They were therefore automatically included in the Phase 2 test scrutiny.
TOEFL iBT	Internet-based Test of English as a Foreign Language	Educational Testing Service, USA	B1-C1	Yes	Yes	Yes (Speaking - semi-direct)	Yes	Yes	Yes	Yes	Yes	No	
PTE (AC)	Pearson Test of English (Academic)	Pearson	B1 -C1	Yes	Yes	Yes (Speaking - semi-direct)	Yes	Yes	Yes	Yes	Yes	No	
CAE	Cambridge English: Advanced (CAE)	Cambridge Assessment	C1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	
OET	Occupational English Test	Cambridge Boxhill Language Assessment, Melbourne, Australia	*N/A	Yes	*No details provided on the webpage	Yes	Yes	*No details provided on the webpage	Yes	*No details provided on the webpage	Yes - Medical	Yes - for Criteria 2, 5 and 7 and check further on CEFR alignment.	*In response to a specific follow-up enquiry, the test provider supplied documentation which showed that the test meets Criteria 2, 5 and 7 and evidence of CEFR alignment (for details see Appendix 4). OET was thus included in the Phase 2 test scrutiny.
ECPE	Examination for the Certificate of Proficiency in English	Cambridge-Michigan	C2	Yes	Yes	Yes	Yes	Yes	Yes (also available in Greece)	**No, but …''Please contact us if you have any questions about the CEFR and CaMLA exams''.	Yes	Yes - for Criterion 7	**In response to a specific follow-up enquiry, the test provider supplied documentation which showed that the test meets Criterion 7. ECPE was included in the Phase 2 test scrutiny in the extended research.
EIKEN	EIKEN Grade Grade 1	Jitsuyō Eigo Ginō Kentei (Test in Practical English Proficiency), Japan	C1 (TOEFL iBT 100)	***No details provided on the webpage	Yes	Yes	Yes	***No details provided on the webpage	Yes	Yes	Yes	Yes - for Criteria 1 and 5. NB: Test only available in Japan and currently under review/revision	*** The test provider supplied documentation which showed that the test meets Criteria 1. Regarding Criterion 5, data are available for Reading, Listening and Writing only. The Speaking component had undergone a revision procedure. The reliability of the Speaking scores from live tests will be available in near future. This test was not included in the Phase 2 analysis but may be a candidate for a supplementary study in the future.

CONFIDENTIAL

Name	Full name of test	Provider	CEFR target level	1. High-level security procedures for test production and delivery	2. A verification system that allows score users to check and confirm test results	3. Assessment of all 4 communicative skills (reading, writing, listening, speaking)	4. Parallel forms of the test available for all test components	5. Evidence of the reliability of the test and the dependability of its scores	6. An approved test centre (or centre network) routinely monitored and inspected for QA	7. Demonstrable evidence that the test measures at the CEFR C1 level and/or above	8. Appropriate test domain (e.g. Academic /professional)	Does CRELLA need to send a request to the test provider asking them to supply missing information regarding specific criteria?	Comments
GEPT	General English Proficiency Test Advanced	LTTC, Taiwan	C1	****No details provided on the webpage	****No details provided on the webpage	Yes	Yes	Yes	Yes	Yes	Yes	Yes - for Criteria 1 and 2. NB: Test only available in Taiwan.	*****The test provider supplied documentation which showed that the test meets Criteria 1 and 2. GEPT Advanced was included in the Phase 2 test scrutiny in the extended research.
IESOL+ ISESOL	International ESOL and International Spoken ESOL (if combined)	City and Guilds, London	C2	Yes	*****No details provided on the webpage	Yes	Yes	*****No details provided on the webpage	Yes	Yes	Yes	Yes – for Criteria 2 and 5	******The test provider supplied documentation which showed that the test meets Criteria 2 and 5. IESOL/ISESOL Level 2 and Level 3 were included in the Phase 2 test scrutiny in the extended research.
IESOL+ ISESOL	International ESOL and International Spoken ESOL (if combined)	City and Guilds, London	C1	Yes	******No details provided on the webpage	Yes	Yes	******No details provided on the webpage	Yes	Yes	Yes	Yes – for Criteria 2 and 5	
MELAB	Michigan English Language Assessment Battery	Cambridge-Michigan	B1-C1	Yes	Yes	Yes but speaking is optional	Yes but the speaking test is not offered at every test center.	Yes	Yes	*******No but …''Please contact us if you have any questions about the CEFR and CaMLA exams''.	Yes	Yes - for Criterion 7. Also request information on CEFR alignment and a list of test centres that offer the speaking test.	*******The test provider supplied documentation which shows that the test partially meets Criterion 7 (for details see Appendix 4). The test was not eligible for inclusion in the Phase 2 analysis.
Can-Test	CAN TEST	University of Ottawa	N/A	No details provided on the webpage	Yes	Yes	Yes	No details provided on the webpage	Yes	No details provided on the webpage	Yes	The test providers were contacted to confirm whether alignment with CEFR has been undertaken. No response received.	A public claim about these tests' (equivalent) CEFR level does not appear to be provided by the test provider. Also, the tests do not appear to meet **several** of the key criteria. The test was not eligible for inclusion in the Phase 2 analysis.
CAEL	Canadian Academic English Language Assessment	Carleton University, Ottawa, Canada	N/A	Yes	Yes	Yes	Yes	No details provided on the webpage	Yes	No details provided on the webpage	Yes	The test providers were contacted to confirm whether alignment with CEFR has been undertaken. No response received.	A public claim about these tests' (equivalent) CEFR level does not appear to be provided by the test provider. Also, the tests do not appear to meet **several** of the key criteria. The test was not eligible for inclusion in the Phase 2 analysis.

Name	Full name of test	Provider	CEFR target level	1. High-level security procedures for test production and delivery	2. A verification system that allows score users to check and confirm test results	3. Assessment of all 4 communicative skills (reading, writing, listening, speaking)	4. Parallel forms of the test available for all test components	5. Evidence of the reliability of the test and the dependability of its scores	6. An approved test centre (or centre network) routinely monitored and inspected for QA	7. Demonstrable evidence that the test measures at the CEFR C1 level and/or above	8. Appropriate test domain (e.g. Academic /professional)	Does CRELLA need to send a request to the test provider asking them to supply missing information regarding specific criteria?	Comments
ISLPR	International Second Language Proficiency Ratings	ISLPR Lang Services Pty Ltd, Australia	N/A / assessed on ISLPR levels	No details provided on the webpage	No details provided on the webpage	Yes	No details provided on the webpage	No details provided on the webpage	Yes	No details provided on the webpage	Yes	The test providers were contacted to confirm whether alignment with CEFR has been undertaken. No response received.	A public claim about these tests' (equivalent) CEFR level does not appear to be provided by the test provider. Also, the tests do not appear to meet **several** of the key criteria. The test was not eligible for inclusion in the Phase 2 analysis.
CET	College English Test Band 6	Ministry of Education of the People's Republic of China	N/A	No details provided on the webpage	No details provided on the webpage	Yes	Yes	No details provided on the webpage	Yes	No details provided on the webpage	Yes	The test providers were contacted to confirm whether alignment with CEFR has been undertaken.	A public claim about these tests' (equivalent) CEFR level does not appear to be provided by the test provider. Also, the tests do not appear to meet **several** of the key criteria. The test providers replied that they are not currently in a position to claim a precise correspondence between their test and the CEFR framework. The test was not eligible for inclusion in the Phase 2 analysis.
CELBAN	Canadian English Language Benchmark Assessment for Nurses	Centre for Canadian Language Benchmarks	N/A / assessed on Canadian Language Benchmarks (CLB) level	No details provided on the webpage	Yes	Yes	Yes	No details provided on the webpage	Yes	No details provided on the webpage	Medical	The test providers were contacted to confirm whether alignment with CEFR has been undertaken.	Nurses are the target test population of the test. Also, the tests do not appear to meet **several** of the key criteria. The test providers replied that they are not currently in a position to claim a precise correspondence between their test and the CEFR framework. The test was not eligible for inclusion in the Phase 2 analysis.
MET	NES Medical English Test	NES Healthcare	No details provided on the webpage	No details provided on the webpage	No details provided on the webpage	NO (Speaking is evaluated during RMO skills testing)	No details provided on the webpage	No details provided on the webpage	No details provided on the webpage	No details provided on the webpage	Medical	No	There is no publically available info at this point in time to determine whether it meets the criteria for selection to Phase 2. Also, the tests do not appear to meet **several** of the key criteria. The test was not eligible for inclusion in the Phase 2 analysis.
	Certificates in ESOL Skills for Life Level 2	Cambridge Assessment	C1	Yes	Yes	Yes	Yes	Yes	Yes	Yes?	No	No	This test targets the domain of General English and for Schools. It is available in the UK only. The test was not eligible for inclusion in the Phase 2 analysis.

Name	Full name of test	Provider	CEFR target level	1. High-level security procedures for test production and delivery	2. A verification system that allows score users to check and confirm test results	3. Assessment of all 4 communicative skills (reading, writing, listening, speaking)	4. Parallel forms of the test available for all test components	5. Evidence of the reliability of the test and the dependability of its scores	6. An approved test centre (or centre network) routinely monitored and inspected for QA	7. Demonstrable evidence that the test measures at the CEFR C1 level and/or above	8. Appropriate test domain (e.g. Academic /professional)	Does CRELLA need to send a request to the test provider asking them to supply missing information regarding specific criteria?	Comments
TOEIC	Test of English for International Communication	Educational Testing Service, USA	A1-C1	No (for more details, see the comments box)		Yes			Yes		Yes	No	Following the broadcast of a BBC Panorama program in February 2014 which highlighted an organized criminal element seeking to circumvent the U.K.'s visa-granting process, ETS has made the decision not to extend their Secure English-Language Testing (SELT) license with the Home Office. As a result, TOEIC® testing will no longer be offered for U.K. visa-granting purposes. http://www.ets.org/toeic/important_update/update_toeic_uk The test was not eligible for inclusion in the Phase 2 analysis.
BULATS	Business Language Testing Service Online Tests	Cambridge Assessment	A1-C2	No details provided on the webpage	No details provided on the webpage	Yes	Yes	No details provided on the webpage	Yes	No details provided on the webpage	Business	No	These tests are set in a specific professional domain, e.g. Legal and Financial so the tests were not eligible for inclusion in the Phase 2 analysis.
ICFE	Cambridge English Financial (ICFE)	Cambridge Assessment	B2-C1	Yes	Yes	Yes	Yes	Yes	Yes	No details provided on the webpage	Financial	No	
ILEC	Cambridge English: Legal (ILEC)	Cambridge Assessment	B2-C1	Yes	Yes	Yes	Yes	Yes	Yes	No details provided on the webpage	Legal	No	
BEC H	Cambridge English: Business Higher	Cambridge Assessment	C1	Yes	Yes	Yes	Yes	Yes	Yes	No details provided on the webpage	Business	No	
ISE 3	Integrated Skills in English (ISE) Examination 3	Trinity College Examinations London	C1	No details provided on the webpage	No details provided on the webpage	No (Portfolio, Writing and Speaking) (New ISE will be launched in 1 September 2015*)	Yes	No details provided on the webpage	Yes	Yes		No	The current test did not assess all 4 individual skill components so it was not considered eligible for inclusion in the Phase 2 analysis. * The new ISE will be launched in 1 September 2015. The new ISE assesses the 4 skill components in two integrated papers. The new version may be a candidate for a supplementary study in the future. At the end of February, the UKIV took a decision that only Trinity's ISE, GESE (Speaking component only) and IELTS would be considered SELTs from 6 April 2015. No evidence has been made available in support of this decision.

Name	Full name of test	Provider	CEFR target level	1. High-level security procedures for test production and delivery	2. A verification system that allows score users to check and confirm test results	3. Assessment of all 4 communicative skills (reading, writing, listening, speaking)	4. Parallel forms of the test available for all test components	5. Evidence of the reliability of the test and the dependability of its scores	6. An approved test centre (or centre network) routinely monitored and inspected for QA	7. Demonstrable evidence that the test measures at the CEFR C1 level and/or above	8. Appropriate test domain (e.g. Academic /professional)	Does CRELLA need to send a request to the test provider asking them to supply missing information regarding specific criteria?	Comments
TOEFL PBT and TWE	TOEFL Paper based Test (is being phased out) (has to be taken with TOEFL TWE) / Test of Written English	Educational Testing Service, USA	B1-C1			No (No speaking)						No	These tests do not assess all 4 individual skill components so the tests were not eligible for inclusion in the Phase 2 analysis.
GESE	Graded Examinations in Spoken English (Grade 10-11)	Trinity College Examinations London	C1			No (Speaking only)			Yes			No	
GESE	Graded Examinations in Spoken English (Grade 12)	Trinity College Examinations London	C2			No (Speaking only)			Yes			No	
USMLE	United States Medical Licensing Examination (3 steps)	The Federation of State Medical Boards and the National Board of Medical Examiners	N/A	Yes	Yes	No	Yes	No details provided on the webpage	Yes	No details provided on the webpage	Medical	No	
BEC P	Cambridge English: Business Preliminary	Cambridge Assessment	B1									No	These tests are targeted to measure at a level BELOW the critical target level of C1 so the tests were not eligible for inclusion in the Phase 2 analysis.
BEC V	Cambridge English: Business Vantage	Cambridge Assessment	B2									No	
FCE	Cambridge English: First (FCE)	Cambridge Assessment	B2									No	
PET	Cambridge English: Preliminary (PET)	Cambridge Assessment	B1									No	

Name	Full name of test	Provider	CEFR target level	1. High-level security procedures for test production and delivery	2. A verification system that allows score users to check and confirm test results	3. Assessment of all 4 communicative skills (reading, writing, listening, speaking)	4. Parallel forms of the test available for all test components	5. Evidence of the reliability of the test and the dependability of its scores	6. An approved test centre (or centre network) routinely monitored and inspected for QA	7. Demonstrable evidence that the test measures at the CEFR C1 level and/or above	8. Appropriate test domain (e.g. Academic /professional)	Does CRELLA need to send a request to the test provider asking them to supply missing information regarding specific criteria?	Comments
	Cambridge IGCSE English - Second Language (BES)	Cambridge Assessment	Cambridge International Level 1 /Level 2 Certificates									No	
	Cambridge IGCSE: English - First Language	Cambridge Assessment	Cambridge International Level 1 /Level 2 Certificates									No	
KET	Cambridge Key (KET)	Cambridge Assessment	A2									No	
	Certificates in ESOL Skills for Life Entry 1-3	Cambridge Assessment	A1-B1									No	
	Certificates in ESOL Skills for Life Level 1	Cambridge Assessment	B2									No	
ECCE	Examination for the Certificate of Competency in English	Cambridge-Michigan	B2									No	These tests are targeted to measure at a level BELOW the critical target level of C1 so the tests were not eligible for inclusion in the Phase 2 analysis.
GEPT - HI	General English Proficiency Test High-Intermediate	LTTC, Taiwan	B2									No	
ISE 1	Integrated Skills in English (ISE) Examination 1	Trinity College Examinations London	B1									No	
ISE 2	Integrated Skills in English (ISE) Examination 2	Trinity College Examinations London	B2									No	
ISE F	Integrated Skills in English (ISE) Examination Foundation	Trinity College Examinations London	A2									No	

NB. Tests shaded in grey were NOT included in the Phase 2 analysis.

Appendix 2 – Example letter of enquiry to test provider (a)

Centre for Research in English Language Learning and Assessment (CRELLA)
University of Bedfordshire
Putteridge Bury
Hitchin Road, Luton
Bedfordshire
LU2 8LE, UK

14 July 2014

Dear language testing colleague

We are writing to request your assistance with a research project that the *Centre for Research in English Language Learning and Assessment (CRELLA)* is currently undertaking.

CRELLA has recently been commissioned by the UK General Medical Council (GMC) to prepare a review of English language tests. The aim of this review is to survey and investigate a range of currently available English language proficiency tests in terms of their suitability for use within the GMC registration process for internationally trained doctors wishing to practice in the UK. The GMC currently uses IELTS as the sole English language proficiency measure for this purpose, but is interested to ascertain whether other English language tests merit consideration for approval alongside IELTS.

Phase 1 of the project has so far identified approximately a dozen English language proficiency tests that appear to meet all (or most) of a set of 8 essential quality criteria as specified by the GMC. Examples of such essential criteria include: the existence of high-level test security procedures; a test score verification system for users; evidence of test reliability and score dependability; and evidence of alignment to the C1 level of the Common European Framework of Reference (CEFR). Subject to the outcomes of Phase 1, a selected number of tests will be carried forward to Phase 2 of the project, where they will be subjected to closer scrutiny and content analysis for their suitability for the GMC purpose outlined above.

For all the tests on the Phase 1 list, information on the 8 essential criteria has thus far been sought and obtained through a process of systematic desk research using the test provider's website and other documentation available in the public domain, i.e. academic volumes, research reports, publicity brochures, etc.

For a few of the tests, however, it has not proved possible to find the relevant information on some of the 8 essential criteria. For this reason, we have decided to contact the test providers individually, with a request to:

- **supply us direct with the relevant information we are seeking,**

CONFIDENTIAL

or

- **suggest to us where we can readily access the outstanding information - online or in print.**

We should be most grateful, therefore, if you could provide us with information on the following criterion/a for your test:

Criterion 2:
A verification system that allows score users to check and confirm test results. [Does this currently exist?]

Criterion 5:
Evidence of the reliability of the test and of the dependability of the test scores.

Criterion 7:
Demonstrable evidence that the test measures at the CEFR C1 level and or above.

As you will appreciate, we are keen to make our survey of potentially suitable English language proficiency tests for the GMC as comprehensive as possible and to ensure, as far as we can, that no test should be ruled out unfairly or unreasonably from closer scrutiny and content analysis in Phase 2 of the study.

We should be most grateful, therefore, if you could send us relevant information concerning your test and Criteria 2, 5 and 7 as indicated above. We would be pleased to receive the information from you by email or regular mail **within 21 days** (i.e. by Monday 4th August) to enable us to proceed with the study according to schedule.

Thank you so much for your time and attention. Please do not hesitate to contact me if you have any questions regarding this letter or the project itself.

We look forward to hearing from you in due course.

With all good wishes

Dr Lynda Taylor
Senior Lecturer in Language Assessment - CRELLA
GMC Project Lead

CONFIDENTIAL

Appendix 3 – Example letter of enquiry to test provider (b)

Centre for Research in English Language Learning and Assessment (CRELLA)
University of Bedfordshire
Putteridge Bury
Hitchin Road, Luton
Bedfordshire
LU2 8LE, UK

14 July 2014

Dear language testing colleague

We are writing to request your assistance with a research project that the *Centre for Research in English Language Learning and Assessment (CRELLA)* is currently undertaking.

CRELLA has recently been commissioned by the UK General Medical Council (GMC) to prepare a review of English language tests. The aim of this review is to survey and investigate a range of currently available English language proficiency tests in terms of their suitability for use within the GMC registration process for internationally trained doctors wishing to practice in the UK. The GMC currently uses IELTS as the sole English language proficiency measure for this purpose, but is interested to ascertain whether other English language tests merit consideration for approval alongside IELTS.

A primary criterion set by the GMC for selecting any test for further consideration is the test's linkage or alignment to the Common European Framework of Reference for Languages (CEFR) at C1 level or above.

For some of the tests we have been reviewing we have so far been unable to confirm, through a process of desk research, whether CEFR linkage has either been claimed or demonstrated empirically by the test provider.

For this reason, we have decided to contact test providers individually, with a request to:

- **supply us direct with the relevant information we are seeking,**
  or
- **suggest to us where we can readily access the outstanding information - online or in print.**

We should be most grateful, therefore, if you could provide us with any information to answer the following 2 questions:

1. *Does your test publicly claim to measure at the CEFR C1 level and or above?*
2. *If so, is any there empirical evidence available to support such a claim?*

As you will appreciate, we are keen to make our survey of potentially suitable English language proficiency tests for the GMC as comprehensive as possible and to ensure, as far as we can, that no test should be ruled out unfairly or unreasonably from closer scrutiny and consideration.

We should be most grateful, therefore, if you could send us relevant information concerning your test and the CEFR as indicated above. We would be pleased to receive the information from you by

CONFIDENTIAL

email or regular mail **within 21 days** (i.e. by Monday 4th August) to enable us to proceed with the study according to schedule.

Thank you so much for your time and attention. Please do not hesitate to contact me if you have any questions regarding this letter or the project itself.

We look forward to hearing from you in due course.

With all good wishes

Dr Lynda Taylor
Senior Lecturer in Language Assessment - CRELLA
GMC Project Lead

CONFIDENTIAL

Appendix 4 - A summary of outcomes from Phase 1 follow-up with test providers (July 2014)

1. ECPE (Cambridge-Michigan Language Assessment)

Information required: *Demonstrable evidence that the test measures at the CEFR C1 level and/or above*

A validation study was conducted in 2005, following procedures outlined in the Council of Europe's *Relating language examinations to the Common European Framework of Reference: learning, teaching, assessment – A Manual.* The report shows that the judges concurred that each language skill is tested in the ECPE at or around level C2 on the CEFR. Only in the case of the listening section were the judges unable to conclusively place a skill at C2 as opposed to C1, and this was because all the activities within the listening skill category lack a C2 descriptor. For listening, the C2 descriptors are "as C1." In virtually all cases for each language skill, the level assigned for ECPE input and tasks was at the highest available descriptor.

The available documentation shows demonstrable evidence to support the test provider's claim that ECPE measures at the CEFR C1 level and/or above, meaning that the ECPE could justifiably be considered for inclusion in the Phase 2 analysis of the equivalence project. However, constraints on the resources available for the project combined with the time pressure to complete the study by end October 2014, suggest that we need to consider carefully how many additional tests to add to the 5 tests already identified for Phase 2 scrutiny through Phase 1 (i.e. IELTS, Cambridge Proficiency, Cambridge Advanced, Internet-based TOEFL and Pearson Test of English (Academic)).

The original CRELLA research proposal anticipated that 5-6 tests could realistically be included in Phase 2 but that any additional tests would necessitate additional time and resources beyond those provided for in the project plan (see page 5 of the original proposal – the section on *Key risks and mitigating factors*). While there remains scope to add one additional test (i.e. test number 6) to the current list, this should ideally be done on a priority basis if more than 1 potential candidate emerges. In our view, OET (see discussion below) constitutes a stronger candidate for inclusion in Phase 2 for two reasons: i) it is a medically-oriented language proficiency test, and ii) it is already used in some parts of the world for professional registration purposes. For these reasons, it is decided to prioritise OET over ECPE at this stage for inclusion in Phase 2, and to recommend that ECPE is not included but that it is considered a candidate for the follow-up study.

2. OET (Cambridge-Boxhill Language Assessment)

Information required:
(1) *Existence of a verification system that allows score users to check and confirm test results*
(2) *Evidence of the reliability of the test and the dependability of its scores*
(3) *Demonstrable evidence that the test measures at the CEFR C1 level and/or above*

CONFIDENTIAL

Cambridge-Boxhill Language Assessment was able to supply documentation (both information in the public domain and proprietary information which is internal and confidential) to show that the OET meets all three of these criteria. The most recently published version of *OET: Information for Health Professional Authorities, Employers and Educators* provides information concerning the test results verification service for score users: "A free and secure online results verification service exists to help organisations and authorities verify candidate results. The service provides access to a secure database through which candidates' results – across a period of three years – can be checked for authenticity. The site is securely encrypted and designed for use in high-stakes environments." (2013, p.6). CRELLA was also given sight of a more extensive public information booklet which is currently in preparation by the test providers (August 2014) and which gives a more extensive guide for stakeholders about how to check OET results online.

The publicly available *OET: Information for Health Professional Authorities, Employers and Educators* (p.10) also gives reliability estimates for the 4 sub-tests (Reading, Writing, Listening and Speaking) and estimates from the most recent 2014 reliability analyses were provided in confidence. CRELLA researchers were also given sight of two internal technical reports (which are still being finalized for external publication) concerning the linkage between OET, IELTS and the CEFR. One reported on a standard setting study, conducted in Cambridge in December 2013, which provides empirical evidence of the alignment of OET subtest scores to the CEFR levels. The second empirical study investigated the relationship between and comparability of scores achieved by several hundred candidates who took both OET and IELTS.

In light of this, and particularly given that OET is a medically-oriented language proficiency test which is already used in some parts of the world for professional registration purposes, OET is included in the Phase 2 analysis alongside the other 5 on the list - IELTS, Cambridge Proficiency, Cambridge Advanced, Internet-based TOEFL and Pearson Test of English (Academic).

3. Eiken Grade 1 (The Eiken Foundation, Japan)

Information required:
(1) *Existence of high-level security procedures for test production and delivery*
(2) *Evidence of the reliability of the test and the dependability of its scores*

The Eiken Foundation was able to provide documentation which shows that there are high-level security procedures for their test production and delivery. In addition, they were able to provide reliability data for the Reading, Listening and Writing components, but not for the Speaking component of the test. This is because the Speaking component has been undergoing a revision procedure. The test provider anticipates that a reliability estimate for speaking scores from live tests will become available in the near future.

In light of this, Eiken Grade 1 is not included in the Phase 2 analysis in this study but could be a candidate for consideration in the future when the reliability of speaking component becomes available. It is also worth noting that Eiken Grade 1 is only available in Japan. According to 2008-2013 data routinely gathered by the GMC and supplied to CRELLA, Japan is not one of the countries from which the GMC receives a high volume of applications for registration.

CONFIDENTIAL

4. GEPT (LTTC, Taiwan)

Information required:
(1) *Existence of high-level security procedures for test production and delivery*
(2) *Existence of a verification system that allows score users to check and confirm test results*

LTTC was able to supply documentation to show that the test meets both these criteria. While GEPT is a fairly strong candidate for consideration alongside IELTS, it is worth noting that the test is currently available in Taiwan only. According to 2008-2013 data routinely gathered by the GMC and supplied to CRELLA, Taiwan is not one of the countries from which the GMC receives a high volume of application for registration. Therefore, GEPT is not included in the Phase 2 analysis of the current study, though it is considered for scrutiny in a follow-up study.

5. MELAB (Cambridge-Michigan Language Assessment)

Information required: *Demonstrable evidence that the test measures at the CEFR C1 level and/or above*

Cambridge-Michigan Language Assessment provided documentation of a study which established the relationship between MELAB scores and TOEFL scores. They then established the hypothetical MELAB-CEFR relationships based upon a few relevant studies in the literature, which include (1) a study by ETS relating TOEFL scores to IELTS scores, (2) a study by IELTS relating IELTS scores to the CEFR, and (3) a study by ETS relating TOEFL scores to the CEFR

Taking into account that i) MELAB is targeted at CEFR levels B2-C1 (i.e. the test will include a substantial amount of material focusing below the critical C level), and ii) the relationship between the test scores and the equivalent CEFR level is partly empirical and partly hypothetical, MELAB is not included in the Phase 2 analysis in this study.

CONFIDENTIAL

Appendix 5 – Detailed analyses of the selected reading tests (IELTS, CPE, CAE, PTE-A, TOEFL iBT, OET, GEPT Advanced, ECPE, IESOL & ISESOL Level 2 and IESOL & ISESOL Level 3)

CONFIDENTIAL

TEST:	International English Language Testing System (IELTS)	COMPONENT:	Reading

Task Features					
Skills focus	To test students' ability to comprehend academic texts and to extract important information from those texts The IELTS Reading test is designed to assess a wide range of reading skills, including how well candidates: - read for the general sense of a passage - read for the main ideas - read for detail - understand inferences and implied meaning - recognise a writer's opinions, attitudes and purpose - follow the development of an argument				
Task description	There are 40 questions. A variety of question types is used, chosen from the following types: multiple choice, identifying information, identifying writer's views/claims, matching information, matching headings, matching features, matching sentence endings, sentence completion, summary completion, note completion, table completion, flow-chart completion, diagram label completion, short-answer questions. Sometimes you will need to give one word as your answer, sometimes a short phrase, and sometimes simply a letter, number or symbol. Mainly receptive, some limited writing involved in SAQ, but only brief answers are required; no more than XXX words. Candidates must be careful when writing answers on the answer sheet because they will lose marks for incorrect spelling and grammar.				
Timing	60 minutes to answer a total of 40 questions on 3 passages. Individual tasks are not timed.				
Weighting	All items equally weighted. Each correct answer receives one mark. Scores out of 40 are converted to the IELTS 9-band scale. Scores are reported in whole and half bands				
Cognitive processing Goal setting	**Expeditious reading: local** (scan/search for specifics)	**Careful reading: local** (understanding sentence)			
	Expeditious reading: global (skim for gist/search for key ideas/detail)	**Careful reading: global** (comprehend main idea(s)/overall text(s))			
Cognitive processing Levels of reading	Word recognition				
	Lexical access				
	Syntactic parsing				
	Establishing propositional meaning (cl./sent. level)				
	Inferencing				
	Building a mental model				
	Creating a text level representation (disc. structure)				
	Creating an intertextual representation (multi-text)				

Features of the Input Text					
Word count	Three different passages to read, each with accompanying questions. Officially candidates have to read 2,150 - 2,750 words in total. There are three sections to the IELTS Academic Reading test, and each contains one long text. Nuttall (1996) argues that a long text is required for candidates to skim for main ideas, scan for specific information, make relevance judgements and distinguish between main points and subsidiary details. Green, Weir and Unaldi (2010) analysed 42 texts making up 14 IELTS reading tests. The passages in their study contained 854 words on average (maximum 1063 words, minimum 589 words).				
Average sentence length	21.89 (This is an average figure. Individual sentences will span a range above and below the average)				
Domain	Social	Work	**Academic**		
Discourse mode	Descriptive	**Historical/biographical**	**Expository**	**Argumentative**	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	**Mostly concrete**	Fairly abstract	Mainly abstract	
Presentation	**Verbal**	Non-verbal (i.e. graphs)	**Both**		
Lexical Level; further criteria	The cumulative coverage reaches 92% at the K3 level. 76.4 K1, 11.36 K2, 3.26 K3. 1.09%>15K (UG mean 4.3) Academic Word level 7.9 (undergraduate mean 10.5). TTR. 53.68. Lexical Density 0.57 Mean number of modifiers per noun phrase 0.98 Mean number of words before main verb in main clause 5.48				

Grammatical level	C1 Grammatical exponents. CEFR : The description in the Common European Framework of Reference (CEFR) level for C1 and C2 is *many complex sentences*				
Readability	Flesch-Kincaid Grade Level 13 (12.64)				
Topic	A broad range of subject areas were represented among the 42 IELTS texts examined by Green et al (2010) with the categories of *Social studies* (10 or 11 texts), *Engineering & technology* (6 or 7) and *Business & administrative studies* (4 or 5) emerging as the most popular topic areas for the test.				
Rhetorical organization	**Explicit**				**Not explicit**
Text genre	There are three reading texts with a variety of question types. Texts are appropriate to, and accessible to, test-takers entering undergraduate or postgraduate courses, or seeking professional registration. The kinds of text used in IELTS are those that introduce academic topics to a general audience, often in the form of articles sourced from newspapers or magazines presenting research findings to a general audience. These include self- contained reports on developments in science and technology and overviews of academic debates. The IELTS texts often present solutions to problems that are likely to be of interest to the general reader. Texts come from books, journals, magazines, newspapers, reports and online resources, written for a non-specialist audience While IELTS passages are at a level of difficulty appropriate to university study, they are not as challenging as s in the more linguistically demanding areas such as the law textbook analysed by Green et al 2010.				
Intended Writer/Reader relationship	The relationship is not specified. The texts will typically be written for a general audience, not a specific reader.				
Reliability *[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored, i.e. correct or incorrect,, is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*					
Kuder-Richardson formula 20 (KR20)[1]	0.9				

CONFIDE

[1] A measure of internal consistency used to estimate reading test reliability.

TEST:	Cambridge English: Proficiency (CPE)	COMPONENT:	Reading

Task Features

Skills focus	CPE candidates are expected to be able to understand in detail a range of texts, both short and long, from the following sources: books (fiction and non-fiction); non-specialist articles from journals, magazines and newspapers; informational materials. They will be able to recognise the purpose and main ideas and details of the texts and the opinions and attitudes expressed in them. They will also recognise the structure of certain texts and be able to follow their development.
Task description	**Reading and Use of English** - 7 parts, 90 minutes - 1 hour 30 minutes 40% of total marks **Use of English** Parts 1–4 test underlying knowledge of vocabulary and grammar. Parts 1-3, 3 texts with 24 items: three short Use of English texts, each with a different task, such as supplying a missing word or forming a new word. Part 1 is mainly lexical, Part 2 is mainly grammatical, and Part 3 is all about how affixes affect words. There are lots of other sub-skills involved such as ability to paraphrase and work out meaning from context. Part 4 is not text based and involves rewriting 6 sentences another way. **Reading Component** Then there are three longer Reading texts: multiple choice, gapped paragraphs and multiple matching Part 5 text with 6 MCQ questions Part 6 text with 7 missing paragraphs Part 7 text with 10 multiple matching items Texts are not for a specialised readership. They are about interesting things in the real world. Some are simply of general interest; others will have an academic slant or work-related flavour. The reading skills needed include very detailed reading, following an argument, coherence and linking, looking for specific information and skim reading.
Timing	53 items administered in 90 minutes
Weighting	Items in Parts 1, 2, 3 and 7 receive 1 mark each; items in Parts 4, 5 and 6 receive 2 marks each.
Cognitive processing Goal setting	**Expeditious reading: local** (scan/search for specifics) / **Careful reading: local** (understanding sentence) **Expeditious reading: global** (skim for gist/search for key ideas/detail) / **Careful reading: global** (comprehend main idea(s)/overall text(s))
	Part 1-4 (30 items): Tests the candidates' sentence level processing abilities. Candidates have to establish propositional meaning within the sentence where the lexical item has been deleted and in part 4 reformulate a sentence including the given word. Part 5: Multiple Choice (6 items): one text with 4-option MCQs. Tests the ability of the candidates to integrate information and select the relevant macro-propositional content from the text. Inferencing required within a sentence and across sentences for a number of items. Part 6: Gapped Text (7 items): a text from which paragraphs have been removed and placed in jumbled order after the text along with one distractor. Candidates must decide from where in the text the paragraphs have been removed. Candidate benefits from reading the whole of the text to see how it fits together. The task involves integration of new information into an ongoing meaning representation as well as, albeit indirectly, establishing a text model. It also involves reading to learn unknown information with this particular example. At CAE/CPE reading to learn must be there according to the CEFR, hence, the need to expose candidates to new information they are unlikely to have previously encountered. Part 7: Multiple matching (10 items): Tests candidates' ability to search for the relevant macro-propositional content from a longer text. The longer text is likely to place greater demands on the cognitive processing since more information will contribute to the mental model that the candidate has to carry forward and that information will be linked in more complex ways. In the writing paper there is a summary task where candidates have to integrate meaning across texts. The literature on processing suggests that this is likely to require a higher level of processing than all the reading tasks we have reviewed so far.
Cognitive processing Levels of reading	**Word recognition** **Lexical access** **Syntactic parsing** **Establishing propositional meaning (cl./sent. level)** **Inferencing** **Building a mental model** **Creating a text level representation (disc. structure)**

	Creating an intertextual representation (multi-text)				
Features of the Input Text					
Word count	Six different passages to read, each with accompanying questions. 3 short passages and 3 long passages 600-1100 words. Maximum no of words for any single passage = 1100.				
Average sentence length	23.39 (This is an average figure. Individual sentences will span a range above and below the average). SD 7.2				
Domain	Social	Work	**Academic**		
Discourse mode	**Descriptive**	**Historical/ biographical**	**Expository**	**Argumentative**	**Instructive**
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete	**Fairly abstract**	Mainly abstract	
Presentation	**Verbal**	Non-verbal (i.e. graphs)	**Both**		
Lexical Level; further criteria	Very wide range of vocabulary including idiomatic expressions and colloquialisms as well as language relating to opinion, persuasion and abstract ideas. Candidates should be challenged by complexity of expression rather than arcane or specialist lexis but at this level they are expected to be able to cope with the full range of general topics that might interest an educated native speaker. As for CAE but with a range and appropriateness of vocabulary which an educated native-speaker might be expected to understand. Texts that can be understood in real- world professional contexts by native speakers Literal/ inferential evaluative/ synthesis/ analytical use. More complex fiction may be used By CPE learners need a very good appreciation of written English. They will be able to operate comfortably in most contexts of work and of higher education. The functions that are tested in CPE typically include those tested in CAE along with, additionally: presenting ideas, reporting, outlining, making proposals, judging priorities, analysing, drawing conclusions and making recommendations The new functions that candidates are asked to comprehend typically include a degree of appreciation of style with regard to the description of people, places or events and the presentation of a point of view. Functions that relate to aspects of presenting a reasoned argument are key at this level. Candidates are expected to comprehend arguments for and against, even when these are presented indirectly. There are only a few discernible differences in functions between the CPE and CAE examinations. The distinction lies more in the quality and range of writing that candidates are exposed to. Candidates at this level need to have a deep appreciation of the full range of functional exponents. They should be able to appreciate language used in a sophisticated way for the full range of personal, literary and academic functions The cumulative coverage reaches 91.11% at the K3 level. 78.95 K1, 8.45 K2, 3.71 K3. 2.09%>15K (UG mean 4.3) It is interesting to note what frequency level of vocabulary is necessary to reach the 95% coverage level suggested by Laufer (1988). This percentage Laufer suggests is necessary for learners to understand the gist of a text (and perhaps be able to inference the meaning of unknown words in the text). In the KET, PET, and FCE levels, learners would need to know the words in the K1-K3 bands. For CAE and CPE this goes up to include the K4-K6 bands. This suggests that examinees will need knowledge of many more words to successfully engage with the CAE and CPE texts. If we use a higher criterion (97%) closer to that suggested by Nation and Gu (2007), we find that the KET and PET readings would require knowledge of words at the K5 level, moving up to the K6 level for the FCE readings, and to the K10 level for the CAE and CPE readings. This suggests that for true ease in reading the passages (at least in lexical terms), examinees require a large vocabulary, even at the lower levels, but especially so at the higher levels. TTR (40 70) 80 sd 0.07 Lexical density .52 Academic Word level 5.82 (SD 2.8) (undergraduate mean 10.5) Mean number of modifiers per noun phrase 0.90 Mean number of words before main verb in main clause 5.26				
Grammatical level	C2 Grammatical exponents. CEFR : The description in the Common European Framework of Reference (CEFR) level for C1 and C2 is *many complex sentences* This level is typified by: many complex sentences; frequent use of modals; some use of ellipsis; complex approaches to referencing; use of synonymy. At Cambridge English: Proficiency level most sentences are long and complex. No restriction on the types of structure employed by the text. Many examples of structures typically used for effect in writing – sentences with several subordinate clauses, for example. The conveying of opinion both directly and indirectly is focused on more than the conveying of information				
Readability	Flesch-Kincaid Grade Level 9.9 Range 5.6 – 16.1 The low estimates for some texts at CPE might be a cause for some				

	concern as according to the CEFR students at these levels should be capable of processing undergraduate level texts which Weir et al (2006) found to be at around 13.5 in terms of Flesch Kincaid estimate
Topic	At CPE level more abstract and academic topics appear and the candidate may be expected to comprehend passages on any non-specialist topic. CPE candidates are expected to be able to operate confidently in a wide variety of social, work-related and study-related situations. In Cambridge English examinations, the sine qua non is that candidates should have sufficient knowledge of the topic to allow normal processing to take place. This means avoiding arcane topics where candidates would be unlikely to possess the schemata necessary for processing them on a global level. Topics are chosen to be accessible and of interest to the broad range of candidates and are not intended to exclude any large group in terms of their standpoint or assumptions. Care is taken to ensure that candidates come to the text equally, no matter where they are from, how old they are, what their background is, and so on (as far as it is possible to control for this). Clearly, it is impossible to interest everybody, but subjects which appeal only to a minority are avoided. The topic should not be biased in favour of any particular section of the test population. Material must not favour candidates with specialised knowledge of a particular subject or have content that would be too specialised or technical for the majority of candidates.
Rhetorical organization	**Explicit** | | | | **Not explicit**
Text genre	There are three long reading texts with a variety of question types. Texts are appropriate to, and accessible to, test-takers entering undergraduate or postgraduate courses, or seeking professional registration. CPE candidates are expected to be able to deal with both short and long texts, from the following sources: newspapers and magazines (e.g., articles, reports); journals; books (fiction and non-fiction extracts); promotional and informational materials (e.g., guides, manuals). Referential, poetic, emotive, conative purposes for CPE texts.
Intended Writer/Reader relationship	In Cambridge English practice, Main Suite exam texts are aimed at the general reader. At higher levels (CAE and CPE) texts assume a certain level of education and understanding of the world as well as of English

Reliability

[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]

Kuder-Richardson formula 20 (KR20)[2]	0.8 – 0.85 (34 items as against 40 items in IELTS)
	For most routine analysis, Cambridge uses the BIGSTEPS analysis program supplemented by in-house programmes for formatting data and handling the output. Items are pretested in specially constructed papers which include anchor items. Because the anchor items are of known difficulty, the analysis that is carried out on the pretest responses allows the new items to be calibrated on a logit scale. This scale produced by BIGSTEPS is re-scaled to produce a conventional ESOL scale, which is used for examination construction purposes. It is this scale which underpins the Local Item Banking System (LIBS). Examinations are constructed from the calibrated tasks in the item bank. Each task, therefore, consists of items of measured (Rasch) difficulty, which are selected from within a specified range to determine the mean difficulty of the task. In operational test production new items are pre-tested in specially constructed papers which include anchor items. Because the anchor items are of known difficulty, the analysis that is carried out on the pre-test responses allows the new items to be calibrated and linked to the common scale. Using this approach to model the predicted facility of new items, the test construction team is able to make a judgement on how examination tasks are likely to perform under live examination conditions. It also allows for a comparison between the "live" values and those modelled, at the time when analysis takes place for grading the examination (see Table 7.1 below). Mean difficulty of Reading Paper KET (A2) Dec 41 40 PET (B1) Dec 55 55 FCE (B2) June 62 62 CAE (C1) June 70 71 CPE (C2) June 78 77 The above table shows no or slight variation between the difficulty of the reading paper at the test construction stage and at the live administration stage which is an outcome of the rigorous procedures in place to ensure stability of difficulty levels at these stages. The standard operating procedures for test construction ensure that tasks selected for Main Suite reading papers fall within the specified range of difficulty and achieve the targeted average for the paper as a whole. These procedures help to ensure comparability of difficulty and maintenance of standards across

[2] A measure of internal consistency used to estimate reading test reliability.

	different versions of the paper and between different administration sessions. The overall targeted difficulty values for Main Suite reading papers on the Cambridge Common Scale are: 41 for KET, 56 for PET, 62 for FCE, 71 for CAE and 78 for CPE. The above table shows that the difficulty of the reading papers administered for each Main Suite examination over a period of two sessions in 2005 remained within acceptable range of variation, i.e., one to two points on either side of the difficulty point on the scale.

CONFIDENTIAL

TEST:	Cambridge: Advanced (CAE)	COMPONENT:	Reading

Task Features

Skills focus

At CAE, as well as global items that require the identification of main ideas, candidates have to understand the whole of the text in one of the tasks. They will be able to understand the detail, tone, purpose, main idea and implication of the texts, and the opinions and attitudes expressed in them. They will recognise text organization features such as exemplification, comparison and reference, and they will also understand how texts are structured and be able to follow text development.

In CAE the tasks focus on global reading to comprehend main ideas and the overall text. Expeditious reading through searching (rather than scanning) is tested in Part 4 but the items are mainly set at the local level.

Part 1 careful reading global: comprehend main ideas
Part 2 careful reading global: comprehend overall text
Part 3 careful reading global: comprehend main ideas
Part 4 expeditious local: search reading.

For CAE Part 2 candidates need to comprehend the information structure of the whole text. The ability to comprehend the whole of the text in CAE and how its parts interrelate is seen to be more difficult than understanding individual propositions and their relationship to the immediate context. CAE Part 2 requires the candidate to recognize macro- and micro- propositions, to establish the hierarchical links between a macro- proposition and the micros that depend upon it and to establish text structure on the basis of links between macro- propositions. Gapped-text items in CAE and CPE are expected to require a substantial amount of processing by the candidates. Ashton (2003: 128) notes '...the task has been designed to test a candidate's ability to understand text structure, and its successful completion is believed to involve candidates in performing at "whole text" level rather than processing identifiable chunks of text in order to answer each item in a set. ....'

The expectation of a candidate at the C1 level is that they 'can read quickly enough to cope with an academic course', i.e. they can cope with expeditious global as well as careful global reading demands, adapting their reading style to meet different reading purposes. The limited number of items of this type at CAE and FCE requires some attention.
In the expeditious reading task although the candidate has to search for the information rather than directly match through scanning, nearly all the answers are contained in single sentences. At this level, in an examination, which can be used for university entrance purposes, the ability to locate global information expeditiously might be expected (see Weir, Devi, Green, Hawkey, Maniski, and Zergarac 2006).

Task description

Total of 34 questions. The test paper has 4 parts.
1. Multiple Choice (6 items): three texts on one theme. Each text has two 4-option MCQs.
2. Gapped Text (6 items): a text from which paragraphs have been removed and placed in a jumbled order after the text along with one distractor. Candidates must decide from where in the text the paragraphs have been removed.
3. Multiple Choice (7 items): a text followed by 4-option MCQs.
4. Multiple Matching (15 items): a text or several short texts preceded by statements or questions. Candidates must match prompts to elements in the text.

Timing

34 items administered in 75 minutes

Weighting

Differential weighting: parts 1 to 3 are weighted differently from part 4. Each question in the former receives two marks whereas in the latter one mark is awarded per question.

Cognitive processing
Goal setting

Expeditious reading: local (scan/search for specifics)	**Careful reading: local** (understanding sentence)
Expeditious reading: global (skim for gist/search for key ideas/detail)	**Careful reading: global** (comprehend main idea(s)/overall text(s))

Part 1: Mostly integration of new information involved. A number of items require inferencing within and between sentences.
Part 2: Candidates benefit from reading the whole of the text to see how it fits together. Involves integration of new information as well as establishing a text level representation. It is more difficult to understand the whole text and how its parts interrelate, than to just understand individual main ideas; rereading is likely to be required to complete the task. This example also involves reading to understand unknown information. At CAE/CPE, the ALTE specification for C1 upwards includes academic reading ability, hence, texts need to contain new information in common with all texts used at university.

	Part 3: Tests candidates' detailed understanding of a text. Candidates may be able to answer some items by using propositional information at the sentence level only but others require the integration of information especially those targeted at the opinions and attitudes expressed in a text. A number of items require Inferencing. **Part 4:** Tests candidates' ability to locate specific information in a text. Information to answer the question is usually located in a single sentence, so candidates only need to establish sentence level propositional meaning with the exception of a few items. In terms of cognitive processing, the activities at CAE are in many parts similar to those at FCE. Both have a large number of items that require the reader to integrate information across sentences.
Cognitive processing Levels of reading	Word recognition
	Lexical access
	Syntactic parsing
	Establishing propositional meaning (cl./sent. level)
	Inferencing
	Building a mental model
	Creating a text level representation (disc. structure)
	Creating an intertextual representation (multi-text)

Features of the Input Text					
Word count	Six different passages to read, each with accompanying questions. Officially candidates have to read 3000 words in total. Maximum no of words for any single passage = 1100.				
Average sentence length	20.42 (This is an average figure. Individual sentences will span a range above and below the average). Range 13.97 – 29.08				
Domain	Social	Work	**Academic**		
Discourse mode	**Descriptive**	**Historical/ biographical**	**Expository**	**Argumentative**	**Instructive**
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete	**Fairly abstract**	Mainly abstract	
Presentation	**Verbal**	Non-verbal (i.e. graphs)	**Both**		
Lexical Level; further criteria	Broad range of vocabulary including idiomatic expressions and colloquialisms as well as language relating to opinion, persuasion and ideas some of which may be unknown to the candidate or less familiar. There are also a number of quite low frequency words. TTR (59 95) 75 Lexical density .52 A number of functions appear for the first time at this level; hypothesising, prioritising and summarising, for example, are functions that are not tested at lower levels. This is because these functions are most typical of text in an academic or professional context and so are typically found in some of the genres new at this level of the Main Suite. Clearly, many of the other functions which are tested in the CAE Reading Paper are also tested at lower levels. As far as the functional dimension is concerned, the difference at CAE level is that – once again – candidates are expected to cope with a richer variety of functional exponent. They are also expected to be developing the ability to appreciate a wider range of style and register and the ways in which functions are expressed in texts of different types. It is at CAE and CPE levels where an abstract idea may be a main focus for a text. Some would argue that this adds a layer of difficulty that may be more than purely linguistic but this is appropriate at CAE and CPE level given the fact that these exams are used as evidence of sufficient language proficiency to enter higher education and given the descriptors of C1 and C2 reading levels in the CEFR. CAE candidates are expected to be able to deal with topics that are more specialised and less personal than those that tend to feature at lower levels. The step up to CAE also involves coping with lexically and conceptually challenging treatment of topic areas. The cumulative coverage reaches 90.50% at the K3 level. 78.67 K1, 8.53 K2, 3.50 K3. 2.55%>15K (UG mean 4.3) Academic Word level 4.58 (undergraduate mean 10.5) Mean number of modifiers per noun phrase 0.92 Mean number of words before main verb in main clause 4.4				
Grammatical level	C1 Grammatical exponents. CEFR : The description in the Common European Framework of Reference				

	(CEFR) level for C1 and C2 is *many complex sentences* This level is typified by:many complex sentences; frequent use of modals; some use of ellipsis; complex approaches to referencing; use of synonymy.
Readability	Flesch-Kincaid Grade Level 9.6. Range 5.7 – 16. The low estimates for some texts at CAE might be a cause for some concern as according to the CEFR students at these levels should be capable of processing undergraduate level texts which Weir et al (2006) found to be at around 13.5 in terms of Flesch Kincaid estimates.
Topic	In Cambridge examinations, the sine qua non is that candidates should have sufficient knowledge of the topic to allow normal processing to take place. This means avoiding arcane topics where candidates would be unlikely to possess the schemata necessary for processing them on a global level. In the writing of tasks, Cambridge considers the following issues: • Are the topics broadly appropriate for the level of candidature from most cultures, experiences and age groups? • Is there any cultural/UK bias (urban/rural, boy/girl, etc.), i.e. does any task favour a candidate of a particular background, age or gender? • Have any cultural knowledge assumptions been made in the treatment of the topic? • Are potentially distressing topics avoided? • Is each topic likely to appeal to a broad base of candidates? • Will any topic 'date' too quickly? • Is there a good range of options in terms of topic and functions? • Are the topics likely to produce answers of the appropriate level and of the required length for the particular candidature, i.e. not too easy or too difficult? • Will test takers have an existing schema, i.e. organised mental framework, to be able to access the topic? Topics are chosen to be accessible and of interest to the broad range of candidates and are not intended to exclude any large group in terms of their standpoint or assumptions. Care is taken to ensure that candidates come to the text equally, no matter where they are from, how old they are, what their background is, and so on (as far as it is possible to control for this). Clearly, it is impossible to interest everybody, but subjects which appeal only to a minority are avoided. The topic should not be biased in favour of any particular section of the test population. Material must not favour candidates with specialised knowledge of a particular subject or have content that would be too specialised or technical for the majority of candidates.
Rhetorical organization	**Explicit** | | | | **Not explicit**
Text genre	There are three reading texts with a variety of question types. Texts are appropriate to, and accessible to, test-takers entering undergraduate or postgraduate courses, or seeking professional registration. CAE candidates are expected to be able to deal with both short and long texts, from the following sources: newspapers and magazines (e.g., articles, reports); journals; books (fiction and non-fiction extracts); promotional and informational materials (e.g., guides, manuals). Referential, poetic, emotive, conative purposes for CAE texts. At this level, texts with overall conative purpose also used at. This coincides with the task purposes as well as the sources and topics used. For example, in CAE, a journal article on social trends where the author is presenting an argument for or against a particular trend may be suitable to assess candidates' understanding of attitude/opinion, text organisation, main idea, or implications.
Intended Writer/Reader relationship	In Cambridge practice, Main Suite exam texts are aimed at the general reader. At higher levels (CAE and CPE) texts assume a certain level of education and understanding of the world as well as of English.
	Reliability *[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*
Kuder-Richardson formula 20 (KR20)[3]	0.8 – 0.85 (34 items as against 40 items in IELTS)

[3] A measure of internal consistency used to estimate reading test reliability.

TEST:	Pearson Test of English Academic (PTE-A)	COMPONENT:	Reading

Task Features	
Skills focus	PTE Academic "is designed to be representative of academic language demands within a university or college setting". (p2) It includes a wide range of texts from different disciplines, including arts, sciences, humanities, economics and business, and "tests your ability to comprehend information provided in a given text". Texts used are taken from published sources. Skills tested include the following: - identifying the topic - identifying the main ideas - evaluating and synthesizing information - classifying and categorizing information
Task description	There are five sections in the reading paper, which focus on different question types, as follows: -Multiple choice, choosing a single answer -Multiple choice, choosing multiple answers -Re-order paragraphs -Reading: fill in the blanks -Reading and writing: Fill in the blanks The last section allows for limited writing, involving writing words from a series of options. There are "15-20 independent and integrated skills items, depending on the items in a given test" (Official guide, p. 7)
Timing	The whole test is timed to last 32-41 minutes in total. Individual tasks are not timed. A timer appears on screen to inform the candidate of time remaining.
Weighting	The maximum possible score for each section is not specified explicitly. In some sections candidates can lose marks for an incorrect answer, in what is termed 'partial credit scoring'. Marks for the reading section as a whole can range from 10-90. PTE Academic scores are aligned to the CEF levels, with 86+ equating to C2, approximately (according to the preliminary alignment scale, p15)

Cognitive processing **Goal setting**		
	Expeditious reading: local (scan/search for specifics)	**Careful reading: local** (understanding sentence)
	Expeditious reading: global (skim for gist/search for key ideas/detail)	**Careful reading: global** (comprehend main idea(s)/overall text(s))

Cognitive processing **Levels of reading**	Word recognition
	Lexical access
	Syntactic parsing
	Establishing propositional meaning (cl./sent. level)
	Inferencing
	Building a mental model
	Creating a text level representation (disc. structure)
	Creating an intertextual representation (multi-text)

Features of the Input Text					
Word count	There are five different passages to read, each with accompanying questions, a total of 1130 words maximum. The longest text can be up to 300 words (section 1, 2, 5) while the shortest is a maximum of 80 (section 4)				
Average sentence length	19.59 (Tokens)				
Domain	**Social**		Work		**Academic**
Discourse mode	**Descriptive**	**Historical/biographical**	**Expository**	Argumentative	**Instructive**
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	**Mostly concrete**		**Fairly abstract**	Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		**Both**
Lexical Level; further criteria	The cumulative coverage reaches 90.11% at the K3 level. 70.3% K1, 11.90% K2, 7.88% K3. 5.11%>15K (UG mean 4.3) TTR 51% Lexical Density 0.58				

	Mean number of modifiers per noun phrase 0.99 Mean number of words before main verb in main clause 4.03				
Grammatical level	(In terms of the CEFR) B2				
Readability	Flesch-Kincaid Grade 11.80				
Topic	Topics include a wide range from different disciplines, including arts, sciences, humanities, economics and business.				
Rhetorical organization	**Explicit**				**Not explicit**
Text genre	There are five reading texts with a variety of question types. Texts are of academic interest or are related to student life. They do not require prior knowledge of the topic. Texts used are taken from published sources, such as textbooks or websites containing information for academic audiences.				
Intended Writer/Reader relationship	The relationship is not specified. The texts are normally written for an informed academic audience, but not a specialist one.				
Reliability *[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*					
Kuder-Richardson formula 20 (KR20)[4]	0.92 (Pearson, 2014)				

CONFIDENTIAL

[4] A measure of internal consistency used to estimate reading test reliability.

TEST:	Internet-based Test of English as a Foreign Language (TOEFL iBT)	COMPONENT:	Reading

Task Features	
Skills focus	The reading section of the TOEFL test "measures your ability to understand university-level academic texts and passages". Specifically, it focuses on three main purposes: - Reading to find information - Basic comprehension - Reading to Learn
Task description	There are 3-4 passages per test, each of which has 12-14 accompanying questions – giving a total of 36-56 questions per paper. Reading questions cover Basic Information skills, Inferencing skills, and Reading to Learn skills. There are 10 question types, summarised as follows: Basic Information and Inferencing questions (12 to 14 questions per set) 1. Factual Information questions (3 to 6 questions per set) 2. Negative Factual Information questions (0 to 2 questions per set) 3. Inference questions (1 to 3 questions per set) 4. Rhetorical Purpose questions (1 to 2 questions per set) 5. Vocabulary questions (3 to 6 questions per set) 6. Reference questions (0 to 2 questions per set) 7. Sentence Simplification questions (0 or 1 question per set) 8. Insert Text question (1 question per set) Reading to Learn questions (1 per set) 9. Prose Summary 10. Fill in a Table
Timing	The whole reading test is timed to last 60-80 minutes in total.
Weighting	Each of the 3-4 passages has between 12-14 questions. All questions are worth 1point, except for the final question(s) for each passage, which can be worth 2 points. Raw scores (up to a maximum of 42) are then converted into scaled scores using a grid, to reduce the maximum score to 30.
Cognitive processing Goal setting	**Expeditious reading: local** (scan/search for specifics) / **Careful reading: local** (understanding sentence) **Expeditious reading: global** (skim for gist/search for key ideas/detail) / **Careful reading: global** (comprehend main idea(s)/overall text(s))
Cognitive processing Levels of reading	Word recognition Lexical access Syntactic parsing **Establishing propositional meaning (cl./sent. level)** **Inferencing** **Building a mental model** **Creating a text level representation (disc. structure)** Creating an intertextual representation (multi-text)

Features of the Input Text					
Word count	There are 3-4 different passages to read, each with accompanying questions, a total of around 2800 words.				
Average sentence length	***19.26*** (Tokens)				
Domain	Social		Work		**Academic**
Discourse mode	Descriptive	**Historical/biographical**	**Expository**	**Argumentative**	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	**Mostly concrete**	**Fairly abstract**		Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		**Both**
Lexical Level; further criteria	The cumulative coverage reaches 89.10% at the K3 level. (67.96= K1, 13.06= K2, 8.08= K3). 3.33%>15K (UG mean 4.3) Academic Word level 7.84 (undergraduate mean 10.5)				

	TTR. 35% Lexical Density 0.59 Mean number of modifiers per noun phrase 1.031 Mean number of words before main verb in main clause 5.689				
Grammatical level	(In terms of the CEFR) C1				
Readability	Flesch-Kincaid Grade 12.07				
Topic	Topics include a wide range from different disciplines. The four main ones listed are Arts, Life Science, Physical Science, Social Science.				
Rhetorical organization	**Explicit**				**Not explicit**
Text genre	Texts used are "excerpts from college-level textbooks".				
Intended Writer/Reader relationship	Since the texts are from textbooks, they are aimed an informed academic audience, but not a specialist one.				
Reliability *[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*					
Kuder-Richardson formula 20 (KR20)[5]	0.85 (ETS, 2011)				

CONFIDENTIAL

[5] A measure of internal consistency used to estimate reading test reliability.

TEST:	Occupational English Test (OET)	COMPONENT:	Reading

Task Features	
Skills focus	The reading section of the OET test assesses candidates' ability to source information from multiple texts, to synthesise information in a meaningful way and to 'skim' and 'scan' material to retrieve information quickly (Part A). Detailed descriptors are as follows: • locate specific information in a range of short texts. • understand the conventions of different text types. • synthesise information from different sources. • summarise information for a non-medical audience. • recognise paraphrasing. • use appropriate spelling and word forms. It also assesses candidates' ability to read and understand comprehensive texts on health-related topics similar to those in academic or professional journals. (Part B) Detailed descriptors are as follows: • understand main ideas. • differentiate main ideas from supporting information. • identify the underlying theme of a paragraph or text. • recognise and understand paraphrase. • understand cohesion between parts of a text. • follow a complex argument that is made over several paragraphs.
Task description	The OET Reading sub-test consists of a short answer response (summary completion) test (Part A) and a multiple-choice questions (MCQ) test (Part B). Part A of the Reading sub-test requires completion of a summary by writing the missing word or several words (but only up to three words) within a strict time limit of 15 minutes. This requires test-takers to skim and scan 3-4 short texts (a total of about 650 words) related to a single topic and to complete a summary paragraph by filling in the missing words. Test-takers are required to read the texts and write responses for 25-35 gaps in total, within a strictly monitored time limit of 15 minutes. In Part B, test-takers are required to read two passages (600-800 words each) on general medical topics and answer 8-10 multiple choice questions for each text (a total of 16-20 questions) within a time limit of 45 minutes.
Timing	60 minutes in total (Part A = 15 minutes. Part B = 45 minutes)
Weighting	Each section of the Reading sub-test is weighted equally. That is, Part A is worth 33.3% and each text in Part B is also worth 33% of the total score. Weighting operates as follows: *"When you sit the reading test, you will get a raw score on Part A (between 25 and 35 depending on the number of items on that particular test) and a raw score on Part B (between 16 and 20 depending on the number of items on that particular test). Your scores will then be converted to a total score for the reading test in which your score for Part A counts for 33.33% (one-third) of the total score, and your score for Part B counts for 66.66% (two-thirds) of the total score. Your total score for the reading test will then be transformed into a grade (A - E)."*

Cognitive processing Goal setting	**Expeditious reading: local** (scan/search for specifics)	**Careful reading: local** (understanding sentence)
	Expeditious reading: global (skim for gist/search for key ideas/detail)	**Careful reading: global** (comprehend main idea(s)/overall text(s))

Cognitive processing Levels of reading	Word recognition
	Lexical access
	Syntactic parsing
	Establishing propositional meaning (cl./sent. level)
	Inferencing
	Building a mental model
	Creating a text level representation (disc. structure)
	Creating an intertextual representation (multi-text)

Features of the Input Text	
Word count	Part A = 3-4 short texts (a total of about 650 words) + Part B = two passages (600-800 words each). **TOTAL of 1250-1450 words**

Average sentence length	***23.52*** (Tokens)				
Domain	Social		Work		Academic
Discourse mode	Descriptive	Historical/bio graphical	Expository	Argumentative	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete		Fairly abstract	Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	The cumulative coverage reaches 90.69% at the K3 level . (68.6%= K1, 13.57%= K2, 8.52%= K3). 2.62%>15K (UG mean 4.3) Academic Word level 8.74 (undergraduate mean 10.5) TTR 32% Lexical Density 0.58 Mean number of modifiers per noun phrase 0.89 Mean number of words before main verb in main clause 2.163				
Grammatical level	(In terms of the CEFR) C1 Grammatical exponents. CEFR : The description in the Common European Framework of Reference (CEFR) level for C1 and C2 is *many complex sentences*				
Readability	Flesch-Kincaid Grade 14.27				
Topic	The topics of the OET Reading sub-tests are chosen to be accessible to all candidates of the OET. They are of general medical interest and cover a wide variety of issues. Reading texts are selected for their concentration of content of a general health literacy nature, elements that can be considered common across multiple professions.				
Rhetorical organization	**Explicit**				**Not explicit**
Text genre	Texts used are "selected for their concentration of content of a general health literacy nature; elements that can be considered common across multiple professions."				
Intended Writer/Reader relationship	Texts are aimed an informed general health audience, but not a specialist one.				
Reliability *[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*					
Kuder-Richardson formula 20 (KR20)[6]	0.85				

[6] A measure of internal consistency used to estimate reading test reliability.

Test	GEPT Advanced	Component	Reading

Task Features

Skills focus	The GEPT (General English Proficiency Test) is a test of advanced general English language proficiency, developed by LTTC in Taiwan. It offers test content not confined to any specific discipline or topic. In terms of reading it attempts: • to assess whether candidates can grasp the main ideas and important details: get a general sense of a passage; understand contextual features and structure; follow the development of an argument; recognise a writer's opinions, attitudes and purpose • and make inferences through careful reading • To assess whether candidates can quickly grasp the main ideas of the texts and find specific information by scanning them In particular, the exam handbook requires that a candidate at this level - can read written English from different sources, including magazines, newspapers, and professional periodicals - can employ different reading strategies according to various reading purposes To test students' ability to comprehend academic texts and to extract important information from those texts.
Task description	There are 20 questions in Part 1 Careful Reading and 20 questions in Part 2 Expeditious Reading; Parts 1 and 2 are timed separately. A variety of question types is used, chosen from the following types: multiple choice, matching information, matching headings, summary completion, sentence insertion, and short-answer questions. Mainly receptive, some limited writing involved in short-answer questions, but only brief answers are required.
Further task focus information	Part 1 consists of four texts. Each of the first three texts is followed by several related short-answer and multiple-choice questions After reading the fourth text, candidates are asked to fill in gaps in a summary passage. Part 2 consists of three sections. In each of the first two sections, candidates are given a bank of possible headings from which to choose the appropriate heading for each paragraph of the text. The third section consists of three short passages. Candidates scan them to find the answers to eight questions.
Timing	50 minutes to answer a total of 20 questions on 4 passages in Part 1 Careful Reading. 20 minutes to answer a total of 20 questions on 3 tasks in Part 2 Expeditious Reading.
Weighting	All items equally weighted.
Cognitive processing Goal setting	Expeditious reading: local (scan/search for specifics) / Careful reading: local (understanding sentence) Expeditious reading: global (skim for gist/search for key ideas/detail) / Careful reading: global (comprehend main idea(s)/overall text(s))
Cognitive processing Levels of reading	Word recognition Lexical access Syntactic parsing Establishing propositional meaning (cl./sent. level) Inferencing Building a mental model Creating a text level representation (disc. structure) Creating an intertextual representation (multi-text)

Features of the Input Text

Word count	Four different passages, each with accompanying questions, to read in Part 1, approximately 2,800-3,000 words in total. Three different tasks, approximately 2,500 words in total, to complete in Part 2, each with accompanying questions. A total of around 5500 words. Each text is an average of around 700 words in length.
Average sentence length	20.997 (This is an average figure. Individual sentences will span a range above and below the average) http://events.cambridgeenglish.org/alte-2014/docs/presentations/alte2014-jessica-wu.pdf

Domain	Social	Work	Academic	
Discourse mode	**Descriptive/Narrative**	Historical/biographical	Expository	Argumentative
Content knowledge	**General**			Specific
Cultural specificity	**Neutral**			Specific
Nature of information	**Only concrete**	Mostly concrete	Fairly abstract	Mainly abstract
Presentation	**Verbal**	Non-verbal (i.e. graphs)	Both	
Lexical Level	The cumulative coverage reaches 86% at the K3 level . 72.13% K1, 6.07% K2, 8.07% K3. 0.06%>15K Academic Word level 8.09% (undergraduate mean 10.5%) .			

	TTR .35 Lexical Density 0.59 Mean number of modifiers per noun phrase 0.98 Mean number of words before main verb in main clause 5.517				
Grammatical level	C1 Grammatical exponents. CEFR : The description in the Common European Framework of Reference (CEFR) level for C1 and C2 is *many complex sentences*				
Readability	Flesch-Kincaid Grade Level 12 (12.007) Flesch reading ease 43.31 http://events.cambridgeenglish.org/alte-2014/docs/presentations/alte2014-jessica-wu.pdf				
Topic	A broad range of subject areas; arts, humanities, social sciences, engineering, technology, and life sciences are the most common topic areas for the test.				
Rhetorical organization	**Explicit**				Not explicit
Text genre	There are nine reading texts with a variety of question types. Texts are appropriate to, and accessible to, learners whose English proficiency is at CEFR B2 to C level. The kinds of text used in the GEPT Advanced Reading test include self- contained reports on developments in science/technology and historical/biographical narratives. Texts come from books, journals, magazines, newspapers, reports and online resources, written for a non-specialist audience.				
Intended Writer/Reader relationship	The relationship is not specified. The texts will typically be written for a general audience, not a specific reader.				
Reliability *[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*					
KR20	0.87 to 0.91 https://www.lttc.ntu.edu.tw/E_LTTC/E_GEPT/files/2014_GEPT_Information_Brochure.pdf				

CONFIDENTIAL

Test	Michigan ECPE	Component	Reading

Task Features

Skills focus

The Examination for the Certificate of Proficiency in English (ECPE) is an advanced (C2) academic EFL exam. The language used in the test is American-English. It is a test of advanced English language proficiency, reflecting skills and content typically used in university or professional contexts.

In terms of Reading, the ECPE focuses on three subgroups of reading skills: global, local, and inferential reading skills and include (among others):

Global	• understanding main idea • identifying speaker's purpose • synthesizing ideas from different parts of the text
Local	• identifying supporting detail • understanding vocabulary • synthesizing details • recognizing restatement
Inferential	• understanding rhetorical function • making an inference • inferring supporting detail • understanding pragmatic implications

Task description	There are four sections (texts) in the reading paper, each accompanied by 5 multiple-choice items, giving a total of 20 questions. Each ECPE reading text consists of "a self-contained passage that could be found in any formal written context". Each is 250 to 400 words in length.
Further task focus information	-
Timing	The Reading test is part of a larger section including Grammar, Cloze and Vocabulary as well as reading. The whole of the section is timed to last 75 minutes.
Weighting	The whole GCVR (Grammar, Cloze, Vocabulary, Reading) section is scored automatically by computer. There is only one correct answer for each item. Correct answers receive one point. Incorrect answers receive zero. There are a total of 120 items in the whole GCVR section: 40 for Reading, 40 for Grammar, 20 for Cloze, and 40 for Vocabulary. Reading therefore makes up around ***29%*** of the total for this Section. In the ECPE as a whole, the GCVR section is one of four, equally weighted (the others being Speaking, Listening and Writing). This means that the Reading section in itself makes up around ***7%*** of the weighting of the whole ECPE test.

Cognitive processing Goal setting		
	Expeditious reading: local (scan/search for specifics)	**Careful reading: local** (understanding sentence)
	Expeditious reading: global (skim for gist/search for key ideas/detail)	**Careful reading: global** (comprehend main idea(s)/overall text(s))

Cognitive processing Levels of reading	
	Word recognition
	Lexical access
	Syntactic parsing
	Establishing propositional meaning (cl./sent. level)
	Inferencing
	Building a mental model
	Creating a text level representation (disc. structure)
	Creating an intertextual representation (multi-text)

Features of the Input Text

Word count	There are four different passages to read, each of 250-400 words, with accompanying questions, a total of 1000-1600 words.				
Average sentence length	19.88 (Tokens)				
Domain	Social		Work		academic
Discourse mode	Descriptive	Historical/biographical	Expository	Argumentative	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete		Fairly abstract	Mainly abstract

<table>
<tr><td>Presentation</td><td>Verbal</td><td>Non-verbal (i.e. graphs)</td><td>Both</td></tr>
<tr><td>Lexical Level</td><td colspan="3">The cumulative coverage reaches 88.36% at the K3 level. (69.24 at K1, 12.07 at K2, 7.05 at K3). 3.2%>15K (UG mean 4.3)
AWL = 7.11%

TTR . 42
Lexical Density 0.58
Mean number of modifiers per noun phrase 0.814
Mean number of words before main verb in main clause 1.178</td></tr>
<tr><td>Grammatical level</td><td colspan="3">(In terms of the CEFR) C2</td></tr>
<tr><td>Readability</td><td colspan="3">Flesch Reading Ease 44.69
Flesch-Kincaid Grade 11.97
Gunning Fog 14.02</td></tr>
<tr><td>Topic</td><td colspan="3">The ECPE focuses on various topics used across the public, occupational, and educational domains as follows:
Public — public spaces (street, shops, restaurants, sports, or entertainment venues) and other social networks outside the home
Occupational — workplace settings (offices, workshops, conferences), etc.
Educational — schools, colleges, classrooms, residence halls, etc.

Test takers “should require no specialized knowledge or experience to understand the content of the items or prompts. Topics should be equally accessible to a range of ages and should represent a variety of opinions.” (Source: ECPE website)</td></tr>
<tr><td>Rhetorical organization</td><td>Explicit</td><td></td><td></td><td>Not explicit</td></tr>
<tr><td>Text genre</td><td colspan="3">General academic, relatively formal</td></tr>
<tr><td>Intended Writer/Reader relationship</td><td colspan="3">The texts are normally written for an informed academic audience, in a formal setting, but not a specialist one., “a self-contained passage that could be found in any formal written context. “</td></tr>
<tr><td colspan="4">Reliability
[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates’ scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach’s alpha coefficient.]</td></tr>
<tr><td>KR20</td><td colspan="3">0.92</td></tr>
</table>

CONFIDENTIAL

<table>
<tr><td>Test</td><td>City and Guilds International ESOL C1 level</td><td>Component</td><td>Reading</td></tr>
<tr><td colspan="4">Task Features</td></tr>
<tr><td>Skills focus</td><td colspan="3">The City and Guilds International ESOL C1 level is an advanced (C1) ESOL exam. It is a test of advanced English language proficiency, reflecting skills and content typically used in university or professional contexts.
In terms of reading, the examination draws on CEFR descriptors, specifically:
• Can understand a wide range of demanding longer texts, both written and spoken, and recognise implicit meaning.
• Can respond to extended information and narratives, follow detailed explanations and complex instructions, adapting response to audience, medium and context.
• Can use language flexibly and understand a range of texts of varying complexity and length for social, academic and professional purposes.

In particular, the exam handbook requires a candidate at this level to
• understand long complex texts, appreciating distinctions in style and purpose
• understand texts from authentic sources, identifying context, content and style
• understand the main ideas of complex texts on a wide range of texts on a wide range of professional, academic and social topics
• gather specific information from different parts of a text or from different texts
• understand descriptions and narratives in which language is used to create different sophisticated effects
• understand lengthy texts containing complex Instructions or explanations Including technical or specialist texts such as instruction manuals and articles
• understand texts concerned with contemporary issues In which the writers adopt particular viewpoints, identifying finer points of attitude and implied, as well as stated, opinions.</td></tr>
<tr><td>Task description</td><td colspan="3">There are four sections in the reading paper, with a total of 7 texts to read. The tasks include: Selection of correct statements about the text, Choose a sentence to fill gaps in the text, Multiple matching to identify information from the texts, MC question (focus on best summary), MC question (focus on text structure), MC question (specific information).</td></tr>
<tr><td>Further task focus information</td><td colspan="3">-</td></tr>
<tr><td>Timing</td><td colspan="3">The total time allocated to Reading and writing is 2 hours 40 minutes. There is no stricture provided beyond that as to how much is allocated to Reading specifically.</td></tr>
<tr><td>Weighting</td><td colspan="3">The exam offers 30 marks for reading, with mark distribution and boundaries as follows:
<table>
<tr><td colspan="4">Marks for Reading</td><td></td></tr>
<tr><td>Part 1</td><td>Part 2</td><td>Part 3</td><td>Part 4</td><td>Total</td></tr>
<tr><td>5</td><td>6</td><td>9</td><td>10</td><td>30</td></tr>
<tr><td colspan="4">First Class Pass</td><td>23-30</td></tr>
<tr><td colspan="4">Pass</td><td>15-22</td></tr>
<tr><td colspan="4">Narrow Fail</td><td>13-14</td></tr>
<tr><td colspan="4">Fail</td><td>0-12</td></tr>
</table>
Listening is allocated 30 marks and writing 24, meaning that Reading is allocated around 36% of the total marks for the paper.</td></tr>
<tr><td rowspan="2">Cognitive processing
Goal setting</td><td colspan="2">Expeditious reading: local
(scan/search for specifics)</td><td>Careful reading: local
(understanding sentence)</td></tr>
<tr><td colspan="2">Expeditious reading: global
(skim for gist/search for key ideas/detail)</td><td>Careful reading: global
(comprehend main idea(s)/overall text(s))</td></tr>
<tr><td rowspan="8">Cognitive processing
Levels of reading</td><td colspan="3">Word recognition</td></tr>
<tr><td colspan="3">Lexical access</td></tr>
<tr><td colspan="3">Syntactic parsing</td></tr>
<tr><td colspan="3">Establishing propositional meaning (cl./sent. level)</td></tr>
<tr><td colspan="3">Inferencing</td></tr>
<tr><td colspan="3">Building a mental model</td></tr>
<tr><td colspan="3">Creating a text level representation (disc. structure)</td></tr>
<tr><td colspan="3">Creating an intertextual representation (multi-text)</td></tr>
</table>

Features of the Input Text					
Word count	There are four different sections. Three of them offer longer texts, each of 200-350 words. The fourth offers four short texts totaling around 600 words. The average quantity of reading per section is therefore c. 300 words, and the total word count for all four sections is around **1200 words.**				
Average sentence length	18.66 (Tokens)				
Domain	Social		Work		academic
Discourse mode	Descriptive	Historical/biographical	Expository	Argumentative	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete		Fairly abstract	Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		Both
Lexical Level	The cumulative coverage reaches 89.75% at the K3 level. (77.04 at K1, 8.39 at K2, 4.32 at K3). 4.41%>15K (UG mean 4.3) Academic word list coverage is 4.19% TTR . 44 Lexical Density 0.58 Mean number of modifiers per noun phrase 0.893 Mean number of words before main verb in main clause 4.573				
Grammatical level	(In terms of the CEFR) C1				
Readability	Flesch Reading Ease 55.11 Flesch-Kincaid Grade 10.74 Gunning Fog 13.58				
Topic	The exam focuses on various topics used across the public, occupational, and educational domains, including shopping, science and society, environment and so on.				
Rhetorical organization	**Explicit**				**Not explicit**
Text genre	General, non-academic				
Intended Writer/Reader relationship	The texts are normally written for an informed general audience.				
Reliability *[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*					
Cronbach's alpha	0.752				

CONFIDENTIAL

<table>
<tr><td>Test</td><td>City and Guilds International ESOL C2 level</td><td>Component</td><td>Reading</td></tr>
<tr><td colspan="4">Task Features</td></tr>
<tr><td>Skills focus</td><td colspan="3">The City and Guilds International ESOL C2 level is an advanced (C2) ESOL exam. It is a test of advanced English language proficiency, reflecting skills and content typically used in university or professional contexts.
In terms of reading, the examination draws on CEFR descriptors at C1 level, specifically:
• Can understand a wide range of demanding longer texts, both written and spoken, and recognise implicit meaning.
• Can respond to extended information and narratives, follow detailed explanations and complex instructions, adapting response to audience, medium and context.
• Can use language flexibly and understand a range of texts of varying complexity and length for social, academic and professional purposes.

It also expects the following, at C2 level:
• Can understand with ease virtually everything heard or read.

In particular, the exam handbook requires a candidate at this level to
• understand with ease virtually all types of authentic written texts of different purposes and style and those dense in complex structures
• understand the main ideas of all forms of written language, including lengthy abstract, structurally or linguistically complex texts or highly idiomatic, literary and non-literary writing on a wide range of professional, academic and social topics
• gather specific information from different parts of a text or from different texts
• understand descriptions and narratives in which language is used to create different sophisticated and subtle effects
• understand lengthy texts containing complex instructions or explanations on specialist subjects
• understand texts concerned with contemporary issues in which the writers adopt particular viewpoints, identifying finer points of attitude and implied opinions, and follow complex lines of reasoning.</td></tr>
<tr><td>Task description</td><td colspan="3">There are four sections in the reading paper, with a total of 7 texts to read. The tasks include: Selection of correct statements about the text, Choose a sentence to fill gaps in the text, Multiple matching to identify information from the texts, a set of ‘wh’ questions to be answered in a few words</td></tr>
<tr><td>Further task focus information</td><td colspan="3">-</td></tr>
<tr><td>Timing</td><td colspan="3">The total time allocated to Reading and writing is 2 hours 40 minutes. There is no stricture provided beyond that as to how much is allocated to Reading specifically.</td></tr>
<tr><td>Weighting</td><td colspan="3">The exam offers 30 marks for reading, with mark distribution and boundaries as follows:
<table>
<tr><td colspan="4">Marks for Reading</td><td></td></tr>
<tr><td>Part 1</td><td>Part 2</td><td>Part 3</td><td>Part 4</td><td>Total</td></tr>
<tr><td>5</td><td>6</td><td>9</td><td>10</td><td>30</td></tr>
<tr><td colspan="4">First Class Pass</td><td>23-30</td></tr>
<tr><td colspan="4">Pass</td><td>15-22</td></tr>
<tr><td colspan="4">Narrow Fail</td><td>13-14</td></tr>
<tr><td colspan="4">Fail</td><td>0-12</td></tr>
</table>
Listening is allocated 30 marks and writing 24, meaning that Reading is allocated around 36% of the total marks for the paper.</td></tr>
<tr><td rowspan="2">Cognitive processing
Goal setting</td><td colspan="2">Expeditious reading: local
(scan/search for specifics)</td><td>Careful reading: local
(understanding sentence)</td></tr>
<tr><td colspan="2">Expeditious reading: global
(skim for gist/search for key ideas/detail)</td><td>Careful reading: global
(comprehend main idea(s)/overall text(s))</td></tr>
<tr><td rowspan="3">Cognitive processing
Levels of reading</td><td colspan="3">Word recognition</td></tr>
<tr><td colspan="3">Lexical access</td></tr>
<tr><td colspan="3">Syntactic parsing</td></tr>
</table>

	Establishing propositional meaning (cl./sent. level)				
	Inferencing				
	Building a mental model				
	Creating a text level representation (disc. structure)				
	Creating an intertextual representation (multi-text)				
Features of the Input Text					
Word count	There are four different sections. Two of them offer longer texts each of around 300 words. The third offers four short texts totaling around 600 words. The fourth offer a longer text of c. 600 words The average quantity of reading per section is therefore c. 450 words, and the total word count for all four sections is around **1800 words.**				
Average sentence length	20.27 (Tokens)				
Domain	Social		Work		academic
Discourse mode	Descriptive	Historical/biographical	Expository	Argumentative	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete		Fairly abstract	Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		Both
Lexical Level	The cumulative coverage reaches 88.9% at the K3 level. (71.5% at K1, 9.64% at K2, 7.76% at K3). 6.09%>15K (UG mean 4.3) Academic word list coverage is 6.27% TTR . 43 Lexical Density 0.55 Mean number of modifiers per noun phrase 1.01 Mean number of words before main verb in main clause 2.945				
Grammatical level	(In terms of the CEFR) C2				
Readability	Flesch Reading Ease 47.41 Flesch-Kincaid Grade 11.68 Gunning Fog 15.24				
Topic	The exam focuses on various topics used across the public, occupational, and educational domains, including shopping, science and society, environment and so on.				
Rhetorical organization	**Explicit**				**Not explicit**
Text genre	General, non-academic				
Intended Writer/Reader relationship	The texts are normally written for an informed general audience.				
Reliability *[An estimate of the reliability of a reading test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a reading test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the reading test as a whole. Commonly used statistics for computing and reporting reading test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*					
Cronbach's alpha	0.701				

Appendix 6 – Detailed analyses of the selected writing tests (IELTS, CPE, CAE, PTE-A, TOEFL iBT, OET, GEPT Advanced, ECPE, IESOL & ISESOL Level 2 and IESOL & ISESOL Level 3)

CONFIDENTIAL

TEST:	International English Language Testing System (IELTS)	COMPONENT:	Writing

Task Setting			
Number of tasks	Two (Task 1 and Task 2)		
Response format and Genre	Task 1: Continuous writing; Information transfer from multiple non-verbal inputs Task 2: A longer piece of continuous writing; Essay		
Task description	The IELTS Writing test is designed to assess a wide range of writing skills, including how well candidates: • write a response appropriately • organise ideas • use a range of vocabulary and grammar accurately In Task 1 candidates are presented with a graph, table, chart or diagram. Candidates are asked to describe, summarise or explain the information in their own words. This might involve describing and explaining data, describing the stages of a process or how something works, or describing an object or event. In Task 2 candidates are asked to write an essay in response to a point of view, argument or problem. Students should find the issues interesting and easy to understand.		
Source texts	N/A		
Domain	**Social**	Work	**Academic**
Topic (Content knowledge)	The issues raised are of general interest to, suitable for and easily understood by candidates entering undergraduate or postgraduate studies or seeking professional registration.		
Purpose	Task 1: • To transferring informing from multiple inputs • To collate different pieces of information in order to describe, summarise or explain the information Task 2: • To write a persuasive essay • To defend or attack a particular argument or opinion, compare or contrast aspects of an argument, and give reasons for the argument.		
Knowledge of criteria	Publish information about how the tasks are scored, including band descriptors for individual tasks, and procedures for scoring, are provided on the official website (http://takeielts.britishcouncil.org/prepare-test/how-prepare-ielts-test) as well as in a booklet *Information for candidates* (http://takeielts.britishcouncil.org/sites/default/files/Information_for_Candidates_booklet.pdf)		
Writer-reader relationship	Candidates are expected to write within a formal context and to demonstrate sensitivity to their audience. However, a particular audience is not always specified in either Task 1 or Task 2.		
Timing	The IELTS Writing test takes 60 minutes. Students should spend 20 minutes on Task 1, and 40 minutes on Task 2. Students need to manage your own time.		
Text length of expected response	Task 1: at least 150 words Task 2: at least 250 words		
Weighting	Task 2 is worth twice as much as Task 1 in the IELTS Writing test.		
Mode	**Paper-based**	Computer-based	
Skills assessed	In both tasks, candidates are assessed on their ability to write a response which is appropriate in terms of content, the organisation of ideas, and the accuracy and range of vocabulary and grammar.		

Cognitive processing

Task 1:

- Macro-planning: goal setting and task representation
- Transforming ideas from non-verbal inputs
- Organising ideas
- Translating
- Micro-planning
- Monitoring and revision

(Bridges, 2010)

- Comprehending non-graphic task instructions
- Comprehending (and interpreting) the components of graphs
- Re-presenting or re-producing the non-graphic and graphic information as continuous discourse in written form in English as a foreign language

(Yu, Rea-Dickins and Kiely, 2011)

Task 2:

- Macro-planning: goal setting and task representation
- Generating ideas
- Organising ideas
- Translating
- Micro-planning
- Monitoring and revision

(Weir, O'Sullivan, Jin and Bax, 2007)

Linguistic demands

Discourse mode (rhetorical task)	**Descriptive**	Narrative	**Expository**	**Argumentative/ persuasive**	Instructive/ Referral

Lexical Level

To achieve Band 7, candidates are expected to use a sufficient range of vocabulary to allow some flexibility and precision and use less common lexical items with some awareness of style and collocation.

Word frequency (calculates the percentage of words in the scripts at each band level that fell into the first 1000 (K1) and 2000 (K2) most frequently used English words, and those that were found in the Academic Word List (AWL))

Task 1

	L1 Chinese			L1 Spanish		
	K1	K2	AWL	K1	K2	AWL
Band 7	65.38	9.58	13.81	60.21	10.26	17.62
Band 8	-	-	-	65.54	7.57	15.60

Task 2

	L1 Chinese			L1 Spanish		
	K1	K2	AWL	K1	K2	AWL
Band 7	57.69	13.59	13.91	56.94	10.85	18.51
Band 8	-	-	-	70.61	7.27	13.48

(Banerjee, Franceschina & Smith, 2007)

Tokens (the total number of words) and **Types** (the total number of different word forms)

	Task 1 Means (SD)		Task 2 Means (SD)	
	Tokens	Types	Tokens	Types
Band 7	223.6 (51.8)	94.2 (20.6)	312.2 (56.8)	159.0 (25.6)
Band 8	230.6 (39.1)	102.6 (26.5)	323.3 (31.5)	160.4 (12.9)

(Banerjee, Franceschina & Smith, 2007)

CONFIDENTIAL

Ratio of types to tokens (This is a measure of lexical variation/diversity because the higher the number of types in relation to tokens the more varied/diverse the vocabulary used.)

	TTR	
	Task 1	Task 2
Band 7	42.9 (6.6)	51.3 (5.1)
Band 8	44.7 (9.0)	49.8 (3.5)

(Banerjee, Franceschina & Smith, 2007)

Lexical density (This measure compares the number of lexical words used to the number of grammatical words.)

	Lexical density	
	Task 1	Task 2
Band 7	0.82 (0.15)	0.75 (0.15)
Band 8	0.82 (0.21)	0.69 (0.13)

(Banerjee, Franceschina & Smith, 2007)

Syntactic complexity

To achieve Band 7, candidates are expected to use a variety of complex structures and produce frequent error-free sentences.

- **T-unit** (Independent clause plus all its associated dependent clauses; minimal terminable unit of text)
- **Main clause** (Clause containing a subject and a predicate which can stand on its own. (predicate = a phrase headed by a TENSED verb)
- **Dependent clause** (Dependent clauses cannot stand on their own, and can only be present when either embedded in or chained to a main clause).

	Mean clause/t-unit	
	Task 1	Task 2
Band 7	1.51	2.70
Band 8	2.07	2.44

	Mean dependent clause/clause	
	Task 1	Task 2
Band 7	0.34	0.50
Band 8	0.61	0.58

(Banerjee, Franceschina & Smith, 2007)

Grammatical accuracy

To achieve Band 7, candidates are expected to have good control of grammar and Punctuation.

Target-like use ratio (Ellis and Barkhuizen, 2005)

	Band 7	Band 8
Usage of 'this'	0.97	1
Usage of 'that'	1	N/A
Usage of 'these'	0.94	1
Usage of 'those'	1	1
Copula (am, is, are)	0.97	0.86
Copula (was, were)	0.97	N/A
Subject-Verb agreement	0.84	1
Passives	0.77	1

(Banerjee, Franceschina & Smith, 2007)

Coherence and cohesion	To achieve Band 7, candidates are expected to logically organise information and ideas; there is clear progression throughout and use a range of cohesive devices appropriately.	
	Frequency of use of demonstratives (this, that, these, those) • Test-takers are more likely to use demonstratives when responding to IELTS Task 2 than to IELTS Task 1. • Use of demonstratives appears to tail off at higher levels of language proficiency, suggesting that other cohesive ties come into use. We would suggest that writers at higher IELTS band levels are more likely to use lexical ties to create cohesion. We would therefore expect performances at higher IELTS band levels to display greater lexical variation and sophistication. (Banerjee, Franceschina & Smith, 2007: 21)	
Scoring		
Scoring approach	**Analytical**	Holistic
Scoring criteria	Candidates are assessed on their performance on each task by certificated IELTS examiners according to the four criteria of the IELTS Writing Test Band Descriptors (task achievement/response, coherence and cohesion, lexical resource, grammatical range and accuracy). Scores are reported in whole and half bands. The public version of the band descriptors can be found at www.ielts.org/researchers/score_processing_and_reporting.aspx	
Reliability	0.81–0.89	
SEM	N/A	

CONFIDENTIAL

TEST:	Cambridge English: Proficiency (CPE)	COMPONENT:	Writing

Task Setting			
Number of tasks	Two (Part 1 and Part 2) http://www.cambridgeenglish.org/images/information-for-candidates-2013.pdf		
Response format and Genre	Task 1: Some material to read (two texts, each approximately 100 words). Using the information in this material, you have to write an essay. Task 2: Continuous writing; Article, letter, report or review		
Task description	In the two parts of the *Cambridge English: Proficiency* Writing paper, you have to show that you can write different types of text in English. Part 1: one compulsory question Part 2: one question from a choice of four questions (2–5). For Questions 2–4, you may have to write an article, a letter, a report or a review. If you read one of the set books and want to write about it, you may decide to try Question 5. You have a choice of two tasks, 5(a) or 5(b) for this question. You may have to write an article, essay, letter, report or a review.		
Source texts	In Part 1, candidates are required to base their answers on two texts, each approximately 100 words. In Part 2 there is one text of no more than 70 words. The texts will be based on a variety of authentic, contemporary sources. Set texts for *Cambridge English: Proficiency* Q5 January 2014 - December 2014: • E. M. Forster: *Howards End* (any edition) Teachers may choose to prepare candidates for questions on this set text by studying the film directed by James Ivory (1992) as well as, or instead of, the novel. • Nick Hornby: *An Education* (screenplay) and film directed by Lone Scherfig (2009). Candidates should not attempt the optional set text question in Part 2 unless they have the necessary understanding of the text to answer the task set. Teachers are best placed to judge which, if any, of the set texts and/or film versions may be appropriate and stimulating for a given teaching situation.		
Domain	**Social**	**Work**	**Academic**
Topic (Content knowledge)	CPE candidates are expected to be able to operate confidently in a wide variety of social, work-related and study related situations. At all levels topics that might offend or otherwise unfairly disadvantage any group of candidates are avoided.		
Purpose	Task 1: To write a discursive essay in which candidates have to summarise and evaluate the key points contained in two texts of approximately 100 words each. Candidates must integrate a summary of these key points, an evaluation of the abstract arguments involved and their own ideas on the topic in a coherent essay. Task 2: In Part 2, candidates choose one from four tasks, one of which offers three set-text options. Candidates are able to select the task and topic which best suits their interests or which they think they can perform best on. The focuses are mainly referential but some also involve conative use of language and occasionally emotive use of language is called for. Candidates are expected to be able to produce the following text types for this part of the writing paper:		
Knowledge of criteria	Published information about how the tasks are scored, including criteria for correctness, and procedures used for scoring, are provided online (https://www.teachers.cambridgeesol.org/ts/exams/academicandprofessional/cpe/writing?tab=marked) as well as in the Cambridge ESOL Examination Handbook (http://cambridgeesolsoutheasteurope.files.wordpress.com/2012/11/new-writing-assessment-criteria-2012-cambridge-english-proficiency.pdf).		
Writer-reader relationship	Candidates are expected to write within an appropriate context in an appropriate register and to demonstrate sensitivity to their audience. Task types include: • A letter which is addressed to a target reader who would need to understand the writer's point of view, for example, an editor selecting appropriate responses for inclusion in a newspaper; a store manager receiving a letter of complaint; • An article which is written for a specified audience which may be an editor of a • newspaper, magazine or newsletter; • A proposal, which is similar to a report and is written for a specified audience. The proposal/report readership more often than not is a superior (e.g. a boss at work) or a peer group (e.g. one's colleagues);		

<table>
<tr><td></td><td>

- An essay which will be structured to suit a particular audience for example a tutor.

</td></tr>
<tr><td>Timing</td><td>1 hour 30 minutes</td></tr>
<tr><td>Text length of expected response</td><td>Task 1: 240–280 words
Task 2: 280–320 words</td></tr>
<tr><td>Weighting</td><td>Parts 1 and 2 are equally weighted. Each question in the paper carries equal marks. (20 marks spread over five band levels per question). The maximum total for both parts is 40 which constitutes 20% of the examination total.</td></tr>
<tr><td>Mode</td><td>Paper-based | Computer-based</td></tr>
<tr><td>Skills assessed</td><td>You’re tested on how well you can put together and develop ideas on a topic, the impression you make on the reader, your use of language, and how well you achieve the purpose for writing
- defending or attacking an argument or opinion, comparing or contrasting aspects of an argument, summarising an argument, explaining a problem and suggesting a solution, or making recommendations having evaluated an idea.</td></tr>
<tr><td>Cognitive processing</td><td>

- Macro-planning (an *assessment of the context*; *an initial activation of certain aspects of the genre)*
- Organising ideas (prioritise and plan the presentation of information)
- Translating (develop ideas by paying attention to a number of linguistic aspects. Higher level translation skills required would appear to include an ability to employ lexical variety and syntactic complexity to express subtle differences of meaning and attitude.
- Micro-planning
- Monitoring and revision (review the appropriateness of the contents and their order, the correctness of sentences, and the appropriateness of words while writing. There is an expectation that takers will operate in the problem spaces relating to both rhetoric and content and will do so both while writing and post writing)

(Shaw and Weir, 2007: Chapter 3)

</td></tr>
<tr><td colspan="2">Linguistic demands</td></tr>
<tr><td>Discourse mode (rhetorical task)</td><td>Descriptive | Narrative | Expository | Argumentative/ persuasion | Instructive/ Referral</td></tr>
<tr><td>Lexical Level</td><td>
At CPE, the lexical range of the input is once again general, non-specialised, without cultural references, and well within the reading competence of the candidates at this level (i.e. below CPE level). Candidates need to be familiar with the vocabulary relevant to the functions of the language of persuasion, description, comparison and recommendation. Trialling ensures that, as far as is possible, the language of the rubric is accessible to all candidates.
One of the main foci of the task-specific mark scheme for Part 1 is the range of language used. For example, in the task specific mark scheme range might capture the following criteria:

Language for expressing and supporting opinions;
Language for making recommendations;
Language for either attacking or defending arguments.

In Part 2, one of the main foci of the task-specific mark scheme is the range of language used. Elements of the general mark scheme include: sophisticated and/or fluent use of vocabulary, collocation and expression appropriate to the task.

Word frequency
<table>
<tr><td></td><td>CPE</td></tr>
<tr><td>K1</td><td>86.51</td></tr>
<tr><td>K2</td><td>5.12</td></tr>
<tr><td>AWL</td><td>2.43</td></tr>
<tr><td>Off-list</td><td>5.94</td></tr>
</table>
(Shaw and Weir, 2007)

Ratio of types to tokens (This is a measure of lexical variation/diversity because the higher the number of types in relation to tokens the more varied/diverse the vocabulary used.)

46.64 (Shaw and Weir, 2007)

Lexical density (Content words/total.)
</td></tr>
</table>

CONFIDENTIAL

	48 (Shaw and Weir, 2007)
Syntactic complexity	• Structures as appropriate to the specified functions and lexis e.g. present perfect, future tense; coherent organisation; linking devices; reasoned argument; justification, developing arguments through complex structures.
Grammatical accuracy	• Uses a range of vocabulary, including less common lexis, effectively and precisely. • Uses a wide range of simple and complex grammatical forms with full control, flexibility and sophistication. Errors, if present, are related to less common words and structures, or occur as slips.
Coherence and cohesion	• Text is a well-organised, coherent whole, using a variety of cohesive devices and organisational patterns with flexibility.
Scoring	
Scoring approach	**Analytical** / Holistic
Scoring criteria	Writing tasks are marked using four criteria: Content, Communicative Achievement, Organisation and Language. **Content**: focuses on how well you have fulfilled the task, in other words if you have done what you were asked to do. **Communicative Achievement:** focuses on how appropriate the writing is for the task and whether you have used the appropriate register, for example formal or informal. **Organisation:** focuses on the way you put together the piece of writing, in other words if it is logical and ordered. **Language:** focuses on vocabulary and grammar. This includes the range of language as well as how accurate it is Responses are marked on each subscale from 0 to 5.
Reliability	0.73
SEM	2.24

CONFID

TEST:	Cambridge English: Advanced (CAE)	COMPONENT:	Writing

Task Setting

Number of tasks	Two (Part 1 and Part 2)
Response format and Genre	2 pieces of continuous writing; Part 1: one compulsory question Part 2: one question from a choice of five (situationally based writing task) Genres include article, competition entry, contribution to a longer piece, essay, information sheet, letter, proposal, report, review
Task description	Part 1: Some material to read (up to 150 words) which may include material taken from advertisements, extracts from letters, emails, postcards, diaries, short articles, etc. Using this information, you may have to write an article, a report, a proposal or a letter. Part 2: A choice of four questions (2–5). You have to read some input material of no more than 80 words which describes a situation, and write one of the following: an article, a competition entry, a contribution to a longer piece, an essay, an information sheet, a letter, a proposal, a report or a review. Each of the two options for Question 5 is based on one of the two set reading texts. You may have to write an article, an essay, a report or a review.
Source texts	In Part 1, Candidates are required to deal with input material of up to 150 words. This may include material taken from advertisements, extracts from letters, emails, postcards, diaries, short articles, etc. Part 2 Q2-4 is a situationally based writing task specified in no more than 80 words. Set texts for Q5 until December 2014: One or both of the following texts may be read as preparation for the two optional set text tasks: • William Golding: *Lord of the Flies* (any edition) Teachers may choose to prepare candidates for questions on this set text by studying one of the film versions as well as, or instead of, the novel. • P D James: *The Lighthouse* (any edition) There is currently no film version of this book. Candidates should not attempt the optional set text question in Part 2 unless they have the necessary understanding of the text to answer the task set. Teachers are best placed to judge which, if any, of the set texts and/or film versions may be appropriate and stimulating for a given teaching situation. **Note:** From 2015 onward, set texts will be discontinued in the revised Cambridge English: Advanced exam.
Domain	**Social** / **Work** / Academic
Topic (Content knowledge)	CAE candidates are expected to be able to deal with topics that are more specialised and less personal than those that tend to feature at lower levels. The step up to CAE also involves coping with lexically challenging topic areas (e.g. the environment, the scientific world, traditions).
Purpose	Part 1: • To write a particular type of text and using language functions, such as evaluating, expressing opinions, hypothesising, justifying, persuading. Part 2: • To write a particular type of text based on a situation you are given.
Knowledge of criteria	Published information about how the tasks are scored, including criteria for correctness, and procedures used for scoring, are provided online (https://www.teachers.cambridgeesol.org/ts/exams/academicandprofessional/cae/writing?tab=marked) as well as in the Cambridge ESOL Examination Handbook (http://www.cambridgeenglish.org/images/cae-handbook-for-teachers.pdf).
Writer-reader relationship	Task types vary and may include formal letters, informal letters, reports, articles, notes or any combination of these. Scope is given to the candidate to assess and define their own relationship to the target reader. • A letter would not be a personal letter to a friend as this would not generate CAE level language; • A proposal or a report is predicated on there being a likely reason for the target reader to elicit the candidate's opinion, i.e. who wants the report/proposal and why is both clear and convincing to the candidate; • A competition entry would require candidates to persuade the 'judges' who are the target readership;

<table>
<tr><td></td><td>

A contribution to a guidebook, etc. should be appropriate for the intended readership of the text;
A work-related task might be addressed to a superior (requesting some personal development from boss e.g. to work abroad); management (making suggestions for the department in which the candidate may work e.g. more equipment); peers (reporting back on a particular work experience e.g. attendance at a trade fair.

(Shaw & Weir, 2007)
</td></tr>
<tr><td>Timing</td><td>1 hour 30 minutes</td></tr>
<tr><td>Text length of expected response</td><td>Part 1: 180–220 words
Part 2: 220–260 words.</td></tr>
<tr><td>Weighting</td><td>Parts 1 and 2 are equally weighted.</td></tr>
<tr><td>Mode</td><td>Paper-based | Computer-based</td></tr>
<tr><td>Skills assessed</td><td>Part 1 tests the candidate's ability to produce a specified task type in response to instructions and one or more texts, which may be supported by visual prompts. In Part 2, candidates will be expected to show that they are sensitive to the style and tone required by a task, and must be prepared to demonstrate appropriate use of two or more of the following functions as specified in the task: describing, evaluating, hypothesising, persuading, expressing opinion, comparing, giving advice, justifying and judging priorities.

Part 1:
comparing/contrasting, complaining, expressing attitude and giving opinions, evaluating, hypothesising, justifying, persuading, prioritising, summarising, advising, apologising, correcting, describing, expanding, explaining, inviting, recommending, requesting, thanking and suggesting

Part 2:
describing, narrating, persuading, complaining, reporting, presenting ideas, making proposals, evaluating, reviewing, outlining, analysing, comparing, contrasting, drawing conclusions, giving/requiring information, giving reasons/explanations, hypothesising, judging priorities, making proposals and recommendations, narrating, persuading and summarising.</td></tr>
<tr><td>Cognitive processing</td><td>

Macro-planning (an assessment of the context; a preliminary representation of the writing outcome; an evaluation of the potential problems in undertaking the task)
Organising ideas (select relevant ideas; prioritise and plan the presentation of information)
Translating (reword or paraphrase language from the input and use complex sentences suitable to the functions required)
Micro-planning
Monitoring and revision (review the appropriateness of the contents and their order, the correctness of sentences, and the appropriateness of words while writing. There is an expectation that takers will operate in the problem spaces relating to both rhetoric and content and will do so both while writing and post writing)

(Shaw and Weir, 2007: Chapter 3)
</td></tr>
<tr><td colspan="2">Linguistic demands</td></tr>
<tr><td>Discourse mode (rhetorical task)</td><td>Descriptive | Narrative | Expository | Argumentative/ persuasive | Instructive/ Referral</td></tr>
<tr><td>Lexical Level</td><td>At CAE level the lexical range of the input will be close to FCE level, i.e. well within the reading competence of the candidates at this level. In addition it will be general, non-specialised, and without cultural references. Trialling of materials for all Cambridge ESOL papers ensures that, as far as is possible, the language of the rubric is accessible to all candidates.

Word frequency
<table>
<tr><td></td><td>CAE</td></tr>
<tr><td>K1</td><td>85.94</td></tr>
<tr><td>K2</td><td>6.09</td></tr>
<tr><td>AWL</td><td>2.13</td></tr>
<tr><td>Off-list</td><td>5.84</td></tr>
</table>
(Shaw and Weir, 2007)
</td></tr>
</table>

	Ratio of types to tokens (This is a measure of lexical variation/diversity because the higher the number of types in relation to tokens the more varied/diverse the vocabulary used.) 46.94 (Shaw and Weir, 2007)
	Lexical density (Content words/total.) 49 (Shaw and Weir, 2007)
Syntactic complexity	• Structure as appropriate to the specified functions and lexis e.g. present simple, present perfect, future tenses; reason and purpose structures; explaining structures; explaining structures; comparatives and superlatives; organising and linking devices; complex sentence structures.
Grammatical accuracy	• Uses a range of vocabulary, including less common lexis, appropriately. • Uses a range of simple and complex grammatical forms with control and flexibility. • Occasional errors may be present but do not impede communication.
Coherence and cohesion	• Text is well organised and coherent, using a variety of cohesive devices and organisational patterns to generally good effect.
Scoring	
Scoring approach	**Analytical** | Holistic
Scoring criteria	Writing tasks are marked using four criteria: Content, Communicative Achievement, Organisation and Language. **Content**: focuses on how well you have fulfilled the task, in other words if you have done what you were asked to do. **Communicative Achievement:** focuses on how appropriate the writing is for the task and whether you have used the appropriate register, for example formal or informal. **Organisation:** focuses on the way you put together the piece of writing, in other words if it is logical and ordered. **Language:** focuses on vocabulary and grammar. This includes the range of language as well as how accurate it is Responses are marked on each subscale from 0 to 5. http://www.cambridgeenglish.org/images/24913-cae-information-for-candidates.pdf
Reliability	0.79
SEM	1.78

CON

TEST:	Pearson Test of English - Academic (PTE-A)	COMPONENT:	Writing

Task Setting

Number of tasks	There are three or four tasks in the Writing section of the test. Additionally four other parts of the test in the Reading and Listening sections also contribute to the score reported for Writing.
Response format and Genre	The *Writing section* of the test consists of a total of three or four tasks: Summarize written text (up to three tasks: summarize passage in one sentence) Essay (one or two essays) Other tasks contributing to Writing scores: Listening and writing: Summarize spoken text; Fill in the blanks and Write from dictation Reading and writing: Fill in the blanks Of these, only Summarize spoken text requires the test taker to compose a text.
Task description	PTE-A tests the ability to produce written English in an academic environment: Summarize written text involves comprehending, analyzing and combining information from a reading passage, and then summarizing key points in writing. Essay involves writing a persuasive or argumentative essay on a given topic. See Appendix for sample tasks.
Source texts	Summarize written text: up to 300 words; Summarize spoken text: 60-90 seconds
Domain	Social / Work / **Academic**
Topic (Content knowledge)	PTE-A test items represent the range of situations students encounter during academic study in an English speaking environment.
Purpose	Summarize written text Read a passage of up to 300 words and summarize it using one sentence. Essay Plan, write and revise an essay about a given topic. Develop a position, organize ideas, present supporting details, and control the elements of standard written English. Summarize spoken text Write a summary of a lecture for a fellow student who was not present at the lecture. Present the key points from the lecture.
Knowledge of criteria	Pearson publishes the scoring criteria in the PTE Academic Score Guide available online and elsewhere. Scoring criteria are listed in task instructions.
Writer-reader relationship	No audience is specified for Summarize written text or for Essay. Audience for Summarize spoken text is a fellow student.
Timing	Summarize written text 10 minutes and 20 minutes Essay 20 minutes Summarize spoken text 20 or 30 minutes
Text length of expected response	Summarize written text has two parts: the first is a one sentence, the second is one paragraph. Essay 200-300 words Summarize spoken text 50-70 words
Weighting	The relative contribution of different parts of the test to the overall Writing score and to the Writing score is not explicitly stated. The test involves a complex scoring system with different parts of the test contributing to more than one reported score. The score report provides an overall score, a score for each communicative skill (writing, speaking, listening, reading) and a score for each of the 'enabling skills' of grammar, oral fluency, pronunciation, spelling, vocabulary, and written discourse. For integrated skills items (that is, those assessing reading and speaking, listening and speaking, reading and writing, listening and writing, or listening and reading) the item score contributes to the score for both of the communicative skills that the item assesses. Thus Summarize spoken text contributes to both the Writing score and the Listening score as well as the enabling skills scores for grammar, spelling and vocabulary. The score range for each skill is 10–90 points.
Mode	Paper-based / **Computer-based**
Skills assessed	

Writing	**Models of Writing discourse**			
Target language Use Situation:	Domain ~~intention~~/	Reproduce	Organize/reorganize	Invent/ Generate

At postgraduate level, students need to produce a range of text types integral to academic performance	purpose / Cognitive processing				
	To learn	Listening & Writing: Fill in the blanks Writing: Write from dictation			
	To inform (inferential)		Reading & Writing: Fill in the blanks Listening & Writing: Summarize spoken text		
	To convince or persuade			Reading & Writing: Summarize written text Writing: Write essay	
What Makes a Task Difficult?	Variables related to the task itself, such as topic, the expected discourse mode of the response, the variable related to the scoring processes such as the background and experience of the raters, the nature of the rating scale, and training				

Summarize written text
Writing a summary; writing under timed conditions; taking notes while reading a text; synthesizing information; writing to meet strict length requirements; communicating the main points of a reading passage in writing; using words and phrases appropriate to the context; using correct grammar
Essay
Writing for a purpose (to learn, to inform, to persuade); supporting an opinion with details, examples and explanations; organizing sentences and paragraphs in a logical way; developing complex ideas within a complete essay; using words and phrases appropriate to the context; using correct grammar; using correct spelling; using correct mechanics; writing under timed conditions
Summarize spoken text
Writing a summary; writing under timed conditions; taking notes whilst listening to a recording; communicating the main points of a lecture in writing; organizing sentences and paragraphs in a logical way; using words and phrases appropriate to the context; using correct grammar; using correct spelling; using correct mechanics.

Cognitive processing

PTE-A involves writing to learn, to inform and to persuade.
Chan (2011) investigated composing processes on the Summarize written text and Essay items and concluded that these involved distinct cognitive processes. A 'knowledge-telling' approach based on personal experience and awareness of the essay genre was used to complete the Essay item. Most test takers did not perform much macro-planning before writing, but did require considerable low level micro-planning, monitoring and revising. The Summarize Written Text item, elicited a more recursive composing process. Most test takers engaged in careful macro-planning as well as some discourse synthesis processes such as selecting and organizing relevant ideas from the source text. However, as the item requires test takers to write only one sentence, micro-planning, monitoring and revising did not come into play.

Linguistic demands

Discourse mode (rhetorical task)	**Descriptive**	**Narrative/ Historical/ Biographical**	**Expository**	**Argumentative**	**Instructive/~~Referral~~**

<table>
<tr><td>Lexical Level</td><td>To achieve the highest score for Vocabulary range on the Essay, the test taker must demonstrate a good command of a broad lexical repertoire, idiomatic expressions and colloquialisms. In Summarizing spoken and written texts, the test taker must make an appropriate choice of words.

The following table from O'Loughlin (2013) relates to responses elicited by three PTE-A task types. It shows the percentage of words (tokens) from these responses that can be found on the academic word list (AWL) of academic vocabulary. O'Loughlin argues that academic text is characterised by at least 4% of academic vocabulary and that 'the higher the percentage of academic tokens elicited by each item, the stronger the claim for the test's lexical validity will be' (O'Loughlin 2013: 14).

<table>
<tr><td>TASK</td><td>No of items</td><td>Mean AWL tokens(%)</td><td>Standard deviation</td><td>AWL token range (%)</td></tr>
<tr><td>Summarize written text</td><td>38</td><td>7.24</td><td>3.89</td><td>0.70-16.36</td></tr>
<tr><td>Summarize spoken text</td><td>38</td><td>5.93</td><td>2.58</td><td>1.34-12.24</td></tr>
<tr><td>Essay</td><td>19</td><td>5.86</td><td>4.06</td><td>0-10.71</td></tr>
</table>
</td></tr>
<tr><td>Syntactic complexity</td><td>To achieve the highest score for General linguistic range on the Essay, the test taker should exhibit smooth mastery of a wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity. There should be no sign that the test taker is restricted in what they want to communicate</td></tr>
<tr><td>Grammatical accuracy</td><td>To achieve the highest score for Grammar, in the Essay the test taker must demonstrate consistent grammatical control of complex language. Errors should be rare and difficult to spot. In Summarizing spoken and written texts, the test taker must use correct grammatical structures.</td></tr>
<tr><td>Coherence and cohesion</td><td>To achieve the highest score for development, structure and coherence in the Essay, the test taker must show good development and logical structure.</td></tr>
<tr><td colspan="2">Scoring</td></tr>
<tr><td>Scoring approach</td><td>Analytical Holistic</td></tr>
<tr><td>Scoring criteria</td><td>Summarize written text and Summarize spoken text are scored for Content, Form, Grammar and Vocabulary. Summarize spoken text is also scored for Spelling
Essay is scored for Development, Structure and Coherence; Form, General linguistic range, Grammar usage and mechanics, Spelling, Vocabulary range.
The written parts of PTE Academic are scored using the Intelligent Essay Assessor™ (IEA), an automated scoring tool that uses Pearson's Knowledge Analysis Technologies™ (KAT™) engine. The KAT engine automatically evaluates the meaning of text by examining whole passages. It is claimed that the KAT engine evaluates writing as accurately as skilled human raters using Latent Semantic Analysis (LSA): an approach that generates semantic similarity of words and passages by analyzing large bodies of relevant text.</td></tr>
<tr><td>Number of raters</td><td>PTE Academic is machine scored. Human raters are only employed in the initial training of the Intelligent Essay Assessor and in cases where off-topic responses or other situations prove problematic for the automated scoring system.</td></tr>
<tr><td>Reliability</td><td>0.91 (automated scoring)</td></tr>
<tr><td>SEM</td><td>N/A</td></tr>
</table>

TEST:	Internet-based Test of English as a Foreign Language (TOEFL iBT)	COMPONENT:	Writing

Task Setting			
Number of tasks	Two (Task 1 and Task 2)		
Response format and Genre	Task 1 is an 'integrated' writing task involving reading a short passage, listening to related material, and then describing how the information relates. Task 2 is an 'independent' writing task involving a discursive essay on a given topic based on the test taker's own experience.		
Task description	The writing test is designed to measure the ability to use writing to communicate in an academic environment. In Task 1, test takers are three minutes to read and take notes if they wish about a reading passage of around 230 to 300 words. They then listen to a lecture (also of 230 to 300 words - played only once) which challenges or supports the information in the reading passage. They are allowed to take notes during the lecture. The reading passage then reappears along with a question and examinees type their responses. The reading passage remains on screen and test takers can use their notes. Test takers write a summary in connected English prose of important points made in the listening passage, and explain how these relate to the key points of the reading passage. In Task 2, test takers write an essay on a topic of general interest that is presented in the form of a brief statement. They respond to the topic using their own experience. Typical essay questions begin with statements such as: —Do you agree or disagree with the following statement? Use reasons and specific details to support your answer. —Some people believe X. Other people believe Y. Which of these two positions do you prefer/agree with? Give reasons and specific details.		
Source texts	Reading passage and related recording (lecture) 230 to 300 words each.		
Domain	**Social**	**Work**	**Academic**
Topic (Content knowledge)	An illustrative list of 185 independent writing topics is provided in *The Official Guide to the TOEFL Test* available for purchase from ETS. These are topics of general interest such as the following: Nowadays, food has become easier to prepare. Has this change improved the way people live? Use specific reasons and examples to support your answer. Do you agree or disagree with the following statement? The best way to travel is in a group led by a tour guide. Use specific reasons and examples to support your answer.		
Purpose	Task 1: to summarize the points in a listening passage and explain how they relate to specific points in a reading passage. To demonstrate quality of writing and completeness and accuracy of content. Task 2: to write an essay giving an opinion in response to a question on an issue: to state, explain, and support an opinion; to organize an essay coherently, demonstrating quality and accuracy of sentence structure and vocabulary.		
Knowledge of criteria	ETS publishes information about how the tasks are scored, including the rating scales for each task. These can be downloaded from the ETS website (https://www.ets.org/Media/Tests/TOEFL/pdf/Writing_Rubrics.pdf) and can be found in other publications for test takers.		
Writer-reader relationship	No audience is specified for either Task 1 or Task 2.		
Timing	The TOEFL Writing test takes 50 minutes: ☐ Integrated Writing task (20 minutes) — read a short passage and listen to a short lecture. Then write in response to what was read and listened to. ☐ Independent Task (30 minutes) — write an essay in response to a Writing topic.		
Text length of expected response	Task 1 test takers are advised to write 150 to 225 words Task 2: test takers are advised to write at least 300 words		
Weighting	Equal weighting is given to both Tasks. Writing scores are aggregated and reported on a scale from 0 to 30. This contributes to a cumulative overall score for the test reported on a scale from 0 to 120.		
Mode	~~Paper-based~~	Computer-based	
Skills assessed	For each Task, the test taker is awarded a single overall score. Performance is judged according to how effectively the test taker addresses the topic; how well organized and well developed the response is; the extent to which it displays unity, progression, and coherence and the accurate use of language: grammar and word choice.		

Cognitive processing	Cumming et al. (2006) analyzed the discourse characteristics of 36 examinees' written responses to prototype independent and integrated essay questions. Cumming found that the discourse characteristics varied as expected, both with writers' proficiency levels and with task types. Discourse features analyzed included text length, lexical sophistication, syntactic complexity, grammatical accuracy, argument structure, orientations to evidence, and verbatim uses of source text. Higher scores (previously assigned by raters) were associated with longer responses, greater lexical sophistication, syntactic complexity, and grammatical accuracy. Cho, Rijmen and Novák (2013) examined the influence of prompt characteristics on scores on 107 integrated writing tasks administered between 2005 and 2009. Two variables related to perceived task difficulty by expert judges (distinctness of ideas within the prompt and difficulty of ideas in the passage), were also identified as potential sources of variation in the average integrated writing scores.				
Linguistic demands					
Discourse mode (rhetorical task)	**Descriptive**	Narrative/ Historical/ Biographical	**Expository**	**Argumentative**	Instructive/ Referral
Lexical Level	To merit a score of 5 on the independent writing task, the test taker must display consistent facility in the use of language including appropriate word choice, and idiomaticity, though the response may have minor lexical errors.				
Syntactic complexity	To merit a score of 5 on the independent writing task, the test taker must demonstrate consistent facility in the use of language including syntactic variety.				
Grammatical accuracy	To merit a score of 5 on the independent writing task, the response must include only minor grammatical errors. To merit a score of 5 on the integrated writing task, language errors present in the response must not result in inaccurate or imprecise presentation of content or connections between ideas.				
Coherence and cohesion	To merit a score of 5 on the integrated writing task, the test taker must coherently and accurately present important information from the lecture in relation to the relevant information presented in the reading. The response must be well organized.				
Scoring					
Scoring approach	Analytical		**Holistic**		
Scoring criteria	n/a				
Number of raters	The Writing section of the TOEFL test uses a combination of human and automated scoring. Four to six different raters score each response. This is done blindly i.e. the raters are not aware of each others' scores.				
Reliability	0.74				
SEM	2.76				

CONFIDENTIAL

TEST:	Occupational English Test (OET)	COMPONENT:	Writing

Task Setting					
Number of tasks	1				
Response format and Genre	Continuous writing; letter				
Task description	Candidates are required to write a letter (referral, transfer or advice) based on case notes which are provided. Sometimes, and particularly for some professions, a different type of letter is required (e.g., a letter of transfer or discharge, or a letter to advise or inform a patient, carer, or group)				
Source texts	Case notes about 2 pages are presented in point forms.				
Domain	Social	**Work**	Academic		
Topic (Content knowledge)	Medical				
Purpose	To write a referral letter				
Knowledge of criteria	Published information about how the tasks are scored, including criteria for correctness, and procedures used for scoring, are provided in the OET Handbook (http://www.occupationalenglishtest.org/DesktopModules/ViewDocument.aspx?Club=oet&DocumentID=5C898FC0-F9A5-4EB9-A5BC-D6057F35E77C)				
Writer-reader relationship	Candidates are expected to write a letter to a medical body.				
Timing	45 minutes (5 minutes for reading and 45 minutes for writing)				
Text length of expected response	180-200 words				
Weighting	N/A				
Mode	**Paper-based**	Computer-based			
Skills assessed	The writing task, taken directly from the workplace context, requires you to select and organise relevant information and present it in a clear, accurate form that is appropriate for the intended reader. Preparing such a letter with only limited time is a reality for practising professionals. The letter is assessed against five criteria: • Overall task fulfilment – including whether the response is of the required length • Appropriateness of language – including the use of appropriate vocabulary and tone in the response, and whether it is organised appropriately • Comprehension of stimulus – including whether the response shows you have understood the situation and provide relevant rather than unnecessary information to your reader • Control of linguistic features (grammar and cohesion) – how effectively you communicate using the grammatical structures and cohesive devices of English • Control of presentation features (spelling, punctuation and layout) – how these areas affect the message you want to communicate				
Cognitive processing	• Macro-planning • Selecting relevant materials from source • Organising ideas • Micro-planning • Monitoring and revision				

Linguistic demands					
Discourse mode (rhetorical task)	**Descriptive**	Narrative	**Expository**	Argumentative/ persuasion	**Instructive/ Referral**
Lexical Level	N/A				

Syntactic complexity	N/A	
Grammatical accuracy	N/A	
Coherence and cohesion	N/A	
Scoring		
Scoring approach	**Analytical**	Holistic
Scoring criteria	The writing is assessed by two qualified and experienced English assessors who have been trained in OET assessment procedures. The assessment is based on the following 5 criteria • overall task fulfilment • appropriateness of language • comprehension of stimulus • linguistic features (grammar and cohesion) • presentation features (spelling, presentation, layout). Assessors give a score from 1 to 6 for each of the five criteria listed above using a detailed set of level descriptors to guide their decisions. A score of 6 is the highest for each criterion. The five criteria are equally weighted in the scoring and analysis process.	
Reliability	0.89	
SEM	N/A	

CONFIDENTIAL

TEST:	General English Test of Proficiency (GEPT) - Advanced	COMPONENT:	Writing

Task Setting

Number of tasks	Two (Part 1 and Part 2)
Response format and Genre	2 pieces of continuous writing; Part 1: essay; Part 2: letter
Task description	The test contains two parts. In Part 1, candidates are required to read two articles and then write an essay. In Part 2, candidates need to interpret two charts before writing a persuasive letter.
Source texts	Part 1: textual: two articles, each of approximately 400 words long, written in an editorial style similar to that found in newspapers and magazines, and offering contrasting opinions on one topic. Part 2: *Non-textual*: two charts, graphs, or tables providing different but related information on one topic.
Domain	**Social** / Work / **Academic**
Topic (Content knowledge)	Candidates are expected to be able to deal with topics that are used across the public, occupational, and educational domains. Topics in the public domain include public spaces (street, shops, restaurants, sports, or entertainment venues) and other social networks outside the home. Occupational topics include workplace settings (offices, workshops, conferences), etc. Topics in the educational domain include schools, colleges, classrooms, residence halls, etc.
Purpose	To write an essay based upon one of two topic choices.
Knowledge of criteria	Published information about how the tasks are scored, including criteria, procedures used for scoring and benchmarked sample scripts are provided online (https://www.lttc.ntu.edu.tw/GEPT1/Advanced/06.test_score.htm)
Writer-reader relationship	The intended readership is implicitly implied by the task context, e.g.to writer to the judges of the writing contest, but no specific information regarding the writer-reader relationship is provided in the task instruction. (Chan et al, 2014)
Timing	1 hour 45 minutes in total
Text length of expected response	Part 1:250 Part 2: 250
Weighting	N/A
Mode	**Paper-based** / **Computer-based**
Skills assessed	• can summarize articles on general and professional topics • can write well-organized and coherent essays, with appropriate lexical and grammatical usage • can express their opinions on a range of topics and discuss them in depth
Cognitive processing	• Task representation • Macro-planning • Organising ideas • Global careful reading • Selective reading • Connecting ideas from multiple sources and generating new representations • Organising ideas in relation to input texts • Organising ideas in relation to own text • Micro-planning • Monitoring and revision Chan et al (2014)

Linguistic demands

Discourse mode (rhetorical task)	**Descriptive**	**Narrative**	**Expository**	**Argumentative/ persuasive**	Instructive/ Referral
Lexical Level	Band 5 descriptors: A wide range of vocabulary is used to complete the tasks effectively. Band 3 descriptors (pass): The range of vocabulary is adequate to complete the tasks and lexical items are used appropriately most of the time.				

Syntactic complexity **Grammatical accuracy**	Band 5 descriptors: A wide range of structures is used to complete the tasks accurately, appropriately and effectively. Band 3 descriptors (pass): Structures used are accurate and appropriate for the most part.	
Coherence and cohesion	Band 5 descriptors: A wide range of structures is used to complete the tasks accurately, appropriately and effectively. Band 3 descriptors (pass): The text is coherent and logically organized in general.	
Scoring		
Scoring approach	**Analytical**	Holistic
Scoring criteria	Candidates' responses are scored according to four criteria: relevance and adequacy; coherence and organization; lexical use; grammatical use. To pass the test, candidates have to score 3 or above out of 5 in each of the four criteria	
Reliability	0.89-0.90 (https://www.lttc.ntu.edu.tw/E_LTTC/E_GEPT/files/2014_GEPT_Information_Brochure.pdf)	
SEM	N/A	

CONFIDENTIAL

TEST:	Examination for the Certificate of Proficiency in English (ECPE)	COMPONENT:	Writing

Task Setting

Number of tasks	One
Response format and Genre	1 piece of continuous writing; Genres: essay; A typical writing task includes three parts: • the context/situation (e.g., "Some psychologists claim that people inherit most of their personality traits from their parents.") • a statement of what the test taker has to discuss (e.g., "Do you believe we can tell more from body language than from what people say?") • instructions of what is to be achieved (e.g., "Give specific details and reasons to support your viewpoint.")
Task description	The ECPE writing section comprises a single essay question. Test takers are given two tasks to choose from. They write only on one essay question. There is no word limit but test takers are advised that their responses will be marked down if they are extremely short.
Source texts	N/A
Domain	**Social** / **Work** / Academic
Topic (Content knowledge)	Candidates are expected to be able to deal with topics that are used across the public, occupational, and educational domains

Public	public spaces (street, shops, restaurants, sports, or entertainment venues) and other social networks outside the home
Occupational	workplace settings (offices, workshops, conferences), etc.
Educational	schools, colleges, classrooms, residence halls, etc.

Purpose	To write an essay based upon one of two topic choices.
Knowledge of criteria	Published information about how the tasks are scored, including criteria, procedures used for scoring and benchmarked sample scripts are provided online (http://www.cambridgemichigan.org/institutions/products-services/tests/proficiency-certification/ecpe/) (http://www.cambridgemichigan.org/wp-content/uploads/2014/11/ECPE-Writing-Benchmarks.pdf).
Writer-reader relationship	No specific information regarding the writer-reader relationship is provided in the task.
Timing	30 minutes
Text length of expected response	No word limit (but candidates are advised to write about one-and-a-half to two pages)
Weighting	N/A
Mode	**Paper-based** / **Computer-based**
Skills assessed	The ECPE writing task gives test takers the opportunity to demonstrate their ability to • develop a topic and organize the content; • make smooth connections between ideas; • use a variety of grammatical constructions accurately and appropriately; and • use a range of vocabulary accurately and appropriately.
Cognitive processing	• Macro-planning • Organising ideas • Translating • Micro-planning • Monitoring and revision
Linguistic demands	
Discourse mode (rhetorical task)	**Descriptive** / **Narrative** / **Expository** / **Argumentative/ persuasive** / Instructive/ Referral

Lexical Level	Band 4 descriptors (exceeds standard): Vocabulary use shows flexibility; is usually appropriate; Any inappropriate vocabulary does not confuse meaning Band 3 descriptors (standard): Adequate vocabulary, but may sometimes be inappropriately used	
Syntactic complexity	Band 4 descriptors (exceeds standard): Both simple and complex syntax adequately used; good morphological control Band 3 descriptors (standard): Both simple and complex syntax present; For some, syntax is cautious but accurate, while others are more fluent but less accurate	
Grammatical accuracy		
Coherence and cohesion	Band 4 descriptors (exceeds standard): Topic clearly and completely developed, with acknowledgment of its complexity; Organization is controlled and shows appropriateness to the material; Few problems with connection Band 3 descriptors (standard): Topic clearly developed, but not always completely or with acknowledgment of its complexity; Organization generally controlled; connection sometimes absent or unsuccessful	
Scoring		
Scoring approach	**Analytical**	Holistic
Scoring criteria	The test taker's writing performance is evaluated independently by two examiners. The 5-band rating scale guides raters to attend to the test taker's performance in the following areas: • Rhetoric (topic development, organization of information, and connection between ideas) • Grammar/Syntax (range and accuracy of grammar) • Vocabulary (range and accuracy of vocabulary)	
Reliability	Rater 1/Rater 2 Agreement: 76.7%; Pass/Fail Agreement: 84.7% (http://www.lsa.umich.edu/UMICH/eli/Home/Test%20Programs/ECPE/Officials/AdminReports/ECPE10MJTestAdminReport.pdf)	
SEM	N/A	

CONFIDE

Test	IESOL C1	Component	Writing

Task Setting			
Number of tasks	There are two tasks in the C1 Expert Writing test: Writing 1 and Writing 2		
Response format and Genre	Writing 1: letter, report, argument or article expressing stance, opinion, justification, argument Writing 2: personal letter, narrative or descriptive composition: persuasion, argument, hypothesis; expressing mood, opinion, justifying, evaluating etc.		
Task description	Writing 1 calls for a letter, report, argument or article. the candidate responds for a specified reader to a written, graphic or visual input Writing 2 calls for a personal letter, a narrative (story) or descriptive composition for a specified reader Although both tasks call for responses 'for a specified reader', I found that readers are not always specified in the sample materials provided. See Appendix for sample tasks.		
Source texts	Brief input to Writing 1 of 75 to 125 words – may also include a visual element Writing 2 prompt only		
Domain	**Social**	Work	Academic
Topic (Content knowledge)	The syllabus published in the Qualification handbook includes a comprehensive list of topics. Topics appearing at C1 level, not covered at B2 include: image; region-specific phenomena; stress; money management; leisure/work ratio; common currency e.g. the euro; migration; anti-social behaviour; ethical shopping; retail therapy; food fashions; diplomatic services; employment agencies; how geography affects people; alternative places to live eg.; underwater, on Mars; universal languages eg. Esperanto; body language; weather and mood; importance of maths in everyday life; education systems; teaching and learning; endangered species; future of the planet; architecture; literature; popular culture; genetic modification; ethics; animal testing; human rights; citizenship; the global village		
Purpose	Writing 1: to produce a formal response [to the input text] for an intended public audience expressing stance, opinion, justification, argument Writing 2: to produce a personal letter, a narrative or descriptive composition involving persuasion, argument, hypothesis; expressing mood, opinion, justifying, evaluating etc		
Knowledge of criteria	Criteria are detailed in the published syllabus: Task Fulfilment, Grammar, Vocabulary and Structure. Task specific rating scales are created for each test form. Task fulfillment: a measure of how far the candidate has achieved/addressed the task. Has the candidate done what was asked (eg - "Write four sentences" – are there four sentences?) Grammar: a measure of the range, appropriacy and accuracy of grammar. Vocabulary: a measure of the range, accuracy and appropriacy of vocabulary. This criterion also includes spelling accuracy. Structure: a measure of coherence and cohesion. How the text is put together? Is there an attempt to link the ideas and to organise them in a coherent manner? This criterion also includes the accurate use of punctuation.		
Writer-reader relationship	Audiences are inconsistently specified in sample materials, but include a close friend, a teacher, readers of a journal and a sponsor of voluntary organisations.		
Timing	Candidates are allowed 2 hours and 40 minutes to complete the Reading and Writing tests. No recommendations are given on the exam paper concerning the timing of sections.		
Text length of expected response	Writing 1: 150 – 200 words Writing 2: 250 - 300 words		
Weighting	For both Writing 1 and Writing 2, a mark is awarded on a scale from 0 to 3 for each of the four criteria giving a possible maximum of 12 points. The two tasks are equally weighted to give 24 points in total. Grades are awarded as follows: First Class Pass 20-24 Pass 12-19 Narrow Fail 11 Fail 0-10		
Mode	Paper-based	~~Computer-based~~	
Skills assessed	According to the Qualifications handbook, The C1 Expert candidate will be able to: – write well-structured, detailed texts on complex subjects clearly communicating ideas,		

	impressions, feelings and opinions – write complex narratives and descriptions varying style appropriately – write a well-structured, detailed argument emphasising important points, expanding and supporting relevant ideas – write formal letters, reports or articles to fulfil a wide range of functions including those requiring a tactful approach – use all punctuation marks accurately and effectively – spell correctly words used in work, study and daily life including commonly used technical words – consistently control grammar to communicate effectively with few errors even when complex structures are employed – use a wide range of vocabulary suitable for purpose – generally use idiomatic expressions appropriately and naturally – use a wide range of complex structures – adapt register effectively and appropriately according to purpose in most contexts – link and organise ideas using a range of linking words and cohesive devices – organise text effectively to reflect the structure of the ideas expressed.				
Cognitive processing	• Macro-planning • Organising ideas • Micro-planning • Monitoring and revision				
Linguistic demands					
Discourse mode (rhetorical task)	**Descriptive**	**Historical/ biographical**	Expository	**Argumentative**	~~Instructive~~
Lexical Level	Candidates should: – use a wide range of vocabulary suitable for purpose – spell correctly words used in work, study and daily life including commonly used technical words – generally use idiomatic expressions appropriately and naturally				
Syntactic complexity	Detailed grammatical specifications are provided covering B2 Communicator up to C2 Mastery. Additional grammar points appearing at C1, but not at lower levels include: word order in complex sentences, including order selected for emphasis; there could be/would be/should be; could have/would have/should have; wider range of conjunctions including on condition that, provided that; comparative clauses; more complex participial clauses; describing action with –ed; all verb forms active and passive; would expressing habit in the past; mixed conditionals; reported speech with full range of; tenses and introductory verbs; extended phrasal verbs such as get; round to, carry on with; question tags using all tenses; imperative + question tag; contracted forms appropriate to this level; should have/might have/may have/could have/must have and negative forms of these, can't have, needn't have; extended range of complex noun phrases; preposition + having + past participle such as having eaten; collocation of intensifiers with absolute and relative adjectives such as absolutely gorgeous, very pretty; accurate use of all punctuation; the correct spelling of words used in work, studies and daily life including familiar technical words; a range of logical markers (in this respect, accordingly); sequence markers (subsequently); a wider range of discourse markers to structure formal and informal speech (can we now turn to) Candidates should: – use a wide range of complex structures				
Grammatical accuracy	Candidates should: – consistently control grammar to communicate effectively with few errors even when complex structures are employed – use all punctuation marks accurately and effectively				
Coherence and cohesion	Candidates should: – link and organise ideas using a range of linking words and cohesive devices – write well-structured, detailed texts on complex subjects clearly communicating ideas, impressions, feelings and opinions – write complex narratives and descriptions varying style appropriately – write a well-structured, detailed argument emphasising important points, expanding and supporting relevant ideas – organise text effectively to reflect the structure of the ideas expressed				
Scoring					
Scoring approach	Analytical		~~Holistic~~		

<table>
<tr><td>Scoring criteria</td><td>The qualifications handbook lists Can Do statements for the level derived from the CEFR and provides the following chart to illustrate the rating scales employed:

<table>
<tr><td></td><td>Task Fulfilment</td><td>Grammar</td><td>Vocabulary</td><td>Structure</td></tr>
<tr><td>3
First Class Pass</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td></tr>
<tr><td>2
Pass</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td></tr>
<tr><td>1
Narrow Fail</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td></tr>
<tr><td>0
Fail</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td></tr>
</table></td></tr>
<tr><td>Number of raters</td><td>1</td></tr>
<tr><td>Reliability</td><td>unavailable</td></tr>
</table>

CONFIDENTIAL

Test	City & Guilds IESOL C2	Component	Writing
Task Setting			
Number of tasks	There are two tasks in the C2 Mastery Writing test: Writing 1 and Writing 2		
Response format and Genre	Writing 1: letter, report, argument or article expressing stance, opinion, justification, argument Writing 2: personal letter, narrative or descriptive composition: persuasion, argument, hypothesis; expressing mood, opinion, justifying, evaluating etc.		
Task description	Writing 1 calls for a letter, report, argument or article. the candidate responds for a specified reader to a written, graphic or visual input Writing 2 calls for a personal letter, a narrative (story) or descriptive composition for a specified reader Although both tasks call for responses 'for a specified reader', I found that readers are not always specified in the sample materials provided. See Appendix for sample tasks.		
Source texts	Brief input to Writing 1 of 50 to 100 words – may also include a visual element Writing 2 prompt only		
Domain	**Social**	Work	Academic
Topic (Content knowledge)	The syllabus published in the Qualification handbook includes a comprehensive list of topics. Topics appearing at C2 level, not covered at C1 include personal learning style; demographics; life plans; the social importance of leisure; travel restrictions & border controls; tolerance & respect; consumerism; genetically modified food; cookery; government; living in hostile environments; language and culture; effect of weather on lifestyle; design; knowledge versus skills; education as a basic human right; individual's/society's responsibilities; youth culture; the limits of human endeavour; world events; world economy		
Purpose	Writing 1: to produce a formal response [to the input text] for an intended public audience expressing stance, opinion, justification, argument Writing 2: to produce a personal letter, a narrative or descriptive composition involving persuasion, argument, hypothesis; expressing mood, opinion, justifying, evaluating etc		
Knowledge of criteria	Criteria are detailed in the published syllabus: Task Fulfilment, Grammar, Vocabulary and Structure. Task specific rating scales are created for each test form. Task fulfillment: a measure of how far the candidate has achieved/addressed the task. Has the candidate done what was asked (eg - "Write four sentences" – are there four sentences?) Grammar: a measure of the range, appropriacy and accuracy of grammar. Vocabulary: a measure of the range, accuracy and appropriacy of vocabulary. This criterion also includes spelling accuracy. Structure: a measure of coherence and cohesion. How the text is put together? Is there an attempt to link the ideas and to organise them in a coherent manner? This criterion also includes the accurate use of punctuation.		
Writer-reader relationship	Audiences are inconsistently specified in sample materials, but include a festival committee, a café owner and readers of newspapers, magazines and newsletters.		
Timing	Candidates are allowed 2 hours and 40 minutes to complete the Reading and Writing tests. No recommendations are given on the exam paper concerning the timing of sections.		
Text length of expected response	Writing 1: 200 – 250 words Writing 2: 250 - 300 words		
Weighting	For both Writing 1 and Writing 2, a mark is awarded on a scale from 0 to 3 for each of the four criteria giving a possible maximum of 12 points. The two tasks are equally weighted to give 24 points in total. Grades are awarded as follows: First Class Pass 20-24 Pass 12-19 Narrow Fail 11 Fail 0-10		
Mode	Paper-based	~~Computer-based~~	
Skills assessed	According to the Qualifications handbook, the C2 Mastery candidate will be able to: – write well-structured texts on complex or abstract subjects – write clear, smoothly flowing, complex formal letters, reports or articles in styles fullyappropriate to purpose and target readership – describe or narrate in an assured natural manner consistently maintaining the style appropriate to purpose		

<table>
<tr><td></td><td>

develop cogent and smoothly flowing arguments
use all punctuation marks accurately and effectively
spell correctly words used in more specialised contexts such as business, academic and international affairs
maintain consistent control of complex grammatical structures
use a very wide range of vocabulary demonstrating ability to distinguish between finer shades of meaning
use idioms appropriately and naturally
use a full range of structures to achieve different styles and purposes
consistently use appropriate register in a wide range of contexts including those requiring sensitive and tactful responses
use a range of linguistic devices to create coherent and cohesive writing
structure texts logically using linguistic markers to enable the reader to understand significant points.

</td></tr>
<tr><td>Cognitive processing</td><td>

Macro-planning
Organising ideas
Micro-planning
Monitoring and revision

</td></tr>
<tr><td colspan="2">Linguistic demands</td></tr>
<tr><td>Discourse mode (rhetorical task)</td><td colspan="5">Descriptive | Historical/ biographical | ~~Expository~~ | Argumentative | ~~Instructive~~</td></tr>
<tr><td>Lexical Level</td><td>Candidates should:

use a very wide range of vocabulary demonstrating ability to distinguish between finer shades of meaning
use idioms appropriately and naturally
spell correctly words used in more specialised contexts such as business, academic and international affairs

</td></tr>
<tr><td>Syntactic complexity</td><td>Detailed grammatical specifications are provided covering B2 Communicator up to C2 Mastery. Additional grammar points appearing at C2, but not at lower levels include:
full range of conjunctions; the correct spelling of words used in more specialized contexts (such as business, academia, international affairs); a full range of discourse markers adapted to context and register. Candidates should:

use a full range of structures to achieve different styles and purposes

</td></tr>
<tr><td>Grammatical accuracy</td><td>Candidates should:

maintain consistent control of complex grammatical structures
use all punctuation marks accurately and effectively

</td></tr>
<tr><td>Coherence and cohesion</td><td>Candidates should:

write well-structured texts on complex or abstract subjects
write clear, smoothly flowing, complex formal letters, reports or articles in styles fully appropriate to purpose and target readership
develop cogent and smoothly flowing arguments
use a range of linguistic devices to create coherent and cohesive writing
structure texts logically using linguistic markers to enable the reader to understand significant points.

</td></tr>
<tr><td colspan="2">Scoring</td></tr>
<tr><td>Scoring approach</td><td colspan="5">Analytical | ~~Holistic~~</td></tr>
<tr><td>Scoring criteria</td><td>The qualifications handbook lists Can Do statements for the level derived from the CEFR and provides the following chart to illustrate the rating scales employed:
<table>
<tr><td></td><td>Task Fulfilment</td><td>Grammar</td><td>Vocabulary</td><td>Structure</td></tr>
<tr><td>3
First Class Pass</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td></tr>
<tr><td>2
Pass</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td></tr>
<tr><td>1
Narrow Fail</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td><td>Description relevant to specific level</td></tr>
<tr><td>0</td><td>Description</td><td>Description</td><td>Description</td><td>Description</td></tr>
</table>
</td></tr>
</table>

	Fail	relevant to specific level	relevant to specific level	relevant to specific level	relevant to specific level	
Number of raters	1					
Reliability	unavailable					

CONFIDENTIAL

Appendix 7 – Detailed analyses of the selected speaking tests (IELTS, CPE, CAE, PTE-A, TOEFL iBT, OET, GEPT Advanced, ECPE, IESOL & ISESOL Level 2 and IESOL & ISESOL Level 3)

CONFIDENTIAL

TEST:	International English Language Testing System (IELTS)	COMPONENT:	Speaking

Features of the Test	
Skills focus	The Speaking Test consists of an oral interview between the candidate and an examiner. All Speaking tests are recorded. There are three parts to the test and each part fulfils a specific function in terms of pattern of interaction, task input and candidate output. The candidates will deal with functions including providing personal information, expressing and justifying opinions, explaining, suggesting, speculating, expressing preferences, comparing, summarizing, narrating, etc.
Task description	Candidates for IELTS Speaking, which is a multi-level test, are expected to engage with three tasks which require them to answer short questions, speak at length on a familiar topic, and discuss more abstract questions (see below for more detail on each part).
Further task focus information	In Part 1 (Introduction and Interview), the examiner introduces himself/herself and checks the candidate's identity. Then the examiner asks the candidate general questions on some familiar topics such as home, family, work, studies, interests. To ensure consistency, questions are taken from a scripted examiner frame. This part of the test focuses on the candidate's ability to communicate opinions and information on everyday topics and common experiences or situations by answering a range of questions. In Part 2 (Long turn), the examiner gives the candidate a task card which asks the candidate to talk about a particular topic, includes points to cover in their talk and instructs the candidate to explain one aspect of the topic. Candidates are given 1 minute to prepare their talk, and are given a pencil and paper to make notes. Using the points on the task card effectively, and making notes during the preparation time, will help the candidate think of appropriate things to say, structure their talk, and keep talking for 2 minutes. The examiner asks the candidate to talk for 1 to 2 minutes, stops the candidate after 2 minutes, and asks one or two questions on the same topic. Part 2 lasts 3-4 minutes, including the preparation time. This part of the test focuses on the candidate's ability to speak at length on a given topic (without further prompts from the examiner), using appropriate language and organising their ideas coherently. It is likely that the candidate will need to draw on their own experience to complete the long turn. In Part 3 (Discussion), the examiner and the candidate discuss issues related to the topic in Part 2 in a more general and abstract way and – where appropriate – in greater depth. This part of the test focuses on the candidate's ability to express and justify opinions and to analyse, discuss and speculate about issues.
Timing	11-14 minutes (Part 1: 4-5 mins; Part 2: 3-4 mins; Part 3: 4-5 minutes)
Scoring and weighting	Speaking performances are assessed by certificated *IELTS* examiners. All *IELTS* examiners hold relevant teaching qualifications and are recruited as examiners by the test centres and approved by British Council or IDP: *IELTS* Australia. IELTS examiners undergo intensive face –to-face training and standardisation to ensure that they can apply the test administration and rating procedures in a valid and reliable manner. Examiners award a band score for each of four criterion areas: Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy and Pronunciation. The four criteria are equally weighted. Scores are reported in whole and half bands. Detailed performance descriptors have been developed which describe spoken performance at the nine *IELTS* bands. Versions of the band descriptors for Speaking have been developed and made available on the *IELTS* website to help stakeholders better understand the level of performance required to attain a particular band score in each of the criterion areas. **Fluency and Coherence** This criterion refers to the ability to talk with normal levels of continuity, rate and effort and to link ideas and language together to form coherent, connected speech. The key indicators of fluency are speech rate and speech continuity. The key indicators of coherence are logical sequencing of sentences, clear marking of stages in a discussion, narration or argument, and the use of cohesive devices (e.g. connectors, pronouns and conjunctions) within and between sentences. **Lexical Resource** This criterion refers to the range of vocabulary the candidate can use and the precision with which meanings and attitudes can be expressed. The key indicators are the variety of words used, the adequacy and appropriacy of the words used and the ability to circumlocute (get round a vocabulary gap by using other words) with or without noticeable hesitation. **Grammatical Range and Accuracy** This criterion refers to the range and the accurate and appropriate use of the candidate's grammatical resource. The key indicators of grammatical range are the length and complexity of the spoken sentences, the appropriate use of subordinate clauses, and the range of sentence structures, especially to move elements around for information focus. The key indicators of grammatical accuracy are the number of grammatical errors in a given amount of speech and the communicative effect of error. **Pronunciation** This criterion refers to the ability to produce comprehensible speech to fulfil the Speaking test

	requirements. The key indicators will be the amount of strain caused to the listener, the amount of the speech which is unintelligible and the noticeability of L1 influence.
Cognitive processing: Levels of speaking	**Conceptualisation**
	Grammatical encoding
	Phono-morphological encoding
	Phonetic encoding
	Self-monitoring
Cognitive processing: Interaction pattern + planning time	Non-reciprocal, i.e. semi-direct (computer/tape-based)
	Reciprocal, i.e. direct (face-to-face)
	Planning time allowed
	No planning time

Features of the Stimulus Task(s)

Domain	**Social**		Work		**Academic**
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	**Instructive**
Content knowledge	General				Specific
Cultural specificity	**Neutral**				Specific
Nature of information	Only concrete	**Mostly concrete**		**Fairly abstract**	Mainly abstract
Presentation	**Verbal**		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	No analysis available and not specified in public domain				
Structural range	No analysis available and not specified in public domain				
Functional range	No analysis available and not specified in public domain				
Topic familiarity	**Familiar**				**Unfamiliar**
Intended Speaker/Listener relationship, inc. interlocutor features	**Speech rate: normal**				
	Variety of accent: not specified				
	Number of speakers: 2				
	Acquaintanceship: not specified				
	Gender: not specified				

Features of the Expected Response(s) – Spoken Output

Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	Instructive
Content knowledge	General				Specific
Cultural specificity	**Neutral**				Specific
Nature of information	Only concrete	**Mostly concrete**		**Fairly abstract**	Mainly abstract
Presentation	**Verbal**		**Non-verbal (i.e. graphs)**		**Both**
Lexical Level; further criteria	Not specified; though Nakatsuhara (2012, p525) reports a study of IELTS test takers' average self-assessment of spoken interaction ability being at B2 level = CEFR level 6/6.5				
Structural range	Not specified; though Nakatsuhara (2012, p525) reports a study of IELTS test takers' average self-assessment of spoken interaction ability being at B2 level = CEFR level 6/6.5				
Functional range	Not specified; though Nakatsuhara (2012, p525) reports a study of IELTS test takers' average self-assessment of spoken interaction ability being at B2 level = CEFR level 6/6.5				
Topic familiarity	**Familiar**				**Unfamiliar**

Estimated reliability of rating spoken output

All rating of IELTS Speaking tests takes place at the test centre by trained examiners

Examiners for the Writing and Speaking sub-tests are recruited and trained in line with agreed standards. They are required to demonstrate that they are marking to standard every two years in addition to on-going monitoring of their performance. Candidates receive scores on a Band Scale from 1 to 9. A profile score is reported for each skill. The four individual scores are averaged and rounded to produce an Overall Band Score. Overall Band Scores and scores for each sub-test (Listening, Reading, Writing and Speaking) are reported in whole bands or half bands.

Reliability of rating is assured through the face-to-face training and certification of examiners and all must undergo a retraining and recertification process every two years. A Professional Support Network (PSN) manages and standardizes the examiner cadre, including face to face examiner monitoring as well as distance monitoring using recordings of the Speaking tests.

A 'jagged profile' system maintains a further check on the global reliability of IELTS performance assessment. Routine targeted double marking identifies the level of divergence (i.e., jagged profile) between Writing and/or Speaking scores and Reading and Listening scores. This process allows for the identification of possible misclassified candidates.

CONFIDENTIAL

The jagged profile system is also combined with 'Targeted sample monitoring' to further identify possible faulty ratings by examiners. Selected centres worldwide are required to provide a sample of examiners' marked tapes. Tapes are then second-marked by a team of IELTS Principal Examiners and assistant Principal Examiners. Principal Examiners monitor for quality of both test conduct and rating, and feedback is returned to each test centre. The outcomes that emerge from these reliability measures feed back into examiner retraining and continually build on quality management and assurance systems for IELTS.

Experimental generalisability studies were carried out as part of the IELTS Speaking and Writing Revision Projects to investigate the reliability of ratings (Shaw, 2004; Taylor & Jones, 2001). More recent G-studies based on examiner certification data showed coefficients of **0.83-0.86** for Speaking.

CONFIDENTIAL

TEST:	Cambridge English: Proficiency (CPE)	COMPONENT:	Speaking

Features of the Test	
Skills focus	The Speaking test is taken face-to-face, with two candidates and two examiners, to provide a more realistic and reliable measure of the ability to use English to communicate. It tests the ability to communicate effectively in face-to-face situations. Candidates are expected to be able to respond to questions and to interact in conversational English.
Task description	The *Cambridge English: Proficiency (CPE)* Speaking test has four parts and is taken together with another candidate. There are two examiners. One of the examiners – the Interlocutor - conducts the test (asks questions, gives out the prompt cards, manages timing, etc). The other examiner – the Assessor - listens to the conversation and takes notes. Part 1 is a Conversation between the interlocutor and each candidate. The examiner asks each of the candidates a series of questions, addressing a question to each candidate in turn, to talk about themselves, expressing opinions or speculating about various topics. Part 2 is a Collaborative Task. The interlocutor gives the candidates some spoken instructions and one or more pictures to look at. First, the two candidates have to talk together in response to a question which focuses on their reaction to aspects of one or more pictures (1 minute). The second part is a decision-making task which elicits discussion between the two candidates (3 minutes). This part focuses on the candidates' skills to sustain an interaction through exchanging ideas, expressing and justifying opinions, agreeing and/or disagreeing, suggesting, speculating, evaluating, reaching a decision through negotiation, etc. Part 3 is a Long Turn and Discussion. The interlocutor gives the first candidate a card with a question and some ideas on it and he or she has to speak for about 2 minutes on their own. Afterwards, the other candidate has to comment and the interlocutor then asks both candidates a question on the same topic. The interlocutor follows the same procedure with the second candidate, and then leads a discussion with both of the candidates. This part involves speaking on one's own for a longer time, expressing and justifying opinions and developing topics.
Timing	16 minutes (for a pair of candidates) (Part 1: 2 minutes; Part 2: 1 minute for talking about pictures and 3 minutes for decision-making part; Part 3: 10 minutes (1 minute per candidate and approx. 4 minutes for discussion).
Scoring and weighting	5 assessment criteria with associated rating scale (0-5) for each one. • Global Achievement score (awarded by Examiner 1 who conducts the test) • Grammatical Resource (awarded by Examiner 2 who observes the test) • Lexical Resource (awarded by Examiner 2 who observes the test) • Discourse Management (awarded by Examiner 2 who observes the test) • Interactive Communication (awarded by Examiner 2 who observes the test) • Pronunciation (awarded by Examiner 2 who observes the test) The 0-5 scale (i.e. 6 bands) includes half-points which can be awarded by the examiners. Verbal band descriptors are provided for bands 1.0, 3.0 and 5.0. Band 3 is seen to represent adequacy at a given level; Band 5.0 represents the top of the range and Band 1.0 is seen as an inadequate performance. The final mark for the Speaking test is achieved by combining Examiner 1's global score with Examiner 2's 5 analytical scores to produce a weighted average, which in turn contributes to the overall score for the examination as a whole when combined with the scores from the Reading, Listening, Writing and Use of English components. Within this process, Examiner 1's score is weighted relative to the 4 subscales used by Examiner 2 so that it makes a more balanced contribution to the overall score when combined.
Cognitive processing: Levels of speaking	**Conceptualisation** **Grammatical encoding** **Phono-morphological encoding** **Phonetic encoding** **Self-monitoring**
Cognitive processing: Interaction pattern +	Non-reciprocal, i.e. semi-direct (computer/tape-based) **Reciprocal, i.e. direct (face-to-face)** Planning time allowed

planning time	**No planning time included**				
Features of the Stimulus Task(s)					
Domain	**Social**		Work		**Academic**
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	**Mostly concrete**		**Fairly abstract**	**Mainly abstract**
Presentation	Verbal		Non-verbal (i.e. graphs)		**Both**
Lexical Level; further criteria	Schmitt (2009) in Taylor (ed, 2011, p155-161) reports findings from an analysis of 10 CPE Speaking task prompts which were considered typical of tasks administered at this level and to constitute a representative sample for the level. All tasks had previously been used in the public domain. The analysis was carried out using Compleat Lexical Tutor (www.lextutor.ca) According to this analysis: • 98.55% of words in the CPE task prompts fell into the range K1- K2, i.e. the first 2,000 most frequent words in English • 2.27% of words in the CPE task prompts could be found on the Academic Word List (Coxhead 2000) • Number of word tokens in the CPE task prompts = 2,685 • Number of word types in the CPE task prompts = 418 • Type:token ratio for the CPE task prompts = 6.42				
Structural range	No analysis found				
Functional range	In the CPE Speaking test tasks are explicitly presented to candidates in terms of the functions which they are required to demonstrate - see the analysis reported by Galaczi and ffrench in Taylor (ed, 2011, pp 163-165). The list of functions for CPE is very similar to that for CAE (i.e. 1 level below the CPE), but a key feature at this level is focused more on speculating, evaluating and developing discussion. The level at which candidates are expected to engage with specific linguistic functions and the breadth and depth of their functional resources are closely linked to the cognitive demands of the task. CPE candidates are expected to be able to deal with the higher cognitive demands of interacting with multiple speakers.				
Topic familiarity	**Familiar**				**Unfamiliar**
Intended Speaker/Listener relationship, inc. interlocutor features	Speech rate: Not specified				
	Variety of accent: Not specified				
	Number of speakers: 3				
	Simulated acquaintanceship: Not specified				
	Gender: Not specified				
Features of the Expected Response(s) – Spoken Output					
Length of input in time					
Domain	**Social**		Work		**Academic**
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	**Mostly concrete**		**Fairly abstract**	Mainly abstract
Presentation	**Verbal**		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	Schmitt (2009) in Taylor (ed, 2011, p155-161) reports findings from an analysis of the spoken output of 3 candidates who took the CPE Speaking test who were considered to represent the middle of the proficiency range at this C2 level. The analysis was carried out using Compleat Lexical Tutor (www.lextutor.ca) According to this analysis: • 97.18% of words in the CAE candidate spoken output fell into the range K1- K2, i.e. the first 2,000 most frequent words in English • 1.41% of words in the CAE candidate spoken output could be found on the Academic Word List (Coxhead 2000) • Mean number of word tokens per examinee = 1,228 • Mean number of word types per examinee = 219 • Type:token ratio for candidate output = 5.61 • Mean word length = 4.18 • Lexical density (content words/total) = 0.43				

Structural range	No analysis found
Functional range	An empirical study which applied O'Sullivan, Weir and Saville's (2002) functional checklist tool to CPE spoken responses confirmed that all 31 micro-functions on the checklist appeared in the spoken output.
Estimated reliability of rating spoken output	
Reliability estimate for the Speaking test is based on inter-rater correlation (i.e. degree of agreement in rating between raters). SEM figures are based on raw scores.	
0.85 (SEM = 4.47) – from 2010 figures (http://www.cambridgeenglish.org/research-and-validation/quality-and-accountability/)	

CONFIDENTIAL

TEST:	Cambridge English: Advanced (CAE)	COMPONENT:	Speaking

Features of the Test	
Skills focus	The Speaking test is taken face-to-face, with two candidates and two examiners, to provide a more realistic and reliable measure of the ability to use English to communicate. It tests the ability to communicate effectively in face-to-face situations. Candidates are expected to be able to respond to questions and to interact in conversational English.
Task description	The *Cambridge English: Advanced (CAE)* Speaking test has four parts and is taken together with another candidate. There are two examiners. One of the examiners – the Interlocutor - conducts the test (asks questions, gives out the prompt cards, manages timing, etc). The other examiner – the Assessor - listens to the conversation and takes notes. Part 1 is a Conversation between the candidates and the interlocutor. The examiner asks questions seeking information about interests, studies, careers, etc. This part involves giving personal information and expressing opinions about various topics. Part 2 is a Long Turn. The interlocutor gives each candidate between two and five photographs and asks them to talk about them. Each candidate is invited to speak about their set of photos for 1 minute without interruption. The interlocutor then asks the other candidate to comment on what the first candidate said for about 30 seconds. The second candidate receives a different set of photographs and the first candidate has to listen and comment when they have finished speaking. The questions about the photographs are written at the top of the page to remind candidates of what they should talk about. This part involves talking on one's own about something: comparing, describing, expressing opinions, speculating. Part 3 is a Collaborative Task. It involves a conversation between the two candidates. The examiner provides some pictures and a task to do. The candidates have to talk together and make a decision. The discussion questions related to the pictures are written at the top of the page to remind candidates what they have to do. This part involves exchanging ideas, expressing and justifying opinions, agreeing and/or disagreeing, suggesting, speculating, evaluating, reaching a decision through negotiation, etc. Part 4 is a Discussion. It involves further discussion between both candidates, based on the topics or issues raised in the task in Part 3. The interlocutor asks both candidates some questions and they discuss them together. This part involves expressing and justifying opinions, agreeing and/or disagreeing.
Timing	15 minutes (for a pair of candidates) (Part 1: 3 minutes; Part 2: 1 minute per candidate, plus a 30-second response from the second candidate; Part 3: 3 minutes; Part 4: 4 minutes).
Scoring and weighting	5 assessment criteria with associated rating scale (0-5) for each one. • Global Achievement score (awarded by Examiner 1 who conducts the test) • Grammar and Vocabulary (awarded by Examiner 2 who observes the test) • Discourse Management (awarded by Examiner 2 who observes the test) • Interactive Communication (awarded by Examiner 2 who observes the test) • Pronunciation (awarded by Examiner 2 who observes the test) The 0-5 scale (i.e. 6 bands) includes half-points which can be awarded by the examiners. Verbal band descriptors are provided for bands 1.0, 3.0 and 5.0. Band 3 is seen to represent adequacy at a given level; Band 5.0 represents the top of the range and Band 1.0 is seen as an inadequate performance. The final mark for the Speaking test is achieved by combining Examiner 1's global score with Examiner 2's 5 analytical scores to produce a weighted average, which in turn contributes to the overall score for the examination as a whole when combined with the scores from the Reading, Listening, Writing and Use of English components. Within this process, Examiner 1's score is weighted relative to the 4 subscales used by Examiner 2 so that it makes a more balanced contribution to the overall score when combined.
Cognitive processing: Levels of speaking	**Conceptualisation**
	Grammatical encoding
	Phono-morphological encoding
	Phonetic encoding
	Self-monitoring
Cognitive processing: Interaction pattern + planning time	Non-reciprocal, i.e. semi-direct (computer/tape-based)
	Reciprocal, i.e. direct (face-to-face)
	Planning time allowed
	No planning time included

Features of the Stimulus Task(s)					
Domain	**Social**		Work		**Academic**
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	**Instructive**
Content knowledge	General				Specific

Cultural specificity	Neutral				Specific
Nature of information	Only concrete	**Mostly concrete**	**Fairly abstract**		Mainly abstract
Presentation	Verbal	Non-verbal (i.e. graphs)	**Both**		
Lexical Level; further criteria	Schmitt (2009) in Taylor (ed, 2011, p155-161) reports findings from an analysis of 15 CAE Speaking task prompts which were considered typical of tasks administered at this level and to constitute a representative sample for the level. All tasks had previously been used in the public domain. The analysis was carried out using Compleat Lexical Tutor (www.lextutor.ca) According to this analysis: • 99.62% of words in the CAE task prompts fell into the range K1- K2, i.e. the first 2,000 most frequent words in English • 1.91% of words in the CAE task prompts could be found on the Academic Word List (Coxhead 2000) • Number of word tokens in the CAE task prompts = 1,312 • Number of word types in the CAE task prompts = 274 • Type:token ratio for the CAE task prompts = 4.79				
Structural range	CEFR level C1?				
Functional range	In the CAE Speaking test tasks are explicitly presented to candidates in terms of the functions which they are required to demonstrate - see the analysis reported by Galaczi and ffrench in Taylor (ed, 2011, pp 163-165). A key feature at this level is the potential for interactional management functions to appear; this is made possible through the face-to-face format of the speaking test and is less likely to be the case in a semi-direct speaking test. The level at which candidates are expected to engage with specific linguistic functions and the breadth and depth of their functional resources are closely linked to the cognitive demands of the task. CAE candidates are expected to be able to deal with the higher cognitive demands of interacting with multiple speakers.				
Topic familiarity	**Familiar**				**Unfamiliar**
Intended Speaker/Listener relationship, inc. interlocutor features	**Speech rate: normal** Variety of accent: not specified **Number of speakers: 3 (or 4 in a trio format)** Acquaintanceship: not specified Gender: not specified				
Features of the Expected Response(s) – Spoken Output					
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	**Instructive**
Presentation	**Verbal**	Non-verbal (i.e. graphs)	Both		
Lexical Level; further criteria	Schmitt (2009) in Taylor (ed, 2011, p155-161) reports findings from an analysis of the spoken output of 5 candidates who took the CAE Speaking test who were considered to represent the middle of the proficiency range at this C1 level. The analysis was carried out using Compleat Lexical Tutor (www.lextutor.ca) According to this analysis: • 97.75% of words in the CAE candidate spoken output fell into the range K1- K2, i.e. the first 2,000 most frequent words in English • 1.80% of words in the CAE candidate spoken output could be found on the Academic Word List (Coxhead 2000) • Mean number of word tokens per examinee = 720 • Mean number of word types per examinee = 127 • Type:token ratio for candidate output = 5.66 • Mean word length = 4.29 Lexical density (content words/total) = 0.44				
Structural range	CEFR level C1?				
Functional range	An empirical study which applied O'Sullivan, Weir and Saville's (2002) functional checklist tool to CAE spoken responses confirmed that all 31 micro-functions on the checklist appeared in the spoken output.				
Estimated reliability of rating spoken output					
Estimate of Speaking test reliability i is based on inter-rater correlation : 0.82. SEM figures are based on raw scores : SEM = 4.31					

TEST:	Pearson Test of English (Academic) (PTE-A)	COMPONENT:	Speaking

Features of the Test	
Skills focus	Speaking skills are tested alongside Writing skills in Part 1 of the 3-part Pearson Test of English Academic as shown below. Section 2 of the speaking and writing part of PTE Academic tests ability to produce spoken English in an academic environment. The Speaking skills tested include: • Speaking for a purpose (to repeat, to inform, to explain) • Reading a text aloud • Supporting an opinion with details, examples and explanations • Organizing an oral presentation in a logical way • Developing complex ideas within a spoken discourse • Using words and phrases appropriate to the context • Using correct grammar • Speaking at a natural rate • Producing fluent speech • Using correct intonation • Using correct pronunciation • Using correct stress • Speaking under timed conditions **Part 1: Speaking and Writing** Section 1 – *Personal introduction* (1 minute) Section 2 – *Read aloud, Repeat sentence, Describe image, Re-tell lecture, Answer short question* (30-35 minutes) Section 3/4 - *Summarise written text* (20 minutes) Section 5 – *Summarise written text* or *Write essay* (10 or 20 minutes) Section 6 – *Write essay* (20 minutes)
Task description	Five item types appear in the speaking section, providing a total of around 40 test items. Total time to complete these items is approx 30-35 minutes, depending on the combination of items in a given test. (Because PTE is an integrated skills test, some task types in Part 1 assess reading and listening skills as well as speaking and writing skills.) *Read aloud* (6-7 items): - a text of up to 60 words appears on screen to be read aloud; timing varies by item depending on text length *Repeat sentence* (10-12 items): - listen to a sentence and repeat it; recording time =3-9 secs; response time = 15 secs *Describe image* (6-7 items): - image appears on screen – describe image in detail; response time = 40 secs *Re-tell lecture* (3-4 items): - listen to or watch a lecture, then retell in own words; recording time = up to 90 secs; response time = 40 secs *Answer short question* (10-12 items): - listen to question and respond with single word or a few words; recording time = 3-9 secs; response time = 10 secs.
Further task focus information	Speaking item types require oral responses using fluent speech, correct intonation, stress and pronunciation that are easily understandable to most regular speakers of the language. PTE recognizes regional and national varieties of English pronunciation to the degree that they are understandable to most regular speakers of the language.
Timing	Approximately 77-93 minutes for the whole of Part 1, of which the Speaking test tasks take up 30-35 minutes (not including the 1-minute Personal Introduction which is not assessed). Approximately 40 items are included in Section 2.
Scoring and weighting	All items in PTE Academic are machine scored. Scores for some items are based on correctness alone, while others are centred on correctness, formal aspects and the quality of the response. Formal aspects refer to the form of the response, e.g. whether it is over or under the word limit for a particular item type, or whether it contains more than one sentence for an item type requiring a one-sentence response. The quality of a response is represented in a number of additional skills called enabling skills. For example, in the item type *Re-tell lecture* the response will be scored on skills such as oral fluency and pronunciation. Scores for item types assessing speaking and writing skills are generated by automated scoring systems (see www.pearsonpte.com). There are two types of scoring: 1) correct or incorrect, and

	2) partial credit (i.e. correct, partially correct or incorrect) The *Read aloud, Repeat sentence, Describe image,* and *Re-tell lecture* tasks are scored with partial credit. The *Answer short question* is scored as Correct/incorrect. The *Personal Introduction* task is not assessed. PTE Academic reports an Overall score, Communicative Skills scores (this includes the Speaking score) and Enabling Skills scores (for Speaking this includes Grammar, Oral fluency, Pronunciation and Vocabulary – those scores on these Grammar and Vocabulary). NB. Scores on Grammar and Vocabulary are informed by the candidate's performance on the Writing tasks in Part 1 of the test. The score range for each skill dimension reported on a scale of 10 – 90 and a skills profile showing results for the Communicative and Enabling Skills is reported on the Test Taker Score Report as well as the Overall score.				
Cognitive processing: Levels of speaking	**Conceptualisation**				
	Grammatical encoding				
	Phono-morphological encoding				
	Phonetic encoding				
	Self-monitoring				
Cognitive processing: Interaction pattern + planning time	**Non-reciprocal, i.e. semi-direct (computer/tape-based)**				
	Reciprocal, i.e. direct (face-to-face)				
	Planning time allowed				
	No planning time				
Features of the Stimulus Task(s)					
Domain	Social		Work		**Academic**
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	**Instructive**
Content knowledge	General				Specific
Cultural specificity	**Neutral**				Specific
Nature of information	**Only concrete**	**Mostly concrete**		**Fairly abstract**	Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		**Both**
Lexical Level; further criteria	CEFR level B1-C1?				
Structural range	CEFR level B1-C1?				
Functional range	CEFR level B1-C1?				
Topic familiarity	**Familiar**				**Unfamiliar**
Intended Speaker/Listener relationship, inc. interlocutor features	**Speech rate: normal**				
	Variety of accent: a mixture of L1/L2 accents represented in spoken input of the interlocutor				
	Number of speakers: 2				
	Acquaintanceship: not specified				
	Gender : not specified				
Features of the Expected Response(s) – Spoken Output					
Domain	Social		Work		**Academic**
Discourse mode	**Descriptive**	**Biographical**	**Expository**	Argumentative	Instructive
Presentation	**Verbal**		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	CEFR level B1-C1?				
Structural range	CEFR level B1-C1?				
Functional range	CEFR level B1-C1?				
Topic familiarity	**Familiar**				**Unfamiliar**
Estimated reliability of rating spoken output					
Estimate of Speaking test reliability: 0.91; SEM = 2.32					

CONFIDENTIAL

TEST:	Internet-based Test of English as a Foreign Language (TOEFL iBT)	COMPONENT:	Speaking
Features of the Test			
Skills focus	The Speaking section is semi-direct, i.e. taken using a PC and a headset, and each candidate speaks into the microphone (attached to the headset) in response to written input on screen and/or oral input via the headset. The candidate's performance is recorded and then sent to be rated by qualified raters. It measures the ability to speak English effectively in academic settings, during class as well as outside the classroom. The tasks in this section resemble the real-life situations that students encounter: • During a class, students are expected to respond to questions, participate in academic discussions, summarize what they read and hear, and express their views on topics under discussion. • Outside the classroom, students participate in casual conversations, express their opinions and communicate with people		
Task description	Candidates are required to speak on a variety of topics that draw on personal experience, campus-based situations and academic content. The Speaking section includes six questions (i.e. tasks). The first two questions are called *Independent Speaking Tasks* because they require to draw entirely on the candidate's own ideas, opinions and experiences. The other four questions are called *Integrated Speaking Tasks* because they require to integrate different English-language skills—listening and speaking, or listening, reading and speaking—just as candidates must during class and outside the classroom. The details of each question is as below: <u>*Independent Tasks*</u> Task 1. Personal Preference: This question asks candidates to express and defend a personal choice from a given category—for example, important people, places, events or activities that you enjoy. Task 2. Choice: This question asks candidates to make and defend a personal choice between two contrasting behaviors or courses of action. <u>*Integrated Tasks: Read, Listen and Speak*</u> Task 3. Campus Situation: This task presents a reading passage (75-100 words) on a campus-related issue and a listening passage (60-80 seconds; 150-180 words) which comments on the issue in the reading passage, and then require candidates to speak. Task 4. Academic Course: This task presents a reading passage (75-100 words) which broadly defines a term, process or idea from an academic subject and a listening passage (60-00 seconds; 150-220 words) which provides examples and specific information to illustrate the term, process or idea from the reading passage. This task requires candidates to combine and convey important information from the reading passage and the lecture excerpt. <u>*Integrated Tasks: Listen and Speak*</u> Task 5. Campus Situation: This task presents a listening passage (60–90 seconds; 180–220 words) which includes a conversation about a student-related problem and two possible solutions. IT then requires candidates to demonstrate an understanding of the problem and to express an opinion about solving the problem. Task 6. Academic Course: This task presents a listening passage (90–120 seconds; 230–280 words) which is an excerpt from a lecture that explains a term or concept and gives concrete examples to illustrate that term or concept. It then requires candidates to summarize the lecture and demonstrate an understanding of the relationship between the examples and the overall topic.		
Timing	20 minutes in total. Preparation time / Response time: • Tasks 1 & 2: 15 sec / 45 sec • Tasks 3 & 4: 30 sec / 60 sec • Tasks 5 & 6: 20 sec / 60 sec		
Scoring and weighting	The overall score is given holistically, although raters listen for various features in the following three analytic categories and assign a single score:		

	1. Delivery: How clear and fluid your speech is, including good pronunciation, natural pacing and natural-sounding intonation patterns. 2. Language Use: How effectively you use grammar and vocabulary to convey your ideas. 3. Topic Development: How fully you answer the question and how coherently you present your ideas. Good responses generally use all or most of the time allotted, and the relationship between ideas and the progression from one idea to the next is clear and easy to follow. Each of six tasks is rated from 0 to 4. The sum is converted to a scaled score of 0 to 30. ETS-certified test scorers rate responses and evaluate how well you develop your topic and deliver your message in English. The response is scored by multiple independent ETS-certified raters.
Cognitive processing: Levels of speaking	**Conceptualisation**
	Grammatical encoding
	Phono-morphological encoding
	Phonetic encoding
	Self-monitoring
Cognitive processing: Interaction pattern + planning time	**Non-reciprocal, i.e. semi-direct (computer/tape-based)**
	Reciprocal, i.e. direct (face-to-face)
	Planning time allowed
	No planning time included

Features of the Stimulus Task(s)					
Domain	**Social**		Work		**Academic**
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	**Instructive**
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	**Mostly concrete**		**Fairly abstract**	Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		**Both**
Lexical Level; further criteria	No analysis found				
Structural range	No analysis found				
Functional range	No analysis found				
Topic familiarity	**Familiar**				**Unfamiliar**
Intended Speaker/Listener relationship, inc. interlocutor features	**Speech rate: Normal speech rate for lectures**				
	Variety of accent: includes native-speaker accents from North America, the United Kingdom, New Zealand or Australia. In the Speaking section, only items 1 and 2 of the six tasks may have accented speech.				
	Number of speakers: 1 or 2 (characters in listening passages)				
	Simulated acquaintanceship: N/A				
	Gender: N/A				

Features of the Expected Response(s) – Spoken Output					
Domain	**Social**		Work		**Academic**
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	**Mostly concrete**		**Fairly abstract**	Mainly abstract
Presentation	**Verbal**		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	Biber and Gray (2013)* which examined the lexical profiles of 960 responses on the independent tasks and 1,919 responses on the integrated tasks. These responses received a score ranging from 1 to 4, but their lexical profiles showed very similar values: - General Service List K1 + K2 list words: 85-88% - AWL words: 3 % - Function words: 9-11%				
Structural range	No analysis found				
Functional range	No analysis found				
Estimated reliability of rating spoken output					

Speaking is based on G-theory and uses G-coefficient (closer to 1 is desirable)
0.88 (SEM = 1.62 on the scale of 0-30 (Speaking)) - data from 2007, found in TOEFL iBT Research Insight Series Vol. 3, p.5.

CONFIDENTIAL

TEST:	Occupational English Test (OET)	COMPONENT:	Speaking

Features of the Test	
Skills focus	The OET Speaking sub-test is a test of the candidate's oral use of English language to communicate in a simulated health-related consultation.
Task description	The consultation is between a patient and a health practitioner, with the candidate taking the role of the practitioner and the interviewer taking the role of the patient or patient's relative or carer (for veterinary science the interviewer is the owner or carer of the animal). The candidate and the interviewer interact face-to-face. The Speaking sub-test consists of two tasks: two simulated consultations in the form of two role-plays. **Structure:** In each interview, the candidate's identity and profession are checked by the interlocutor and there is a short warm-up conversation about their professional background (which is unassessed). **Role-plays:** There are two role-plays, and the candidate has 2-3 minutes to prepare for each. The role-plays take about five minutes each. The candidates receive information for each role-play on a card, which they keep while they do the role-play. They may write notes on the card if they want. The card explains the situation and what they are required to do. If they have any questions about the content of the role-play or how a role-play works, they can ask the interlocutor before starting. Candidates are assessed on their performance in the two roleplays only. The whole interview is recorded and it is this audio recording that is assessed: the interlocutor is not assessing the candidate. The interlocutor follows a script so that the interview structure is closely similar for each candidate. He/she also has detailed information to use in each role-play.
Further task focus information	The Speaking sub-test is specific to each profession, based on a typical workplace situation and the demands of the profession. It is designed to reflect common tasks performed in the workplace. 12 professions are represented: dentistry, dietetics, medicine, nursing, occupational therapy, optometry, pharmacy, physiotherapy, podiatry, radiography, speech pathology, and veterinary science. The medical information needed to carry out the role-play is provided on the candidate card. The role-play is designed so that knowledge of a particular medical condition or treatment will not advantage or disadvantage the candidate in any way. Candidates are assessed on how effectively they deal with the communicative situation on the card, NOT on their knowledge of the specific medical topic.
Timing	(approximately) 20 minutes
Scoring and weighting	A summary of the five assessment criteria against which candidates are assessed in the Speaking subtest is as follows: 1. *Overall Communicative Effectiveness* – including how well candidates are able to maintain meaningful Interaction. 2. *Intelligibility* – including Pronunciation, intonation, stress, rhythm and accent. 3. *Fluency* – including the rate (speed) and natural flow of a candidate's speech. 4. *Appropriateness* – including the use of suitable professional language and the ability to explain in simple terms as necessary; also, how appropriately candidates use language to communicate with the patient, in consideration of the scenario of each role-play. 5. *Resources of Grammar and Expression* – including the accuracy and range of the language candidates use; how effectively and naturally candidates communicate. Candidate performances are scored against the five assessment criteria listed above and each criterion consists of six grade levels. The Speaking sub-test is scored by experienced assessors who receive ongoing training, monitoring and feedback on their performance after each administration of the test. Assessors listen to the audio recording of each interview and give a score from 1 to 6 for each of the five criteria, using a detailed set of level descriptors to guide their decisions. A score of 6 is the highest for each criterion. The assessor gives a set of five scores for each role-play and then decides on a final set of five scores for the overall performance across the two role-plays. The five criteria are equally weighted in the scoring and analysis process. Each candidate's recording is graded by two assessors independently. Neither assessor knows the scores the other assessor gives or the scores awarded to the candidate for other sub-tests. Sets of raw scores provided by each rater are sent to the Language Testing Research Centre (LTRC) where the test analysis program FACETS (Linacre, 2010) relates task difficulty, candidate ability and rater harshness/leniency to each other. This allows it to adjust candidates' scores for the effect of an overly harsh or lenient rater. FACETS produces a fair-score that takes rater harshness and task difficulty into account. It also identifies raters who are overly harsh or lenient, or who rate inconsistently or too conservatively. These raters can then be retrained.

	Following established practice, the 'fair score' generated by the statistical analysis of the two sets of scores from two independent assessors of each candidate's Speaking recording becomes the final grade through a direct conversion.
Cognitive processing: Levels of speaking	**Conceptualisation**
	Grammatical encoding
	Phono-morphological encoding
	Phonetic encoding
	Self-monitoring
Cognitive processing: Interaction pattern + planning time	Non-reciprocal, i.e. semi-direct (computer/tape-based)
	Reciprocal, i.e. direct (face-to-face)
	Planning time allowed
	No planning time

Features of the Stimulus Task(s)					
Domain	**Social**		**Work**		Academic
Discourse mode	**Descriptive**	**Biographical**	**Expository**	Argumentative	**Instructive**
Content knowledge	General				**Specific**
Cultural specificity	**Neutral**				Specific
Nature of information	Only concrete	**Mostly concrete**	**Fairly abstract**		Mainly abstract
Presentation	**Verbal**		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	No analysis available and not specified.				
Structural range	No analysis available and not specified.				
Functional range	No analysis available and not specified.				
Topic familiarity	**Familiar**				Unfamiliar
Intended Speaker/Listener relationship, inc. interlocutor features	**Speech rate: normal**				
	Variety of accent: not specified				
	Number of speakers: 2				
	Acquaintanceship: healthcare professional to patient/relative/carer (simulated scenario)				
	Gender: not specified				

Features of the Expected Response(s) – Spoken Output											
Domain	**Social**				**Work**			Academic			
Discourse mode	**Descriptive**		**Biographical**		**Expository**		**Argumentative**		**Instructive**		
Presentation	**Verbal**				Non-verbal (i.e. graphs)			Both			
Lexical Range	**K1**	**K2**	**K3**	**K4**	**K5**	**K6**	**K7**	**K8**	**K9**	**K10**	
Lexical Level; further criteria	CEFR level B2-C2?										
Structural range	CEFR level B2-C2?										
Functional range	CEFR level B2-C2?										
Topic familiarity	**Familiar**									**Unfamiliar**	

Estimated reliability of rating spoken output

The reliability of OET sub-tests is calculated and monitored regularly. Responses to the Speaking are double-scored and analysed using multi-faceted Rasch measurement (McNamara 1996) with the program FACETS (Linacre 2010). This enables scores adjusted for rater severity to be produced, and provides detailed analyses of rater quality, including consistency. Reports on the satisfactory quality (or otherwise) of the measurement on individual candidates' performances are also produced, thus flagging individual performances for third rating. A statistic called Root Mean Square Error (RMSE), the logit equivalent of the classical Standard Error of Measurement (SEM) is available from the FACETS analysis.

The reliability of the Speaking sub-test given by the above analysis is reported to be **0.89**.

For a recent Speaking sub-test, the RMSE for individual candidates was **0.78** logits (about one fifth of a score point)

TEST:	ECPE	COMPONENT:	Speaking
Features of the Test			
Skills focus	The ECPE assesses linguistic, discoursal, sociolinguistic, and pragmatic elements of the English language. The examination focuses on skills and content at the C2 level of the Common European Framework of Reference (CEFR). Throughout ECPE speaking test, test takers are given the opportunity to: - ask and answer questions; - orally summarize information from written descriptions of people, places, or things; - provide suggestions or recommendations; - explain opinions and decisions; - negotiate a decision; - present a position or decision; and - justify a decision.		
Task description	The ECPE Speaking test is a structured two-on-two (two test takers and two examiners) test. The test consists of five stages that require test takers to interact with each other and with examiners. The five stages of the test build on each other; the linguistic and interactional demands become increasingly more difficult as the test progresses. In Stage 1 (Introduction), Examiner 1 and test takers introduce themselves. Test takers are expected to actively participate in the conversation by providing expanded responses and also by asking each other and Examiner 1 questions. The goal of this stage is for test takers to become comfortable interacting with each other. In Stage 2 (Explaining and Recommending), each test taker is given an information sheet with descriptions of two options (for a total of four different options in the paired format). Test takers are allowed to keep the information sheet during the test and can take notes on their information sheet. However, during this stage, test takers are not permitted to see their partners' information sheet. Test takers are given time to read through the information. They then take turns explaining the descriptions of their two options in their own words to each other. They must listen carefully to each other during the explanations because afterward test takers will make a recommendation to their partner of the best option from the two options presented by their partner. Test takers may take notes while their partner is explaining. At the end of this stage, each test taker decides which of the two options on their information sheet is the best. The goal of Stage 2 is for test takers to learn what all the options are, to make a recommendation to their partner, and to choose one of their own options as the best. In Stage 3 (Consensus Reaching), test takers compare and contrast the options they have individually chosen and discuss the advantages and disadvantages of each option. Test takers are still not allowed to look at each other's piece of paper. The goal of this stage is for the two test takers to come to an agreement on one single option. In Stage 4 (Presenting and Convincing), test takers formally present the option they have chosen to Examiner 2, who is in a position of power (for example, Examiner 2 may be the principal of a school that needs to hire a new teacher). Test takers are given time to collaborate and plan the short presentation. At this point, they may look at each other's paper, if they wish. Each test taker presents different reasons for deciding on a particular option and explains why those reasons are important. The goal of this stage is for the test takers to present and to convince Examiner 2 that the option they have chosen is the best one. In Stage 5 (Justifying and Defending), based on the presentations made during Stage 4, Examiner 2 questions the test takers about the decision they have made and about the reasons for that decision. The goal of this final stage is for each test taker to address questions and challenges presented by Examiner 2.		
Further task focus information	There are two examiners present during the entire test. Examiner 1 conducts Stages 1 through 4. Examiner 2 participates during Stage 4 and Stage 5. Throughout the majority of the test, the participation of the examiners is minimal. During Stage 1 and Stage 5, the examiners will be participants in the speaking activity. However, examiner involvement during Stages 2 through 4 does not extend beyond giving directions and answering questions pertaining to test directions. **TOPICS** Test takers should require no specialized knowledge or experience to engage in the speaking tasks. Topics should be equally accessible to a range of ages and must involve a task that requires a decision-making process (e.g. selecting a television advertising campaign); it should be the type of decision that concludes		

	with a person of authority (e.g. the president of the company) being presented information that might be used to make a final decision. Each of these decision-making tasks is one that may occur in reality.
Timing	Approx. 35 mins in the paired format (Approx. 45 mins in the three-way format) (Stage 1 : 3-5 mins, Stage 2 : 5-7 mins, Stage 3 : 5-7 mins, Stage 4 : 5-7 mins, Stage 5 : 5-7 mins)
Scoring and weighting	A test taker's linguistic ability is evaluated independently by two examiners and separately from the other test taker's ability. The test also assesses a test taker's ability to participate in extended, coherent discourse with the other test taker. This means that a test taker should not try to dominate during any stage of the test. At the same time, a test taker should not play a passive role. Test takers are expected to contribute equally to the speaking activity. The stages of the speaking test are designed so that each test taker is provided ample opportunity to speak individually and to engage in discussions with others. Test takers are evaluated at every stage of the test. The rating scale, a five-band measure, guides raters to attend to a test taker's performance in the following areas: ▪ Discourse and Interaction (Development, Functional Range, and Listening Comprehension) ▪ Linguistic Resources (Range and Accuracy of Vocabulary and Grammar) ▪ Delivery and Intelligibility Each of the two examiners assign six independent scores – one score for each of the evaluation criteria for Stages 1,2, and 3 combined, and three more scores for Stages 4 and 5 combined. As in the other sections of the ECPT, the total score will be converted to allow the final speaking score to be reported on a scale of 0-1000. ECPE Speaking scores (as well as other section scores) are reported in five bands: Honors (H) 840-1000; Pass (P) 750-835; Low Pass (LP) 650-745; Borderline Fail (BF) 610-645; Fail (F) 0-605. ECPE test takers who achieve an average score of 650 or higher from the whole test will be awarded a certificate.
Cognitive processing: Levels of speaking	**Conceptualisation**
	Grammatical encoding
	Phono-morphological encoding
	Phonetic encoding
	Self-monitoring
Cognitive processing: Interaction pattern + planning time	Non-reciprocal, i.e. semi-direct (computer/tape-based)
	Reciprocal, i.e. direct (face-to-face)
	Planning time allowed
	No planning time

Features of the Stimulus Task(s)					
Domain	Social		Work		Academic
Discourse mode	Descriptive	Biographical	Expository	Argumentative	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete		Fairly abstract	Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	*No analysis available and not specified in public domain*				
Structural range	*No analysis available and not specified in public domain*				
Functional range	*No analysis available and not specified in public domain*				
Topic familiarity	Familiar				Unfamiliar
Intended Speaker/Listener relationship, inc. interlocutor features	Speech rate: normal				
	Variety of accent: not specified				
	Number of speakers: 4 (2 examiners and 2 candidates – but each examiner speaks at different stages of the test; if a test centre has an odd number of candidates, a group of 3 will be formed, making the total number of speakers 5)				
	Acquaintanceship: not specified				
	Gender: not specified				

Features of the Expected Response(s) – Spoken Output					
Discourse mode	Descriptive	Biographical	Expository	Argumentative	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete		Fairly abstract	Mainly abstract
Presentation	Verbal		Non-verbal (i.e. graphs)		Both

Lexical Level; further criteria	Plough et al (2001) examined 39 test takers' language elicited by the ECPE Speaking test. The speaking performances were transcribed and then analysed for language complexity (as measured by AS-units), vocabulary range (type token ratio), and range of language functions. Standardized type token ratio= 29.05 (Plough et al, 2011)
Structural range	Clause/AS unit=1.60; Sub-clause/AS unit=0.26 (Plough et al, 2011)
Functional range	

Function	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Total (Frequency)
Acknowledgement (issue, opinion)	0	0	9	0	2	11
Agreement	0	10	63	0	16	89
Comparison	10	51	37	4	6	108 (4%)
Concession	0	3	10	1	11	25
Contrast	12	45	35	2	17	111 (4%)
Confirmation (of information, task)	0	14	1	3	0	18
Convincing	0	0	2	29	0	31
Disagreement	0	4	15	0	17	36
Emphasis	2	13	27	18	17	77
Evaluative comment	25	28	16	10	4	83
Explanation	44	74	74	46	46	284 (11%)
Example	5	4	8	3	4	24
General information	24	0	2	0	1	27
Hedge/Softener	0	6	31	6	19	62
Hypothesis/Speculation	9	27	14	6	25	81
Interactive move (offer [to take] floor, request floor, opinion, information)	2	21	39	4	2	68
Justification	0	0	0	0	44	44
Negotiation	0	2	271	0	0	273 (11%)
New supporting information	0	1	4	10	16	31
Opinion	17	135	109	31	59	351 (14%)
Organization of discourse	2	63	12	39	3	119 (5%)
Offering solution	0	2	4	1	12	19
Paraphrasing	1	10	4	0	1	16
Personal information	137	0	0	1	0	138 (5%)
Preference	0	3	0	1	0	4
Presenting	0	0	0	105	0	105 (4%)
Providing (clarification of information; content; task)	0	8	10	0	1	19
Request (clarification of information/ task/ question; repetition; language)	18	42	9	6	11	86
Recommending	0	39	0	2	0	41
Repetition	0	1	1	0	0	2
Resolution	0	0	0	0	2	2
Rhetorical question	0	0	2	0	0	2
Sentence completion	0	5	4	0	2	11
Summarizing	0	159	4	0	0	163 (6%)
Total						2561

*margin for error on totals +/- 4

(Plough et al, 2011)

Topic familiarity	Familiar					Unfamiliar

Estimated reliability of rating spoken output

- Inter-rater agreement: 88.78% (June 2013), 88.19% (November 2013)
- Pass/fail agreement: 94.58% (June 2013), 94.24% (November 2013)

The final score awarded for each test taker is the results of a discussion by both speaking test examiners who have each independently evaluated the speaking performance and then arrive at a consensus rating (ECPE 2013 Report).

TEST:	GEPT Advanced	COMPONENT:	Speaking
Features of the Test			
Skills focus	GEPT Advanced is aligned to the C1 level (Effective Operational Proficiency) of the CEFR, and test-takers who pass this level have English language abilities which enable them to communicate fluently with only occasional errors related to language accuracy and appropriateness, and to handle academic or professional requirements and situations. The Speaking test in GEPT Advanced is taken face-to-face, unlike the three lower levels of the test that are semi-direct. It offers a paired oral interview, comprising both monologic and dialogic elements, such as self-introduction, presentation of a topic, information exchange and discussion between two candidates. The face-to-face paired interview aims to test the candidates' ability to: - participate in discussion on abstract topics or issues, and - fluently express their opinions about social issues.		
Task description	The GEPT Advanced Speaking test has three parts and is taken together with another candidate (If a test centre has an odd number of candidates, a group of three will be formed). There are two examiners. One of the examiners – the Interlocutor - conducts the test (ask questions, gives out the prompt cards, manages timing, etc) and awards a global score to each of the candidates. The other examiner – the Assessor- listens to the conversation and awards 6 analytic scores, 3 part scores and a global score to each of the candidates. Part 1 is an Interview task. Following greetings and self-introduction of both the interlocutor and assessor, the interlocutor checks both of the candidates' names. The interlocutor then asks the candidates (Candidates A and B) to talk about themselves for one minute each. The two candidates receive different questions related to their personal information, and Candidate A is asked a question first *(e.g. Candidate A: 'Please tell us about your favorite book.'; Candidate B: ' Would you please tell us about your personality?').* The interlocutor then asks the candidates three questions seeking their opinions on personal and social/abstract matters *(e.g. Q1: 'Would you prefer to live by yourself or with roommates? Please explain.' Q2: 'What characteristics do you think a successful leader should have?' Q3: 'If you were to write a book, what subject would you be interested in writing about?').* The three questions are the same for both candidates, but they are asked to respond to the questions individually. The order of responses is altered each time. The first question is answered by Candidate B, followed by Candidate A. The second question is answered by Candidate A first, and the last question is answered by B first. The response time of each question is 30 seconds. Part 2 is a Discussion task. The interlocutor passes out the candidates one prompt card each with both verbal and non-verbal information (e.g. chart, table), but different information is missing in the two prompt cards. Candidates are not permitted to see their partners' cards. The interlocutor sets a situation *(e.g. You and your friends are going biking this weekend. [Candidate A], you have information on three bike paths, and [Candidate B], you have different information on the same bike paths.)* Then, the candidates are given 2 minutes to exchange information to complete their charts. Candidate A is asked to initiate the information exchange. In the next 2 minutes, they discuss a topic related to the prompt card *(e.g. the advantages and disadvantages of each path)*. Candidate B is asked to initiate the discussion. The interlocutor then takes back the cards and asks one question each to the two candidates individually. The questions are related to the same topic. They have one minute each to respond, and Candidate A is asked the first question *(e.g. [Candidate A], what might be some reasons why cycling has become increasingly popular in recent years? / [Candidate B], do you think bicyclists should be required to wear helmets in Taiwan?')* Part 3 is a Presentation task. This part involves individual presentations and summarisations of the partners' talks. The presentation starts from Candidate A. The interlocutor gives Candidate A a card with a question on it *(e.g. Do you think improving the quality of life for senior citizens is commanding more attention now than it did twenty years ago?).* Candidate B also receives a copy. Candidate A is given 2 minutes to plan what he/she is going to say, and is asked to present a talk to the partner (Candidate B) for 2 minutes. Candidate B needs to listen carefully during the presentation and takes notes. After the presentation, Candidate B summarises Candidate A's talk for 45 seconds. The interlocutor then takes back the cards, and the same procedure is repeated for Candidate B to present on a different topic and for Candidate A to summarise the presentation *(e.g. What might be some reasons why the media nowadays focus excessively on gossip and scandal?).*		
Further task focus information	The speaking test is available only for those who passed the first stage of the GEPT Advanced test, i.e. Reading and Listening. This contributes to controlling proficiency levels of the two candidates paired in the test. When a test centre has an odd numbers of candidates, a group of three is formed to do the test.		
Timing	Approx. 25 minutes (Part 1: approx. 6 mins; Part 2: approx. 7 mins; Part 3: approx. 11 minutes)		

Scoring and weighting	Two examiners are present at live examinations (i.e. Interlocutor and Assessor); the interlocutor gives one global score using a holistic scale, and the Assessor gives 6 analytic scores as well as 3 part scores and one global score. The six analytic categories are: - Pronunciation, - Relevance and adequacy, - Lexical use, - Grammatical use, - Fluency, and - Coherence Both the holistic and analytic rating scales have 5 Bands. (see GEPT Advanced Level Research Report, pages S-15 and S-16). Each candidate therefore receives 11 scores in total (i.e. one global score from the Interlocutor and 6 analytic scores, 3 part scores and one global score from the Assessor). In order to pass the GEPT Advanced Speaking test, they must receive a global score of 3 or above from both the interlocutor and assessor, and only one analytical score of 2 is allowed. According to the 2013 score report, the average scores for the first stage were 73.20 for the listening test and 65.06 for the reading test, with a passing rate of 26%. The average band scores for the second stage were 2.2 for the writing test and 2.7 for the speaking test, with a passing rate of 19% (https://www.lttc.ntu.edu.tw/academics/GEPT_ScoreR_Doc/Score%20Data%20Summary%20for%202013%20GEPT.pdf).
Cognitive processing: Levels of speaking	**Conceptualisation**
	Grammatical encoding
	Phono-morphological encoding
	Phonetic encoding
	Self-monitoring
Cognitive processing: Interaction pattern + planning time	Non-reciprocal, i.e. semi-direct (computer/tape-based)
	Reciprocal, i.e. direct (face-to-face)
	Planning time allowed
	No planning time

Features of the Stimulus Task(s)					
Domain	Social	Work	Academic		
Discourse mode	Descriptive	Biographical	Expository	Argumentative	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific
Nature of information	Only concrete	Mostly concrete	Fairly abstract	Mainly abstract	
Presentation	Verbal	Non-verbal (i.e. graphs)	Both		
Lexical Level; further criteria	The interlocutors are trained to strictly follow the interlocutor frame (see https://www.lttc.ntu.edu.tw/GEPT1/Advanced/speaking/%E9%AB%98%E7%B4%9A%E5%8F%A3%E8%AA%AA%E6%B8%AC%E9%A9%97%E6%B5%81%E7%A8%8B.pdf and https://www.youtube.com/watch?v=_0MKdXvda1U), and language complexity of the information is B2 or lower to ensure that all candidates understand the rubrics (Wu, 2015, personal communication).				
Structural range	See above.				
Functional range	See above.				
Topic familiarity	Familiar				Unfamiliar
Intended Speaker/Listener relationship, inc. interlocutor features	Speech rate: not specified				
	Variety of accent: not specified				
	Number of speakers: 3 (or 4 in a trio format)				
	Acquaintanceship: not specified				
	Gender: not specified				

Features of the Expected Response(s) – Spoken Output					
Discourse mode	Descriptive	Biographical	Expository	Argumentative	Instructive
Content knowledge	General				Specific
Cultural specificity	Neutral				Specific

Nature of information	Only concrete	Mostly concrete	Fairly abstract	Mainly abstract
Presentation	Verbal	Non-verbal (i.e. graphs)	Both	
Lexical Level; further criteria	Candidates should be able to use an adequate range of vocabulary and grammar (with occasional errors) to complete the tasks, which is expected to be at B2 to C1 level (Wu, 2015, personal communication).			
Structural range	See above.			
Functional range	All functions in O'Sullivan et al.'s (2001) language function lists are to be elicited: [Informational Functions] -Providing personal information -Expressing opinions -Elaborating -Justifying opinions -Comparing -Speculating/Hypothesizing -Staging/Organization -Describing -Summarizing -Suggesting -Expressing preferences [Interactional Functions] -agreeing -disagreeing -modifying -asking for opinions -persuading -asking for information -conversational repair -negotiating meaning [Managing Interaction] -initiating -changing -reciprocating -deciding			
Topic familiarity	Familiar			Unfamiliar
Estimated reliability of rating spoken output				
The inter-rater reliability indices for the writing and speaking tests are between 0.89 and 0.90, while the discrepancy rate falls between 0.014 and 0.017 (https://www.lttc.ntu.edu.tw/E_LTTC/E_GEPT/quality.htm).				

CONFIDENTIAL

TEST:	ISESOL Level 2 (C1 - Expert)	COMPONENT:	Speaking

Features of the Test

Skills focus

The City & Guilds International Spoken ESOL (ISESOL) examinations constitute a suite of tests of spoken language ability in English. The tests can serve either as a stand-alone examination of spoken English or as a complement to the International ESOL series of examinations (English for Speakers for other Languages) which offer a communicative approach to the testing of listening, reading and writing at 6 levels. The 6 levels were developed with reference to levels A1, A2, B1, B2, C1 and C2 of the Common European Framework of Reference for Languages (CEFR) devised by the Council of Europe.

The International ESOL and International Spoken ESOL tests are administered separately with candidates being awarded separate Certificates for the Written (IESOL) and Spoken (ISESOL) components. However, the two certificates can be combined to form the City and Guilds International ESOL (IESOL) Diploma at a required level, e.g. Expert (C1).

The ISESOL test takes the form of a one-to-one structured interview between a candidate and trained interlocutor. The level of spoken competence expected at the Expert (C1) level is described as follows:

- Can understand a wide range of demanding longer texts, both written and spoken, and recognise implicit meaning.
- Can use the language fluently and spontaneously without much obvious searching for expressions.
- Can respond to extended information and narratives, follow detailed explanations and complex instructions, adapting response to audience, medium and context.
- Can engage in discussion in a variety of situations making clear and effective contributions.
- Can use language flexibly and understand a range of texts of varying complexity and length for social, academic and professional purposes.
- Can produce clear well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
- Can communicate ideas and opinions effectively, using length, format and style appropriate to purpose, content and audience.

Task description

The tasks in the examination are designed to test the use of English in real-life situations and the test contains 4 parts, as described below:

Speaking Part 1: to communicate personal information (3 mins from 01/01/2015)

To communicate opinions and ideas on a variety of topics

- Give and spell name
- Give nationality
- Answer 5 questions

The aim of this part is to settle the candidate and to elicit personal and everyday information. The interlocutor first asks the candidate's name, asks for the spelling of their family name and then asks for the candidate's nationality. It is not necessary for any information to be written down. The interlocutor then selects further questions from the list provided under five topic headings. The questions allow the candidate to respond by giving personal information, ideas and opinions on a range of topics and should produce a natural interaction in the time allowed. The interlocutor may expand the questioning to help the candidate produce sufficient language.

Speaking Part 2: to communicate appropriately in social situations (3 mins from 01/01/2015)

To communicate in real-life situations using a range of functional language to elicit or respond as appropriate, and also to show the ability to use a wide range of language functions

- Two or three situations are presented by the interlocutor and candidates are required to respond to and initiate interactions.

The aim of this part is to test the candidate's use of functional language in a range of real-life situations. The interlocutor and candidate enact at least two situations. The Interlocutor may need to assume a different persona, but the candidate is never required to do so. The given situations will require the candidate to enter into exchanges of greater length and complexity. The interlocutor reads aloud the exact words given for the chosen situation which signals the start of the dialogue. The interlocutor first chooses one from four given situations, to which the candidate responds. The interlocutor then chooses one from four given situations which require the candidate to initiate the interaction. The interlocutor allows the candidate to guide and maintain the interaction. The candidate's contribution is expected to demonstrate an awareness of the tone and register appropriate to the situations.

	Speaking Part 3: to exchange information and opinions (4 mins) To engage in discussion and to justify, challenge and persuade when expressing and eliciting opinion • Hold a short discussion to make a plan, arrange or decide something using written text as the prompt The aim of this part is to test the candidate's ability to use English to give and receive information in order to perform a communicative task. The task topics do not require the candidate to have specialist knowledge but candidates are required to have an awareness of and opinions about social and contemporary issues. The interlocutor and candidate exchange information to perform a task. It is the language used in the interaction that is most important, not the ability to complete the task in the given time. However, the interaction should move towards achievement of the task set which typically takes the form of a discussion or negotiation based on written prompts. The candidate is asked to rank or prioritise, analyse, eliminate or hypothesise. Candidates will be required to expand and exemplify, justify, challenge and ask the interlocutor for clarification. **Speaking Part 4: to present a topic (5 mins from 01/01/2015)** To speak in depth on one topic from a wide range of abstract and complex subjects • After 30 seconds of preparation time talk about a topic provided by the interlocutor and answer follow-up questions The aim of this part is to test a candidate's ability to speak at greater length with minimal participation by the interlocutor. This part of the test gives candidates the opportunity to demonstrate their range and fluency in English. The interlocutor chooses one of three given topics and asks the candidate to talk about it. The candidate is given pencil and paper to make notes during the 30 seconds of preparation time. The Interlocutor ensures that the chosen topic is not one which has already been touched on in the previous three parts of the test. The candidate will be asked to answer follow-up questions to their presentation or may be interrupted with questions from the interlocutor.
Further task focus information	**Speaking syllabus** The candidate is expected to be able to: · interact fluently and spontaneously on all subjects except the most abstract with some degree of searching for expressions or use of avoidance strategies · communicate personal information, opinions, feelings and ideas on all topics except the most complex · communicate in almost all social situations, including those requiring tact and diplomacy, using a wide range of language functions · engage in discussion to express and elicit opinion, persuade someone to a point of view, try to reach agreement or perform a specific task · explain, describe, narrate, account for and report on complex subjects integrating subthemes · speak clearly on complex themes, developing particular points and rounding off appropriately.
Timing	[from Jan 01 2015] 15 minutes in total (Part 1 = 3 mins ; Part 2 = 3 mins ; Part 3 = 4 mins ; Part 4 = 5 mins)
Scoring and weighting	All speaking tests are recorded and, after the examinations have been conducted, the recordings are sent directly to City & Guilds in the form of sound files to be marked centrally by a closed group of raters/examiners. Examiners listen to the recording and mark the candidate based on a set of criteria and descriptors of performance for each level. [NB the Interlocutor in the live speaking test does not assess the candidate's performance.] The cadre of City and Guilds raters/examiners (who do not know the candidates they are marking) are regularly standardised through training to ensure consistency and objectivity of assessment that is benchmarked against the CEFR. From 1 January 2015, the following 4 categories are used for assessing a candidate's spoken performance: - Task fulfilment and Coherence - Accuracy and Range of Grammar - Accuracy and Range of Vocabulary - Pronunciation, Fluency and Intonation Three performance grades are reported: First Class Pass, Pass, and Fail
Cognitive processing: Levels of speaking	**Conceptualisation**
	Grammatical encoding
	Phono-morphological encoding
	Phonetic encoding
	Self-monitoring
Cognitive processing:	Non-reciprocal, i.e. semi-direct (computer/tape-based)
	Reciprocal, i.e. direct (face-to-face)

Interaction pattern + planning time	**Planning time allowed**				
	No planning time				
Features of the Stimulus Task(s)					
Domain	**Social**		Work		Academic
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	**Instructive**
Content knowledge	**General**				Specific
Cultural specificity	**Neutral**				Specific
Nature of information	Only concrete	**Mostly concrete**	**Fairly abstract**		Mainly abstract
Presentation	**Verbal**		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	*No analysis available and not specified in public domain*				
Structural range	*Specified in Grammar Syllabus*				
Functional range	*Specified in Functions Syllabus*				
Topic familiarity	**Familiar**				Unfamiliar
Intended Speaker/Listener relationship, inc. interlocutor features	**Speech rate: normal**				
	Variety of accent: not specified				
	Number of speakers: 2				
	Acquaintanceship: not specified				
	Gender: not specified				
Features of the Expected Response(s) – Spoken Output					
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	Instructive
Content knowledge	**General**				Specific
Cultural specificity	**Neutral**				Specific
Nature of information	Only concrete	**Mostly concrete**	**Fairly abstract**		Mainly abstract
Presentation	**Verbal**		Non-verbal (i.e. graphs)		Both
Lexical Level; further criteria	*Not specified and no analysis available*				
Structural range	*Not specified and no analysis available*				
Functional range	*Not specified and no analysis available*				
Topic familiarity	**Familiar**				**Unfamiliar**

Estimated reliability of rating spoken output

The examination board reports that reliability of the rating of the subjectively scored components (i.e. speaking and writing tests) is monitored via the online standardisation facility of the Scoris system used for online marking. The program provides instant feedback on marker performance based on the scores awarded on the seeded (benchmark) papers. Markers deviating in their performance from the standardised scores of the benchmark papers are excluded from marking and can resume their marking activity after several rounds of retraining and standardisation. Markers receive regular feedback on their performance. Marker performance reports include data on scores awarded on the benchmark papers and a standardisation report. Markers also receive an annual FACETS report on their activities which provides additional information about marker and marking characteristics.

No overall value for inter-rater consistency in the speaking tests was reported by the examination board.

CONFIDENTIAL

TEST:	ISESOL Level 3 (C2 - Mastery)	COMPONENT:	Speaking

Features of the Test

Skills focus

The City & Guilds International Spoken ESOL (ISESOL) examinations constitute a suite of tests of spoken language ability in English. The tests can serve either as a stand-alone examination of spoken English or as a complement to the International ESOL series of examinations (English for Speakers for other Languages) which offer a communicative approach to the testing of listening, reading and writing at 6 levels. The 6 levels were developed with reference to levels A1, A2, B1, B2, C1 and C2 of the Common European Framework of Reference for Languages (CEFR) devised by the Council of Europe.

The International ESOL and International Spoken ESOL tests are administered separately with candidates being awarded separate Certificates for the Written (IESOL) and Spoken (ISESOL) components. However, the two certificates can be combined to form the City and Guilds International ESOL (IESOL) Diploma at a required level, e.g. Mastery (C2).

The ISESOL test takes the form of a one-to-one structured interview between a candidate and trained interlocutor. The level of spoken competence expected at the Mastery (C2) level is described as follows:

- Can understand with ease virtually everything heard or read
- Can summarise information from different spoken or written sources, reconstructing arguments and accounts in a coherent presentation
- Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations

Task description

The tasks in the examination are designed to test the use of English in real-life situations and the test contains 4 parts, as described below:

Speaking Part 1: to communicate personal information (4 mins)

To communicate opinions and ideas on a variety of topics

- Give and spell name
- Give nationality
- Answer 5 questions

The aim of this part is to settle the candidate and to elicit personal and everyday information. The interlocutor first asks the candidate's name, asks for the spelling of their family name and then asks for the candidate's nationality. It is not necessary for any information to be written down. The interlocutor then selects further questions from the list provided under five topic headings. The questions allow the candidate to respond by giving personal information, ideas and opinions on a range of topics and should produce a natural interaction in the time allowed. The interlocutor may expand the questioning to help the candidate produce sufficient language.

Speaking Part 2: to communicate appropriately in social situations (4 mins)

To communicate in real-life situations using a range of functional language to elicit or respond as appropriate, and also to show the ability to use a wide range of language functions

- Two or three situations are presented by the interlocutor and candidates are required to respond to and initiate interactions.

The aim of this part is to test the candidate's use of functional language in a range of real-life situations. The interlocutor and candidate enact at least two situations. The Interlocutor may need to assume a different persona, but the candidate is never required to do so. The given situations will require the candidate to enter into exchanges of greater length and complexity. The interlocutor reads aloud the exact words given for the chosen situation which signals the start of the dialogue. The interlocutor first chooses one from four given situations, to which the candidate responds. The interlocutor then chooses one from four given situations which require the candidate to initiate the interaction. The interlocutor allows the candidate to guide and maintain the interaction. The candidate's contribution is expected to demonstrate an awareness of the tone and register appropriate to the situations.

Speaking Part 3: to exchange information and opinions (4 mins)

To engage in discussion and to justify, challenge and persuade when expressing and eliciting opinion

- Hold a short discussion to make a plan, arrange or decide something using written text as the prompt

The aim of this part is to test the candidate's ability to use English to give and receive information in order to perform a communicative task. The task topics do not require the candidate to have specialist knowledge but candidates are required to have an awareness of and opinions about social and contemporary issues. The interlocutor and candidate exchange information to perform a task. It is the

	language used in the interaction that is most important, not the ability to complete the task in the given time. However, the interaction should move towards achievement of the task set which typically takes the form of a discussion or negotiation based on written prompts. The candidate is asked to rank or prioritise, analyse, eliminate or hypothesise. Candidates will be required to expand and exemplify, justify, challenge and ask the interlocutor for clarification. **Speaking Part 4: to present a topic (5 mins)** To speak in depth on one topic from a wide range of abstract and complex subjects • After 30 seconds of preparation time talk about a topic provided by the interlocutor and answer follow-up questions The aim of this part is to test a candidate's ability to speak at greater length with minimal participation by the interlocutor. This part of the test gives candidates the opportunity to demonstrate their range and fluency in English. The interlocutor chooses one of three given topics and asks the candidate to talk about it. The candidate is given pencil and paper to make notes during the 30 seconds of preparation time. The Interlocutor ensures that the chosen topic is not one which has already been touched on in the previous three parts of the test. The candidate will be asked to answer follow-up questions to their presentation or may be interrupted with questions from the interlocutor.
Further task focus information	**Speaking syllabus** The candidate is expected to be able to: · interact fluently and spontaneously on all subjects except the most abstract with some degree of searching for expressions or use of avoidance strategies · communicate personal information, opinions, feelings and ideas on all topics except the most complex · communicate in almost all social situations, including those requiring tact and diplomacy, using a wide range of language functions · engage in discussion to express and elicit opinion, persuade someone to a point of view, try to reach agreement or perform a specific task · explain, describe, narrate, account for and report on complex subjects integrating subthemes · speak clearly on complex themes, developing particular points and rounding off appropriately.
Timing	[from 1 January 2015] 17 minutes in total (Part 1 = 4 mins ; Part 2 = 4 mins ; Part 3 = 4 mins ; Part 4 = 5 mins)
Scoring and weighting	All speaking tests are recorded and, after the examinations have been conducted, the recordings are sent directly to City & Guilds in the form of sound files to be marked centrally by a closed group of raters/examiners. Examiners listen to the recording and mark the candidate based on a set of criteria and descriptors of performance for each level. [NB the Interlocutor in the live speaking test does not assess the candidate's performance.] The cadre of City and Guilds raters/examiners (who do not know the candidates they are marking) are regularly standardised through training to ensure consistency and objectivity of assessment that is benchmarked against the CEFR. From January 2015, the following 4 categories are used for assessing a candidate's spoken performance: - Task fulfilment and Coherence - Accuracy and Range of Grammar - Accuracy and Range of Vocabulary - Pronunciation, Fluency and Intonation Three performance grades are reported: First Class Pass, Pass, and Fail
Cognitive processing: Levels of speaking	**Conceptualisation**
	Grammatical encoding
	Phono-morphological encoding
	Phonetic encoding
	Self-monitoring
Cognitive processing: Interaction pattern + planning time	Non-reciprocal, i.e. semi-direct (computer/tape-based)
	Reciprocal, i.e. direct (face-to-face)
	Planning time allowed
	No planning time

Features of the Stimulus Task(s)					
Domain	**Social**		Work		Academic
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	**Instructive**
Content knowledge	**General**				Specific

Cultural specificity	**Neutral**				Specific
Nature of information	Only concrete	Mostly concrete	**Fairly abstract**		Mainly abstract
Presentation	**Verbal**	Non-verbal (i.e. graphs)	Both		
Lexical Level; further criteria	*No analysis available and not specified in public domain*				
Structural range	*Specified in Grammar Syllabus*				
Functional range	*Specified in Functions Syllabus*				
Topic familiarity	**Familiar**				**Unfamiliar**
Intended Speaker/Listener relationship, inc. interlocutor features	**Speech rate: normal**				
	Variety of accent: not specified				
	Number of speakers: 2				
	Acquaintanceship: not specified				
	Gender: not specified				
Features of the Expected Response(s) – Spoken Output					
Discourse mode	**Descriptive**	**Biographical**	**Expository**	**Argumentative**	Instructive
Content knowledge	**General**				Specific
Cultural specificity	**Neutral**				Specific
Nature of information	Only concrete	Mostly concrete	**Fairly abstract**		Mainly abstract
Presentation	**Verbal**	Non-verbal (i.e. graphs)	Both		
Lexical Level; further criteria	*Not specified and no analysis available*				
Structural range	*Not specified and no analysis available*				
Functional range	*Not specified and no analysis available*				
Topic familiarity	**Familiar**				**Unfamiliar**
Estimated reliability of rating spoken output					
The examination board reports that reliability of the rating of the subjectively scored components (i.e. speaking and writing tests) is monitored via the online standardisation facility of the Scoris system used for online marking. The program provides instant feedback on marker performance based on the scores awarded on the seeded (benchmark) papers. Markers deviating in their performance from the standardised scores of the benchmark papers are excluded from marking and can resume their marking activity after several rounds of retraining and standardisation. Markers receive regular feedback on their performance. Marker performance reports include data on scores awarded on the benchmark papers and a standardisation report. Markers also receive an annual FACETS report on their activities which provides additional information about marker and marking characteristics. No overall value for inter-rater consistency in the speaking tests was reported by the examination board.					

CONFIDENTIAL

Appendix 8 – Detailed analyses of the selected listening tests (IELTS, CPE, CAE, PTE-A, TOEFL iBT, OET, GEPT Advanced, ECPE, IESOL & ISESOL Level 2 and IESOL & ISESOL Level 3)

CONFIDENTIAL

TEST: International English Language Testing System (IELTS)	**COMPONENT : Listening**

Task Features				
Target level	Designed to test across a range of levels from B1 to C1. Sections 1 and 2 may be below criterial level of C1			
Skills focus	The IELTS Listening test is designed with a view to assessing a range of operations. They include how well candidates: o understand main ideas and detailed factual information o understand the opinions and attitudes of speakers o understand the purpose of an utterance o follow the development of ideas			
Structure	Four different passages to listen to, each with 10 accompanying questions			
Task description	40 questions. A variety of question types is used, chosen from the following types: * 3 option multiple choice * gap filling / form completion * labelling diagrams and maps * multiple matching * 'choose from a list'. Variety of question types avoids favouring candidates with particular cognitive attributes. Quite often there are two different formats within the same section. Responses in gap filling are a maximum of three words and may be numbers or letters. Candidates can lose marks for incorrect spelling and grammar.			
Timing	Around 30 minutes.			
Item density	Approx. one item per 75 words; one item per 24 secs running time.			
Delivery and marking	Paper delivered. Answers transferred to answer sheet. Section 2 marked by trained raters using a mark scheme.			
Weighting	All items equally weighted. Each correct answer receives one mark. Scores out of 40 are converted to the IELTS 9-band scale. Scores are reported in whole and half bands			
Presentation	Only one hearing of the recording permitted Items available before the recording is played, with pre-reading time of 45 secs.			
Cognitive processing: Goal setting	Factual information Constructing meaning in context (relevance, speaker goals, inference) Constructing a discourse model (line of argument, topic links)			
Cognitive processing: Levels of listening targeted by items	Word recognition			
	Lexical access			
	Syntactic parsing			
	Establishing the bare meaning of a utterance			
	Relating new information to the context			
	Inference			
	Making referential links			
	Interpreting speaker attitude and speaker goals			
	Integrating information into a wider record of the whole conversation, lecture or broadcast			
Features of the Input				
Input as text				
	Section 1	**Section 2**	**Section 3**	**Section 4**
Domain	Social interaction (campus theme)	Social presentation	Academic interaction (seminar, tutorial, interview)	Academic lecture
Interaction	Dialogue (2 speakers)	Monologue	Dialogue (2-4 speakers)	Monologue
Discourse mode	Interactional/ transactional * Informational * Instructive * Descriptive	* Informational * Instructive * Descriptive	Transactional * Expository * Discursive	*** Narrative** *** Expository** *** Analytical** * Discursive * Persuasive
Information	General	General	General	Specific
Topic	Campus / student life Directions / travel planning Personal information	Travel information Tour guides Practical details	Planning presentations / seminar discussion / discussing research/	History Wildlife Sports science

			tutorial reviews of material / specialist interview	Environment Criminology Business (sample)
Cultural specificity	Specific to campus	Target culture	Academic conventions	Lecture conventions
Nature of information	Only concrete	Mostly concrete	Concrete Analytical Argumentational	Concrete Analytical Argumentational
Text length	Passages contain up to approx 700 words. Running time 3 to 4 minutes.			
Utterance length	Not specified in the criteria. Sample of recordings suggests up to about 16 words in an utterance			
Information density and complexity	High			
Input as recorded material				
Origin	Scripted and studio recorded.			
Transmission	Audio. No visual support. Context and topic provided in rubric.			
Articulation	(sample) Relatively natural in intonation, pausing etc.			
Speech rate	Dialogue: Standard conversational Monologue: standard presentational with paced delivery			
Accent	Range of accents: British, American, Australian/NZ. Accents may be mixed within a dialogue			
Language content				
Lexical level	**K1** **K2** **K3** K4-K10			
Grammatical level	A high level of syntactic complexity. A high proportion of subordination, both within utterances and across utterances			
Reliability *[An estimate of the reliability of a listening test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a listening test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the listening test as a whole. Commonly used statistics for computing and reporting listening test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*				
KR20	0.91			
SEM	0.39 of a band width on the IELTS scale of 1-9			

CONFID

TEST: Cambridge English: Proficiency (CPE)	COMPONENT: Listening

Task Features

Target level	Recorded content and items targeted at specific levels (C1/C2).
Skills focus	To test students' ability to comprehend general texts and to extract important information from those texts. The CPE Listening test is designed with a view to assessing a range of operations, mainly characterised by the provider in general 'listening for' terms. They include how well candidates listen for : gist –detail – function – purpose – topic – identity of speaker/ addressee,- feeling – attitude – opinion (stated and unstated – specific information - detail – inference - main points - interpreting context
Structure	Section 1: 3 short interactional exchanges, with 2 multiple-choice questions on each. Section 2: Monologue with a sentence completion task of 9 items Section 3: Interactional exchange of 2-3 speakers, with 5 multiple-choice questions Section 4: 5 short thematically linked monologues, which have to be matched against 10 items from a choice of 16 in a multiple-matching task
Task description	30 questions in all. A variety of question types is used, chosen from the following types: * 3 option multiple choice * sentence completion * 4 option multiple choice * multiple matching Variety of question types avoids favouring candidates with particular cognitive attributes Responses in gap filling can be words or phrases. No indication of length in the rubrics or Handbook; but max. length seems to be 4 words. Candidates can lose marks for incorrect spelling and grammar..
Timing	Approx 40 minutes.
Item density	[sample] One item per 30 secs running time (Parts 1 and 2); 40 secs running time (Part 3); 17 secs. (Part 4)
Delivery and marking	Paper-delivered in test centre. Answers transferred to answer sheet. Also available by computer delivery Section 2 marked by trained raters using a mark scheme.
Weighting	All items equally weighted. Each correct answer receives one mark.
Presentation	Two hearings of the recording permitted Items available before the recording is played, with 15 secs pre-reading time for clips in Section 1 and 45 secs to 1 min. elsewhere.
Cognitive processing: Goal setting	Factual information Constructing meaning in context (relevance, speaker goals, inference) Constructing a discourse model (line of argument, topic links)
Cognitive processing: Levels of listening targeted by items *(ital = not targeted)*	*Word recognition* *Lexical access* *Syntactic parsing* **Establishing the bare meaning of a utterance** **Relating new information to the context** **Inference** **Making referential links** **Interpreting speaker attitude and speaker goals** **Integrating information into a wider record of the whole conversation, lecture or broadcast**

Features of the Input

Input as text

	Section 1	**Section 2**	**Section 3**	**Section 4**
Domain	Interviews, discussions, conversations, radio plays, talks, speeches, lectures, commentaries, documentaries, instructions			
Interaction	Dialogue (2 speakers)	Monologue or cued monologue	Dialogue (2-3 speakers)	Monologue
Discourse mode	Interactional/ transactional * Informational * Instructive * Descriptive	* Informational * Instructive * Descriptive	Transactional * Expository * Discursive	*** Narrative** *** Expository** *** Analytical** * Discursive * Persuasive
Information	General	General	General	Specific
Topic	General	General	General	General

Cultural specificity	Non-specific	Non-specific	Non-specific	Non-specific
Abstractness	Mostly concrete	Mostly concrete Analytical Argumentational	Concrete Argumentational	Concrete Analytical Argumentational
Text length and duration	+/- 210 words per clip; 1 min. running time	500-700 words; max 4 mins	500-700 words; max. 4 mins	+/- 120 words per clip; 30 secs running time
Utterance length	Not specified in the criteria. Sample of recordings suggests up to about 16 words in an utterance			
Information density and complexity	High			
Input as recorded material				
Origin	Scripted and studio recorded / Authentic material adapted and re-recorded.			
Transmission	Audio. No visual support. Context and topic provided in rubric.			
Articulation	(sample) Relatively natural in intonation, pausing etc.			
Speech rate	Dialogue: standard conversational / interview rate Monologue: standard presentational rate with paced delivery			
Accent	Range of accents (including regional ones). Accents may be mixed within a dialogue			
Language content				
Lexical level	K1 86.94% K2 5.36% K3 2.50% K4 1.16% K5 0.79% K6-10 1.48% K11-20 0.52% (Elliott & Wilson 2013: 222)			
Grammatical level	A high level of syntactic complexity. A high proportion of subordination, both within utterances and across utterances			
Reliability *[An estimate of the reliability of a listening test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect,, is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a listening test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the listening test as a whole. Commonly used statistics for computing and reporting listening test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*				
KR20	0.90+*			
SEM	0.09* [* previous version of CPE]			

Sources for sampled material: Field, J. (2013) 'Cognitive validity', Elliott, M. & Wilson, J. 'Context validity' and Geranpayeh, A. 'Scoring validity'. In Geranpayeh, A & Taylor, L. (eds) *Examining Listening*. Cambridge: Cambridge University Press

Note: New version of CPE launched in 2013. General specifications relate to the new version; some details (e.g. recording length) may be based on samples of the former version.

CONFIDE

TEST: Cambridge English: Advanced (CAE)		COMPONENT: Listening		
Task Features				
Target level	Recorded content and items targeted at a single level (C1).			
Skills focus	To test students' ability to comprehend general texts and to extract important information from those texts. Candidates are expected to be able to show understanding of some of the following: 'agreement – attitude – course of action – detail– feeling – function – gist – context – man points – opinion – purpose – specific information'. These targets are not linked to specific items or question types			
Structure	Section 1: 3 short interactional exchanges, with 2 multiple-choice questions on each. Section 2: Monologue with a gap-filling task of 8 items Section 3: Interactional exchange of 2-3 speakers, with 6 multiple-choice questions Section 4: 5 short thematically linked monologues, which have to be matched against 10 items from a choice of 16 in a multiple-matching task			
Task description	30 questions in all. A variety of question types is used, chosen from the following types: * 3 option multiple choice * sentence completion * 4 option multiple choice * multiple matching Variety of question types avoids favouring candidates with particular cognitive attributes Responses in gap filling can be words or phrases. No indication of length in the rubrics or Handbook; but max. length seems to be 4 words. Candidates can lose marks for incorrect spelling and grammar.			
Timing	Approx. 40 minutes.			
Item density	[sample] One item per 30 secs running time (Parts 1 and 2); 37 secs running time (Part 3); 17 secs. (Part 4)			
Delivery and marking	Paper-delivered in test centre. Answers transferred to answer sheet. Section 2 marked by trained raters using a mark scheme.			
Weighting	All items equally weighted. Each correct answer receives one mark.			
Presentation	Two hearings of the recording permitted Items available before the recording is played, with 15 secs pre-reading time for clips in Section 1 and 45 secs to 1 min. elsewhere.			
Cognitive processing: Goal setting	Factual information Constructing meaning in context (relevance, speaker goals, inference) Constructing a discourse model (line of argument, topic links)			
Cognitive processing: Levels of listening targeted by items *(ital = not targeted)*	*Word recognition*			
	Lexical access			
	Syntactic parsing			
	Establishing the bare meaning of a utterance			
	Relating new information to the context			
	Inference			
	Making referential links			
	Interpreting speaker attitude and speaker goals			
	Integrating information into a wider record of the whole conversation, lecture or broadcast			
Features of the Input				
Input as text				
	Section 1	**Section 2**	**Section 3**	**Section 4**
Domain	lectures, talks, interviews, discussions, radio broadcasts, speeches, announcements, anecdotes			
Interaction	Dialogue (2 speakers)	Monologue	Dialogue (2-3 speakers)	Monologue
Discourse mode	* Narrative * Descriptive * Expository	* Narrative * Descriptive * Expository * Analytical * Discursive	Transactional * Narrative * Expository * Discursive	* Narrative * Expository * Persuasive
Information	General	General	General	General
Topic	General	General	General	General
Cultural specificity	Non-specific	Non-specific	Non-specific	Non-specific
Abstractness	Mostly concrete	Mostly concrete Analytical	Concrete Argumentational	Concrete Analytical

		Argumentational		Argumentational
Text length and duration	+/- 180 words per clip; 1 min. running time	500-700 words; max 4 mins	500-700 words; max. 4 mins	+/- 120 words per clip; 30 secs running time
Utterance length	Not specified in the criteria.			
Information density and complexity	High			
Input as recorded material				
Origin	Scripted and studio recorded / Authentic material adapted and re-recorded.			
Transmission	Audio. No visual support. Context and topic provided in rubric.			
Articulation	(sample) Relatively natural in intonation, pausing etc.			
Speech rate	Dialogue: Standard conversational / interview Monologue: standard presentational with paced delivery			
Accent	Range of accents (including regional ones). Accents may be mixed within a dialogue			
Language content				
Lexical level	**K1 87.4%** **K2 6.05%** **K3 2.36%** K4 0.68% K5 0.39% K6-10 0.53% K11-20 0.21% (Elliott & Wilson 2013: 222)			
Grammatical level	A high level of syntactic complexity. A high proportion of subordination, both within utterances and across utterances			
Reliability *[An estimate of the reliability of a listening test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a listening test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the listening test as a whole. Commonly used statistics for computing and reporting listening test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*				
KR20	0.90+			

Sources for sampled material: Field, J. (2013) 'Cognitive validity' , Elliott, M. & Wilson, J. 'Context validity' and Geranpayeh, A. 'Scoring validity'. In Geranpayeh, A & Taylor, L. (eds) *Examining Listening*. Cambridge: Cambridge University Press

CONFIDEN

TEST: Pearson Test of English Academic (PTE-A)	**COMPONENT: Listening**
Task Features	
Target level	Scores on test are matched against a range of CEFR levels from A1 to C1
Skills focus	Tests students' ability to comprehend very short texts containing expository material and to use other skills to apply the information derived (including mapping from speech to writing and from writing to speech) *Includes integrated skills testing: listening into written summary and reading into listening* A miscellaneous set of processes and strategies are identified by the provider as targeted by the test: - identifying topic or main ideas – summarizing main idea – identifying supporting points – understanding academic vocabulary –inferring word meaning – identifying words appropriate to the context – understanding implicit and explicit information / concrete and abstract information – classifying and categorising information - following sequencing of information – forming a conclusion – predicting how a speaker will continue – identifying errors in a transcription – identifying speaker's purpose, style, tone or attitude - inferring context, purpose or tone - recognising variations in tone, speed and accent – identifying specific details, facts, opinions – identifying the overall organisation of information and connections between pieces of information – evaluating and synthesising information
Structure	2 sections: 1. Summarise a listening passage in writing (20-30 mins) 2. 15-22 items of short clips followed by a question (23-28 mins)
Task description	Some listening tasks; some **integrated skills** tasks **Section 1.** (2-3 items) Listening into writing (summary of 50-70 words) *Recdg length*: 60-90 secs **Section 2.** (15-22 items) A variety of question types is used: • multiple choice (5-7 options); choose multiple options (2 or more) *Recdg length*: 40-90 secs • type blank words in a written transcript (sample: 6-7 words) *Recdg length*: 30-60 secs • read short texts; select one that correctly summarises listening passage (sample: choose between four written passages of approx 40 words) *Recdg length*: 30-90 secs • 4 option multiple choice; choose a single option *Recdg length*: 30-60 secs • select missing word or words to complete the recording (from 3-5 options) *Recdg length*: 20-70 secs • highlight incorrect words in a transcription of the listening passage *Recdg length*: 60-90 secs • listen to a dictated sentence and type it *Recdg length*: 3-5 secs Variation in question types avoids favouring candidates with particular cognitive attributes; but task types may produce processes that are not representative of these employed in academic listening.
Timing	45- 57 minutes.
Item density	Varies across different formats
Delivery and marking	Computer delivered by internet. Answers machine marked – including evaluation of written summary
Weighting	Weighting of marks within Section 1 and between the two sections is not entirely clear. All items are machine marked – including those in Section 1. *Section 1* (summary): score is based on 'the quality of your writing and on how well your response presents the key points in the lecture'. Specified factors are: content (key points) – length – grammar – vocabulary – spelling. Appear to be weighted towards the language used in the writing task. *Section 2*. All items seem to bear equal weight. However, as with Section 1, skills other than listening are factored into the scores: • fill the blanks: correct spelling as well as correct word • highlight correct summary: reading as well as listening • highlight errors in transcript: reading as well as listening • dictated sentence: correct spelling as well as correct word
Presentation	Only one hearing of the recording permitted
Cognitive processing: **Goal setting** *(ital = limited coverage)*	Lexical-level identification Factual information Mapping from spoken to written texts and vice versa *Constructing meaning in context (relevance, speaker goals, inference)* *Constructing a discourse model (line of argument, topic links)*
Cognitive processing: **Levels of listening targeted by items** *(ital = limited coverage)*	**Word recognition** **Lexical access** **Syntactic parsing** Establishing the bare meaning of a utterance *Relating new information to the context* *Inference* *Making referential links*

	Interpreting speaker attitude and speaker goals	
	Integrating information into a mental record of the whole conversation, lecture or broadcast	
Features of the Input		
Input as text		
	Section 1	**Section2**
Domain	Expository material – extract from academic lecture. Close to written text in style and structure	Social presentation Academic interaction (seminar, tutorial, interview) Academic lecture
Interaction	Monologue	Monologue
Discourse mode	Narrative Expository Discursive	Narrative Expository Discursive
Information	General; not too subject-specific	General; not too subject-specific
Topic	Areas of academic study	Areas of academic study
Cultural specificity	Lecture conventions	Lecture conventions
Nature of information	Concrete Analytical Argumentational	Concrete Analytical Argumentational
Text length	Short clips. Maximum running time: 90 secs.	
Utterance length	Not specified in the criteria.	
Information density and complexity	Medium to high	
Input as recorded material		
Origin	Scripted and studio recorded. Resembles written material in style and structure	
Transmission	Audio. No visual support. *No information about content and topic provided in rubric*. Computer delivered test	
Articulation	Often a read aloud style	
Speech rate	Monologue: standard presentational rate with paced delivery,	
Accent	Range of L1 accents.	
Language content		
Lexical level	**K1** **K2** **K3** K4-K10	
Grammatical level	A high level of syntactic complexity. A high proportion of subordination, both within utterances and across utterances. Generally resembles the grammatical sentence-based structure of written discourse.	
Reliability *[An estimate of the reliability of a listening test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a listening test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the listening test as a whole. Commonly used statistics for computing and reporting listening test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*		
KR20	Not found	

TEST: Internet-based Test of English as a Foreign Language (TOEFL iBT)	COMPONENT: Listening

Task Features	
Target level	Lectures: Recorded content and items targeted at specific level (high C2). Conversation: Recorded content and items targeted at level C1/C2
Skills focus	To test the ability to comprehend subject-specific texts and texts relating to academic contexts – and the ability to extract information (including detail) from those texts. Specifications cite the content of the recorded material: lectures, classroom discussions and conversations. Questions require candidates to report complex facts, to interpret facts, to use inference and to interpret speaker intentions (often pragmatic) . Specifications cite the following skills: ☐ identifying the main idea ☐ determining purpose ☐understanding detail ☐ determining attitude ☐understanding implication ☐making inferences ☐connecting content
Structure	6-9 tasks 4-6 lectures • A minimum of 2 lectures, followed by questions • A minimum of 2 lectures, with student comments and questions interspersed, followed by questions 2-3 conversations about academic study or academic contexts. • A minimum of 2 questions between two people - such as two students or a student and a professor, a librarian, a counsellor, etc; followed by questions
Task description	40-51 questions. All are 4-option multiple choice (a few using graphics rather than text). The question sometimes requires candidate to choose 2 out of 6 options. Use of only one question format may favour candidates with particular cognitive attributes *Lectures* are followed by 6 MCQs (including one where a clip from the lecture is replayed for interpretation) *Conversations* are followed by 5 MCQs (including one where a clip is replayed for interpretation)
Timing	60-90 minutes.
Item density	[sample] Lectures: One item per 24 secs running time plus one item on clip Conversations: One item per 30 secs running time
Delivery and marking	Computer-delivered via internet. Answers computer marked
Weighting	All items equally weighted. Each correct answer receives one mark.
Presentation	Only one hearing of the recording permitted. Candidate listens to the recording while taking notes. Questions are then shown after the recording. One question has a repeated clip from the main recording attached to it.
Cognitive processing: Goal setting	Factual information Constructing meaning in context (relevance, speaker goals, inference) Constructing a discourse model (line of argument, topic links)
Cognitive processing: Levels of listening targeted by items	Word recognition
	Lexical access
	Syntactic parsing
	Establishing the bare meaning of a utterance
	Relating new information to the context
	Inference
	Making referential links
	Interpreting speaker attitude and speaker goals
	Integrating information into a wider record of the whole conversation, lecture or broadcast

Features of the Input			
Input as text			
	Lectures		**Interactional and transactional conversations**
Domain	Lectures Lectures with student interventions		Classroom discussions and conversations in an academic context
Interaction	Monologue	Monologue with interventions by 1-3 student voices and lecturer responses	Dialogue (2-3 speakers)
Discourse mode	* Narrative * Expository * Analytical * Discursive		Interactional / Transactional * Instructional * Advisory / explanatory * Descriptive

	* Persuasive * Explanatory	* Expository * Discursive
Information	Subject-specific	Practical and social needs, discussion of course content
Topic	Academic level, specialised. Biology, social science, anthropology, American history, literature, art history, life sciences, drama, political science, physics, agriculture	Campus facilities, project planning, course content planning
Cultural specificity	Culture of academic discourse	US campus environment
Abstractness	Complex argumentation Analytical Detailed	Concrete
Text length and duration	500-800 words; max 5 mins	500-700 words; Approx. 2 mins
Utterance length	Not specified in the criteria.	
Information density and complexity	Very information dense with complex lines of argument. Lack of redundancy and repetition. Heavy conceptual demands and memory load	Information dense. May require understanding of context Lack of redundancy and repetition.
Input as recorded material		
Origin	Scripted and studio recorded. Some may be authentic material that has been re-recorded; some has the information density of a reading passage that has been converted to a spoken monologue.	
Transmission	Audio. Visual support by means of still photos. Limited information provided in rubric.	
Articulation	(sample) Relatively unnatural in intonation, pausing etc.	
Speech rate	Lecture: Quite often delivered at a high speech rate for a presentational context. Delivery not paced to counter the information density. Conversation: standard to fast rate for interactional discourse.	
Accent	US accents. From March 2013, includes some GB and Aus/NZ. Accents may be mixed within a dialogue	
Language content		
Lexical level	**K1** **K2** **K3** K4-K10	
Grammatical level	A very high level of syntactic complexity, more closely resembling a written text in the lecture material. A high proportion of subordination, both within utterances and across utterances.	
Reliability *[An estimate of the reliability of a listening test, which usually consists of discrete test items that are dichotomously scored (correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a listening test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the listening test as a whole. Commonly used statistics for computing and reporting listening test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*		
KR20	Not found	

Source for sampled material: http://www.ets.org/toefl/ibt/prepare/quick_prep/

TEST: Occupational English Test (OET)	COMPONENT: Listening
Task Features	
Target level	Level C2. Aimed at higher levels of health professionals and medical students
Skills focus	Test specifically designed for health professionals, regardless of professional level *Includes integrated skills testing: listening into notetaking* In Part A, candidates are required to demonstrate that they can: • understand the content of a consultation between a health practitioner and a patient/client. • extract relevant information in a consultation. • take notes while listening for relevant information in real time. In Part B, candidates are required to demonstrate that they can: • understand the main ideas and important information in a short academic talk on a health-related topic • understand inferred meaning in a short talk, including recognising the communicative functions of utterances and inferring the meaning of unfamiliar lexical items from the context.
Structure	Two recordings of around 15 mins are divided into short sections (usually around 100 to 200 words / 30 to 60 secs in Part A and 80 to 150 / 25 to 45 secs in Part B). Each has a multi-item task for candidate to complete The test consists of two parts (A and B) with 20-28 tasks overall. Each task contains from 2 to 9 items. Sample paper contains a total of 81 items in Part A and 66 in Part B. Mean of 6 items per task. Part A (20–25 minutes) assesses candidates' ability to follow facts during a consultation between a health professional and a patient. Part B (20–25 minutes) assesses candidates' ability to understand a lecture on a health-related topic
Task description	Part A: responses are in the form of notes that the candidate writes under headings provided. The number of bullet points under each heading indicates how many points are expected. Part B: A range of open-ended and fixed choice task types is used. They include: • multiple-choice questions • sentence completion • short-answer questions • summary completion • lecture notes completion • table/flow-chart/mind-map completion • word replacement response • matching responses Candidates hear each recording once and are expected to write their answers while listening.
Timing	Approx. 50 minutes with 20-28 tasks in total, each consisting of up to 9 items. Total recording time is 30 mins. Test takers are given 60 secs before beginning to read the whole Part and 20 secs to re before listening (= approx 9 mins). There is thus pressure on candidates to complete answers as they listen; and little time for checking answers (5-20 secs in Part A and 5-20 secs in part B, though at the end 2 mins is allowed for a general review.
Item density	Items appear extremely dense: in Part A average of 1 per 11 secs of recording running time; in Part B average of 1 per 13.5 secs. A total of 147 items in the sample examined.
Delivery and marking	Paper-delivered. Answers written while listening. Marked by assessors. Problematic scripts are dealt with by an experienced assessor and all critical borderline scripts are double-marked. Before each administration, assessors attend a formal meeting to discuss the marking guide.
Weighting	All items equally weighted. Each correct answer receives one mark. The score for each Part is converted so that each part accounts for 50% of the total Listening score, which is then converted to Grades A to E. Candidates need a Grade B (i.e. 'High level of performance; i.e. able to use English with fluency and accuracy adequate for professional needs') across all skills to pass the test.
Presentation	Only one hearing of the recording permitted Items available before the recording is played, with 60 secs pre-reading time for each Section A and B (total of 81 items in A and 66 in B) – plus 10-20 secs immediately before listening
Cognitive processing: Goal setting	Factual information Constructing meaning in context (relevance, speaker goals, inference) Constructing a discourse model (line of argument, topic links)
Cognitive processing: Levels of listening targeted by items	Word recognition
	Lexical access
	Syntactic parsing
	Establishing the bare meaning of a utterance
	Relating new information to the context

(ital = limited coverage)	**Inference**	
	Making referential links	
	Interpreting speaker attitude and speaker goals	
	Integrating information into a record of the whole conversation or lecture	
Features of the Input		
Input as text		
	Part A	**Part B**
Domain	Social interaction (clinic-based)	Academic-style lecture
Interaction	Dialogue (2 speakers)	Monologue
Discourse mode	* Informational * Interactional * Instructive * Descriptive	* Informational * Descriptive
Information	Highly specific to doctor-patient context Subject-specific terminology	Specific to Medicine Subject-specific terminology. Some background assumed
Topic	Patients' accounts of health problems	Background on medical conditions
Cultural specificity	Specific to clinics Some local culture (cricket) and idioms	Lecture conventions
Nature of information	Concrete Analytical	Mostly concrete Analytical
Text length	15 min recordings divided into short sections of: (Part A) around 100 to 200 words / 30 to 60 secs (Part B) 80 to 150 words / 25 to 45 secs	
Utterance length	-	
Information density and complexity	Information dense in Part A. Very information dense in Part B, though some hesitations and pauses	
Input as recorded material		
Origin	Studio recorded; but sounds very naturalistic (improvised?)	
Transmission	Audio. No visual support. Context and topic provided in rubric.	
Articulation	Very natural	
Speech rate	Speech rate not very controlled and sometimes reaches a high speed both in dialogue and in monologue, considering information density.	
Accent	Australian-accented	
Language content		
Lexical level	Not specified. Much subject- specific terminology	
Grammatical level	Not specified	
Reliability *[An estimate of the reliability of a listening test, which usually consists of discrete test items that are dichotomously scored (i.e. correct or incorrect), is typically reported using an estimate of internal consistency on a 0-1 scale. The internal consistency of a listening test indicates the degree to which candidates' scores on the individual items are consistent with their total score on the listening test as a whole. Commonly used statistics for computing and reporting listening test reliability are Kuder-Richardson formula KR-20, and Cronbach's alpha coefficient.]*		
KR20	0.89	

Test	GEPT Advanced	Component	Listening
Task Features			
Target level	Recorded content and items targeted at specific level (C1). In fact, much of the material (in terms of content, lexical level of recorded material and of items, speed of delivery, information density, cognitive complexity of tasks) is at level C2.		
Skills focus	To test students' ability to handle academic or professional requirements and situations. This component tests the ability to understand monologues and dialogues of varying lengths, including 'conversations on all sorts of topics, professional lectures, speeches and news reports'		
Structure	Section 1: 15 short conversational exchanges or monologues, followed by an oral question on each. Section 2: 2 long conversations, with 12 items in total (mainly free response but some multiple choice) Section 3: 2 presentational monologues, with 13 items in total (mainly free response but some multiple choice)		
Task description	40 questions in all. Section 1: oral question with written 4 option multiple choice Sections 2 and 3: free written response (word, phrase or complete sentence) to an open-ended question in writing. Also limited number of written questions with 4 option multiple choice. Free responses can be words. phrases or sentences. *No indication of maximum length in rubrics or Handbook.* Limited options given on answer sheet; responses are marked by trained raters. Unclear if candidates are penalised for incorrect spelling and grammar.		
Timing	Approx 45 minutes.		
Item density	[sample] One item per 20 secs running time (Part 1); 20-28 secs running time (Part 2); 17-30 secs. (Part 3)		
Delivery and marking	Paper-delivered in test centre. Answers written on answer sheet. Sections 2 and 3 marked by trained raters. Answers are double marked		
Weighting	All items equally weighted. Each correct answer receives one mark.		
Presentation	Only one hearing of each recording permitted Section 1: Oral question after recording with written MCQ options. No pre-reading time. Sections 2 and 3: Items available before each recording is played, with 30-45 secs to read up to 8 items		
Cognitive processing: **Goal setting**	Factual information (often inferred) Constructing meaning in context (relevance, speaker goals) *Heavy reliance on inference* Constructing a discourse model (topic links)		
Cognitive processing: **Levels of listening targeted by items** *(ital = not targeted)*	*Word recognition*		
	Lexical access		
	Syntactic parsing		
	Establishing the bare meaning of a utterance		
	Relating new information to the context		
	Inference		
	Making referential links		
	Interpreting speaker attitude and speaker goals		
	Integrating information into a wider record of the whole conversation, lecture or broadcast		
Features of the Input			
Input as text			
	Section 1	**Section2**	**Section 3**
Type	Conversations, broadcast monologues, news, announcements	Conversation Interview	Lecture Presentation Documentary
Interaction	Dialogue (2 speakers) Monologue	Dialogue (2 speakers)	Monologue
Discourse mode	Interactional/ transactional * Narrative * Informational * Expository * Discursive	Transactional * Expository	Transactional * Expository * Discursive
Information	General /specific	General	Topic-specific
Topic	General	General	Specific
Cultural specificity	Non-specific	Non-specific	Non-specific

Abstractness	Mostly concrete	Mostly concrete Analytical	Complex exposition Argumentative
Text length and duration	+/- 70 words per clip; 20 secs running time	500 words; 2.5 mins	500 words; 2.5 mins
Utterance length	Not specified in the criteria. Sample of recordings suggests up to about 20 words in an utterance		
Information density and complexity	High; sometimes very high		
Input as recorded material			
Origin	Scripted and studio recorded / Authentic material adapted and re-recorded.		
Transmission	Audio. No visual support. Limited information provided in spoken rubric.		
Articulation	(sample) Not wholly natural. Some characteristics of read-aloud scripted material.		
Speech rate	(sample) Dialogue: quite rapid conversational / interview rate Monologue: quite rapid presentational rate without paced delivery		
Accent	(sample) Standard American only. Limited number of voices used.		
Language content			
Lexical level	**Recorded material in sample contains a very high proportion of low-frequency lexical items, including many that are technical or highly subject-specific. Wording of questions (including multiple-choice options) includes a high proportion of low-frequency lexical items, including many that are specialised or idiomatic.** **Correct answers to questions may thus be dependent upon recognising low-frequency or even specialist lexical terms**		
Grammatical level	Sections 1 and 2 reflect the structure of conversational speech A high level of syntactic complexity in Section 3 including subordination		
Reliability			
KR20	0.88		
SEM	2.65 items (Research Report 2002)		

CONFIDE

Test	**ECPE**		**Component**	**Listening**
Task Features				
Target level	Recorded content and items targeted at a single level (C2). But some of the items and tasks would appear to be at C1 level.			
Skills focus	To test students' ability to 'understand conversational questions, … grasp both the gist and details of a conversation delivered in Standard American English general texts and … understand extended spoken discourse' * Global skills: understanding main idea - identifying speaker's purpose - synthesizing ideas from different parts of the text *Local skills: identifying supporting detail - understanding vocabulary - synthesizing details -recognizing restatement * Inferential skills: understanding rhetorical function - making an inference - inferring supporting detail - understanding pragmatic implications			
Structure	Section 1: 15 short interactional exchanges, with a 3 option multiple-choice question on each. Section 2: 20 uncontextualised single sentence stems (statements or questions) with 3 option multiple choice of best response Section 3: 3 sections from radio interviews with five 3-option multiple-choice questions on each			
Task description	50 questions in all. A single question type (3 option multiple-choice) is used Though this provides objective scores, it runs the risk of favouring candidates with particular cognitive attributes			
Timing	35-40 minutes.			
Item density	Part 1: One item per clip of 7 to 15 secs; Part 2: one item per question of 3 to 4 secs; Part 3: one item per 36 secs			
Delivery and marking	Paper-delivered in test centre. Space for notes on question sheet. Answers on answer sheet.			
Weighting	All items equally weighted. Each correct answer receives one mark.			
Presentation	Only one hearing of the recording permitted Items available before the recording is played. Apparently, no pre-reading time; but space on question sheet for notes.			
Cognitive processing: Goal setting	Gist Factual information Constructing meaning in context (relevance, speaker goals, inference) Implied meaning Constructing a discourse model (line of argument, topic links)			
Cognitive processing: Levels of listening targeted by items *(ital = not targeted)*	*Word recognition*			
	Lexical access			
	Syntactic parsing			
	Establishing the bare meaning of a utterance			
	Relating new information to the context			
	Inference; indirect meaning			
	Making referential links			
	Interpreting speaker attitude and speaker goals			
	Integrating information into a wider record of the whole conversation, lecture or broadcast			
Features of the Input				
Input as text				
	Section 1	**Section 2**	**Section 3**	
Type	Everyday conversations	Decontexualised questions	Radio interviews Lectures	
Domain	Everyday conversation, anecdotes, everyday questions, interviews, lectures, talks, discussions, announcements * Public places, social networks outside the home, * Occupational workplace settings (offices, workshops, conferences), etc. * Educational setting (schools, colleges, classrooms, residence halls, etc.			
Interaction	Dialogue (2 speakers)	Monologue	Dialogue (up to 4 speakers: interviewer and specialist)	
Discourse mode	Interactional * Narrative * Descriptive * Instructional * Expository	Questions with implicit context	Transactional * Narrative * Descriptive * Expository * Discursive	
Information	General (partly campus related / professional)	General	Subject-specific	
Topic	General	General	Specific	
Cultural specificity	Mainly non-specific; some US	Non-specific	Non-specific	

	relevant			
Abstractness	Mostly concrete; sometimes with implicit speaker intentions	Concrete	Concrete Analytical Abstract	
Text length and duration	20-50 words per clip; 5 to 15 secs. running time	10-15 words 2-4 secs	550 words; 3-4 mins	
Utterance length	Not specified in the criteria.			
Information density and complexity	High			
Input as recorded material				
Origin	Scripted and studio recorded / Authentic material adapted and re-recorded.			
Transmission	Audio. No visual support. No context provided in rubric. *All recorded material is without prior context*			
Articulation	(single sample) Sections 1 and 2: Relatively natural in intonation, pausing etc. Section 3: read-aloud style - not natural presentational style			
Speech rate	Sections 1 and 2: Relatively fast conversational style Section3: standard presentational with paced delivery			
Accent and voice	Accent: Only Standard American. Section 1: 2 voices Section 2: One voice for entire set of questions reduces candidate's need to keep adjusting to new voices. Section 3: Up to 4 voices			
Language content				
Lexical level	Generally mainstream lexis at a frequency level suitable for C level candidates. Some subject-specific terminology in Section 3; but responses to questions are not dependent upon knowing or interpreting it			
Grammatical level	Some syntactic complexity in Section 3; subordination within and across utterances			
Reliability				
KR20	0.86/0.87			
SEM	37.15/36.24 [June 2013/Nov 2013]			

CONFID

Test	**IESOL Mastery**	**Component**	**Listening**

Task Features	
Target level	Recorded content and items targeted at a single level (C2).
Skills focus	The test taker is expected to: * follow short conversations (formal and informal) in a range of situations - understanding gist, context, purpose, function and recognising attitudes, feelings and opinions; * understand sophisticated narratives, explanations and arguments; * recognise the function of short utterances which may feature idiomatic or colloquial English * follow a discussion (possibly academic or professional) and identify gist, detail, purpose and key ideas; distinguish fact and opinion * extract key information from announcements, media broadcasts, presentations and lectures on abstract and concrete topics of general, academic and professional interest * follow a complex argument even when it is not clearly structured * recognise how changes of intonation, pitch or stress affect meaning and reflect mood and attitude * understand a very wide range of vocabulary including terms used in academic and professional discourse. * understand the main ideas of complex extended discourse even when the topic is unfamiliar. * extract most points of detail from extended discourse even when the topic is unfamiliar.
Structure	Section 1: 8 sentences, each with a different function and context. Choose appropriate response Section 2: 3 conversations consisting of two speakers. Two items for each; identify aspects of conversation Section 3: monologue presentation with gap-filling task (1-5 words) Section 4: Discussion between two speakers. Eight 4-option multiple choice questions
Task description	30 items in total Section 1: Identify appropriate response to cue: 8 four-option multiple choice Section 2: Identify aspects of conversation (topic, purpose, context, speakers, gist, relationship between speakers, roles, functions, attitudes, feeling and opinions): 6 four-option multiple choice Section 3: Gap-filling task: identify factual information. 8 gaps in set of notes. Section 4: Listen for gist, examples, fact, opinion, contrast, purpose, key ideas, attitude, cause and effect. 8 four-option multiple choice items
Timing	30 minutes.
Item density	Part 1: 1 item per 5 sec utterance Part 2: 1 item per 25 secs Part 3: 1 item per 22-30 secs Part 4: 1 item per 22-30 secs
Delivery and marking	Paper-delivered in test centre. Answers on question sheet.
Weighting	All items equally weighted. Each correct answer receives one mark.
Presentation	Two hearings of the recording permitted Items available before the recording is played. Pre-reading and completion time allowed.
Cognitive processing: Goal setting	Local factual information Constructing meaning in context (relevance, speaker goals, speaker attitude, inference) Constructing a discourse model (topic links)
Cognitive processing: Levels of listening targeted by items *(ital = not targeted)*	*Word recognition*
	Lexical access
	Syntactic parsing
	Establishing the bare meaning of a utterance
	Relating new information to the context
	Inference; indirect meaning
	Making referential links
	Interpreting speaker attitude and speaker goals
	Linking information

Features of the Input				
Input as text				
	Section 1	**Section 2**	**Section 3**	**Section 4**
Domain	Conversational cues – highly idiomatic	Short everyday conversations	Factual presentation	Discussion, expression of views /attitudes
Interaction	Monologue cue	Dialogue 8 turns	Monologue	Dialogue (up to 3 speakers)
Discourse mode	Decontextualised cue	Interactional * Narrative * Descriptive * Instructional	Expository	Transactional * Narrative * Descriptive * Expository

				* Discursive
Information	Indirect cues	General	Subject-specific	General
Topic	General	General	Specific	General
Cultural specificity	UK specific idioms	Non-specific	Non-specific	Non-specific
Abstractness	Decontextualsied	Concrete	Concrete Analytical Abstract	Concrete Discursive
Text length and duration	6-12 words per clip; 5 secs. running time	8 turns 135-160 words 50 secs running time	550-600 words; 3-4 mins	34 turns 650 words 3-4 mins
Utterance length	Not specified in the criteria.			
Information density and complexity	High			
Input as recorded material				
Origin	Scripted and studio recorded.			
Transmission	Audio. No visual support. No contextualisation or topic setting provided in rubric.			
Articulation	(single sample) Relatively natural in intonation, pausing etc.			
Speech rate	Section 1,2,4: Slow-medium conversational rate Section3: standard presentational rate with paced delivery			
Accent and voice	Only standard British Section 1: voice change minimalised across 8 items Section 2: 2 voices. Section 3: Monologue Section 4: 3-4 voices			
Language content				
Lexical level	Section 1 makes excessive use of low frequency idioms which candidates would be unlikely to know Elsewhere, mainstream lexis at a frequency level suitable for C level candidates. Avoids excessive subject-specific terminology in Section 3			
Grammatical level	Some syntactic complexity in Section 3; subordination within and across utterances			
Reliability				
KR20	0.86/0.87			
SEM	37.15/36.24 [June 2013/Nov 2013]			

CONFI

Test IESOL E	**Component Listening**
Task Features	
Target level	Recorded content and items targeted at a single level (C1).
Skills focus	The test taker is expected to: * follow extended speech delivered at natural speed even when the meaning may be implicit rather than explicit * follow short conversations (formal and informal) in a range of situations - understanding gist, context, purpose, function and recognising attitudes, feelings and opinions; * understand complex narratives, sequences and explanations; * recognise the function of short utterances which may contain idiomatic English; * follow a discussion and identify gist, detail, purpose and key ideas; distinguish fact and opinion * extract key information from announcements, media broadcasts, presentations and lectures on abstract and concrete topics of general interest * follow a complex argument. * recognise how intonation, pitch or stress affect meaning and reflect mood and attitude; * understand a broad range of vocabulary including idiomatic and colloquial expressions;. * recognise degrees of formality in different types of utterances used by speakers in a range of situations;. * understand the main ideas of complex extended discourse even when the content is not well structured; * extract most points of detail from extended discourse even when the content is not well structured.
Structure	Section 1: 8 incomplete conversations. Choose appropriate continuation. May include idiomatic expressions, changes of register, intonation indicating attitude. Section 2: 3 conversations consisting of two speakers. Two items for each; identify aspects of conversation Section 3: monologue presentation with gap-filling task Section 4: Discussion between two speakers. Eight 4-option multiple choice questions
Task description	30 items in total Section 1: Identify appropriate end of conversation: 8 four-option multiple choice. *Lack of clear task focus* Section 2: Identify aspects of conversation (topic, purpose, context, speakers, gist, relationship between speakers, roles, functions, attitudes, feeling and opinions): 6 four-option multiple choice Section 3: Gap-filling task: identify factual information. 8 gaps in set of notes. Section 4: Listen for gist, examples, fact, opinion, contrast, purpose, key ideas, attitude, cause and effect. 8 four-option multiple choice items [Sections 2 to 4 identical to C2 level Mastery]
Timing	30 minutes.
Item density	Part 1: 1 item per 5-8 sec conversation Part 2: 1 item per 25 secs Part 3: 1 item per 24 secs Part 4: 1 item per 24 secs
Delivery and marking	Paper-delivered in test centre. Answers on question sheet.
Weighting	All items equally weighted. Each correct answer receives one mark.
Presentation	Two hearings of the recording permitted Items available before the recording is played. Pre-reading and completion time allowed.
Cognitive processing: Goal setting	Local factual information Constructing meaning in context (relevance, speaker goals, speaker attitude, inference) Constructing a discourse model (topic links)
Cognitive processing: Levels of listening targeted by items *(ital = not targeted)*	*Word recognition* *Lexical access* *Syntactic parsing* **Establishing the bare meaning of a utterance** **Relating new information to the context** **Inference; indirect meaning** *Making referential links* **Interpreting speaker attitude and speaker goals** **Linking information**

Features of the Input				
Input as text				
	Section 1	**Section 2**	**Section 3**	**Section 4**
Domain	Conversational cues for a response	Short everyday conversations	Factual presentation	Discussion, expression of views /attitudes / information

Interaction	Dialogue	Dialogue 7-11 turns 3 speakers	Monologue	Dialogue (2 speakers)
Discourse mode	Conversational context	Interactional * Narrative * Descriptive * Instructional	Expository	Transactional * Narrative * Descriptive * Expository * Discursive
Information	General	General	Subject-specific	General
Topic	General	General	Specific	General
Cultural specificity	Mainly non-specific; but some GB specific idioms in questions	Non-specific	Non-specific	General
Abstractness	Mostly concrete; sometimes with implicit speaker intentions.	Concrete	Concrete Analytical Abstract	
Text length and duration	3 turns 14-30 words per clip; 4-9 secs. running time	7-11 turns (*3 speakers*) 145-160 words 50 secs running time	550-600 words; 3-4 mins	38 turns (2 speakers) 650 words 3-4 mins
Utterance length	Not specified in the criteria.			
Information density and complexity	High			
Input as recorded material				
Origin	Scripted and studio recorded.			
Transmission	Audio. No visual support. No contextualisation or topic setting provided in rubric.			
Articulation	(sample) Generally quite natural in intonation, pausing etc. Section 2 contained slow and artificial delivery			
Speech rate	Section 1, 4: Medium conversational rate Section 2: slow Section3: standard presentational with paced delivery			
Accent and voice	Standard British, with some attempts at mild accentuation. Section 1: Voices kept to a minimum across clips *Section 2: 3 voices*. Section 3: Monologue Section 4: 2 voices			
Language content				
Lexical level	Section 1 makes excessive use of low frequency idioms which candidates would be unlikely to know . Strangely, they especially occur in the written questions - loading difficulty on to reading Section 2: mainstream lexis at a frequency level suitable for C level candidates. Some excessive subject-specific terminology in Sections 3 and 4 and idioms in Section 4 – sufficient to render the recordings difficult to follow .			
Grammatical level	Some syntactic complexity in Section 3; subordination within and across utterances			
Reliability				
KR20	0.86/0.87			
SEM	37.15/36.24 [June 2013/Nov 2013]			

Appendix 9 - Survey of English Language Proficiency (ELP) tests used for professional registration purposes

PROFESSIONAL HEALTHCARE BODIES THAT USE IELTS AND SIMILAR ELP TESTS FOR REGISTRATION PURPOSES (as at 31 October 2014)			
Countries	**Registration bodies**	**Website address**	**Policy and practice concerning ELP requirements (May-Oct 2014)**
UNITED KINGDOM			
	Faculty of Public Health	http://www.fph.org.uk/	no information found
	General Dental Council - Dentists	http://www.gdc-uk.org/	IELTS Academic - overall band score 7 with min 6.5 in all 4 skills - one sitting - TRF not more than 2 years old
	General Dental Council - Dental professionals (nurses, etc)	http://www.gdc-uk.org/	IELTS Academic - overall band score 7 with min 6.5 in all 4 skills - one sitting - TRF not more than 2 years old
	General Chiropractic Council	http://www.gcc-uk.org/	no mention of separate ELP requirement - GCC Test of Competence includes 4 assessment components in which 'ability to communicate clearly, concisely and effectively' is evaluated (e.g. within a Clinical OSCE)
	General Optical Council	http://www.optical.org/	For registration as optometrist in the UK: obtain min score of IELTS 7 with no score lower than 6 in skill components, except for Speaking which must be min 7.
	General Osteopathic Council	http://www.osteopathy.org.uk/	IELTS Academic (preferred) - overall band score 7 or above, with min 7.0 in all 4 skills; other tests accepted: Cambridge English: Advanced (CAE); IGCSE 0500 (1st language) or 0510 (2nd language) Grade C; Hong Kong CEE Syllabus B, Grad A; TOEFL iBT -250; TOEIC - 810
	General Pharmaceutical Council	http://www.pharmacyregulation.org/	(if non-exempt) a min score of 7 in each category of the academic level IELTS English language test at the same sitting of the test
	Pharmaceutical Society of Northern Ireland	http://www.psni.org.uk/	no information found
	College of Optometrists	http://www.college-optometrists.org/	no information found
	Health and Care Professions Council - Speech and Lang Therapists	http://www.hcpc-uk.org/	be able to communicate in English to the standard equivalent to level 8 of IELTS with no element below 7.5; this requirement is stricter for speech and language therapists than for all other health professions

CONFIDENTIAL

	Health and Care Professions Council - All other professions	http://www.hcpc-uk.org/	be able to communicate in English to the standard equivalent to level 7 of the International English Language Testing System, with no element below 6.5
	Nursing and Midwifery Council	http://www.nmc-uk.org/	all non-EU-trained applicants must complete the academic version of IELTS and achieve at least a 7 in listening, reading, writing and speaking. (NB previously 6.5 - raised to 7 in Feb 2007 following public consultations for nursing in 2003 and for midwifery in 2005)
	Royal Pharmaceutical Society of GB	http://www.rpharms.com/	no information found
	Royal College of Radiologists	https://www.rcr.ac.uk/	no information found
	College of Emergency Medicine	http://www.collemergencymed.ac.uk/	no information found
	Royal College of General Practitioners	http://www.rcgp.org.uk/	no information found
	Faculty of Intensive Care Medicine	http://www.ficm.ac.uk/	no information found
	Joint Royal Colleges of Physicians Training Board	http://www.jrcptb.org.uk/	no information found
	Joint Committee of Surgical Training	http://www.jcst.org/	no information found
	Royal College of Anaesthetists	http://www.rcoa.ac.uk/	All FRCA examinations are conducted in English. Where a candidate's first language is not English, the College recommends that an applicant's communication and written skills in English are at least equivalent to the level at which a doctor could successfully pass the PLAB tests or at a level where they would expect to gain an average score of approximately 7.5/9 across all bands of the IELTS. We do not require you to provide us with a certificate proving competency in English.
	Royal College of Obstetricians and Gynaecologists	http://www.rcog.org.uk/	To apply for the Medical Training Initiative (MTI) scheme you must have an IELTS certificate: a score of AT LEAST 7.0 in each of the four domains of the test (reading, writing, speaking and listening) AS WELL AS a score of NO LESS THAN 7.5 overall - achieved within a single sitting and the IELTS certificate MUST be in date (it has a 2-year validity) on your first day of employment in the UK.
	Faculty of	http://www.fom.ac.uk/	no information found

CONFIDENTIAL

	Occupational Medicine		
	Royal College of Ophthalmologists	http://www.rcophth.ac.uk/	RCOphth runs a dual sponsorship scheme to allow international medical graduates to undertake targeted training in ophthalmology; but this requires registration with the GMC, meeting GMC IELTS requirements (Min overall 7.5 and at least 7.0 in all four skills (from June 2014). It appears that overseas doctors need to obtain a 'Licence to Practise' from the GMC.
	Royal College of Paediatrics and Child Health	http://www.rcpch.ac.uk/	MRCPCH exams: candidates are expected to hold GMC minimum standards of English Language to pass the exam
	Royal College of Pathologists	http://www.rcpath.org/	International medical graduates need to register with the GMC, implying that they will need to meet GMC English Language requirements.
	Royal College of Psychiatrists	http://www.rcpsych.ac.uk	Handbook for overseas doctors 2013 - GMC English Language requirements have to be met
	The Foundation Programme	http://www.foundationprogramme.nhs.uk	Applicants must demonstrate English language skills as set out in the GMC's "Good Medical Practice" guide (taken from UK Foundation Programme Person Specification). The Reference forms from referees in support of an applicant indicate an evaluation of whether he/she has demonstrable skills in written and spoken English that are adequate to enable effective communication about medical topics with patients and colleagues.
	Royal College of Surgeons	http://rcseng.ac.uk	For general dental practitioners seeking admission to health authority lists in the NHS, a min of 6.0 in each skill component of IELTS Academic, or equivalent on various other tests, including: AQA CELS score 3; pass at Cambridge BEC Vantage; TOEFL 580. (NB the table of English language testing qualifications is not at all clear)
AUSTRALIA			
	Australian Medical Council Inc	http://www.amc.org.au	(website information not reviewed since July 2010) The following tests are accepted: IELTS (minimum score of 7 in each of the four components; overall pass in the OET (Occupational English Test) with grades A or B only in each of the four components; successful completion of the NZREX (New Zealand Registration Examination); successful completion of the PLAB test (Professional and Linguistic Assessments Board of the GMC in the UK). These must be obtained at one sitting and within two years prior to applying for registration.
	Australian Nursing and Midwifery Council	http://www.nursingmidwiferyboard.gov.au	(new model to assess qualifications of international nurses and midwives introduced 10.02.14) Required standards: IELTS (academic) minimum score of 7 in each of the four components. OET minimum score of B grade in each of the four components. Both exam scores must be obtained at one sitting, and within the last two years except under certain special circumstances.

CONFIDENTIAL

	Australian Pharmacy Examining Council Inc	http://www.pharmacycouncil.org.au	(as at 01.01.2012) EL standard required: IELTS (academic) overall score of 7.5, with minimum score of 7 in each of the four components in one sitting; OET - overall pass with grades A or B in each of the four components in one sitting. Results must have been obtained within the last two years.
	Australian Health Practitioner Regulation Agency - national agency responsible for the registration and accreditation of 9 of the 12 healthcare professions in Australia	http://www.ahpra.gov.au	(as at 01.07.2010) The following tests are accepted: IELTS (minimum score of 7 in each of the four components; overall pass in the OET (Occupational English Test) with grades A or B only in each of the four components. Results must have been obtained within the last two years.
	Chinese Medicine Board of Australia	http://www.chinesemedicineboard.gov.au	(from 01.07.2012)The following tests were accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 6.5; OET overall pass with grades A or B only in each of 4 components; TOEFL with TSE and TWE - min score = 237, with 4.5 in TWE. Transitional arrangements in place to 30 June 2015 for those not meeting the above: IELTS (Ac) with min 6.0 in each skill component or OET with min B in each skill component. Results must have been obtained within the last two years at one sitting.
	Chiropractic Board of Australia	http://www.chiropraticboard.gov.au	(from 01.07.2010) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.
	Dental Board of Australia	http://www.dentalboard.gov.au	(from 01.07.2010) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.
	Medical Board of Australia	http://www.medicalboard.gov.au	(from 01.07.2010) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components; successful completion of NZREX; successful completion of PLAB Results must have been obtained within the last two years at one sitting.
	Medical Radiation Practice Board of Australia	http://medicalradiationpracticeboard.gov.au	(from 01.07.2012) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.

CONFIDENTIAL

	Nursing and Midwifery Board of Australia	http://www.nursingmidwiferyboard.gov.au	(from 19.09.2011) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components; other EL tests that may be approved by the board from time to time. Results must have been obtained within the last two years at one sitting.
	Occupational Therapy Board of Australia	http://www.occupationaltherapyboard.gov.au	(from 01.07.2012) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.
	Optometry Board of Australia	http://www.optometryboard.gov.au	(from 01.07.2010) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.
	Osteopathy Board of Australia	http://www.osteopathyboard.gov.au	(from 01.07.2010) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.
	Pharmacy Board of Australia	http://www.pharmacyboard.gov.au	(from 01.07.2010) The following tests are accepted: IELTS (Academic) - overall score min 7.5 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.
	Physiotherapy Board of Australia	http://www.physiotherapyboard.gov.au	(from 01.07.2010) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.
	Podiatry Board of Australia	http://www.podiatryboard.gov.au	(from 01.07.2010) The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.
	Australasian and New Zealand Podiatry Accreditation Council	http://anzpac.org.au/	For Australia: minimum score of 7.0 overall in IELTS academic with min 7.0 in each of 4 skill modules, or OET overall pass with grades A or B only in each of 4 components. For NZ: minimum score of 7.5 overall in IELTS academic with min 7.0 in each of 4 skill modules.
	Australian Dental Council	http://adc.org.au/	The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B only in each of 4 components. Results must have been obtained within the last two years at one sitting.

CONFIDENTIAL

	Australian Institute of Radiography	http://air.asn.au	The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with min grade B in all 4 components. Results must have been obtained within the last two years at one sitting.
	Australian Physiotherapy Council	http://physiocouncil.com.au	The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B in all 4 components. Results must have been obtained within the last two years at one sitting.
	Occupational Therapy Council of Australia and New Zealand	http://www.otcouncil.com.au	The following tests are accepted: IELTS (Academic) - overall score min 7.0 with no individual component score below 7.0; OET overall pass with grades A or B in all 4 components. Results must have been obtained within the last two years at one sitting.
	Speech Pathology Australia	http://speechpathologyaustralia.org.au	A comprehensive knowledge of English is essential to the practice of speech pathology in Australia. Applicants must provide evidence through achieving at least 3 A grades and 1 B grade in OET, and must agree that the examination tapes and reports can be reviewed by the organisation.
NEW ZEALAND			
	Nursing Council of New Zealand	http://www.nursingcouncil.org.nz	(as at 11.01.2013) The following tests are accepted: IELTS (minimum score of 7 in each of the four components; overall pass in the OET (Occupational English Test) with grades A or B only in each of the four components. No changes are anticipated until the completion of research being conducted by Dr Leslie Hawthorn (commissioned by AHPRA).
	Pharmacy Council of New Zealand	http://www.pharmacycouncil.org.nz	(as for Australia) EL standard required: IELTS (academic) overall score of 7.5, with minimum score of 7 in each of the four components in one sitting; OET - overall pass with grades A or B in each of the four components in one sitting. Results must have been obtained within the last two years.
	New Zealand Medical Council	http://www.mcnz.org.nz	(Changed February 2014) IELTS (academic) minimum scores 7.5 in Speaking and Listening, 7.0 in Writing and Reading, achieved within one result. OET (medical module) with A or B in each of the four components within one result.
	Dental Council of New Zealand	http://dcnz.org.nz	IELTS Academic: 7.5 or more with at least 7 in each skill component (for dentists, dental specialists, therapists, hygienists, clinical technicians); average score of 7.5 for dental technicians; at least band 7 in all 4 components for orthodontic auxiliaries. OET: a pass with A or B grades in each component for all categories of specialism.
	Physiotherapy Board of New Zealand	http://physioboard.org.nz/	IELTS Academic: 7.5 or more with at least 7 in each skill component. OET: a pass with A or B grades in each component for all categories of specialism.
USA		-	

CONFIDENTIAL

	Commission on Graduates of Foreign Nursing Schools	http://www.cgfns.org	no obvious mention of English Language requirements - this seems to be an organisation set up to verify credentials for applicants, according to the requirements of individual US state boards.
	International Commission on Healthcare Professionals	http://www.ichp.org	(takeielts.britishcouncil.org site: ichp website does not appear to be accessible) Philadelphia USA requires IELTS 6.5 (academic) with no speaking score below 7.0 for Clinical Laboratory Scientist, Physician Assistant and Speech Language Pathologist and Audiologist; and IELTS 6.0 (no speaking below 7.0) for Clinical Laboratory Technician.
	Educational Commission for Foreign Medical Graduates	http://ecfmg.org	USMLE used for medical licensure contains element testing English language and communication skills
	Federation of State Medical Boards	http://www.fsmb.org	no information found
	American Medical Association	http://ama-assn.org/	no information found
CANADA			
	College of Physicians and Surgeons of British Columbia	http://www.cpsbc.ca	TOEFL iBT: total score of 96 and with 24 in each of the 4 sections IELTS Academic : min score of 7 in each component with overall band score of no less than 7
	College of Physicians and Surgeons of Alberta	http://www.cpsa.ab.ca	IELTS Academic: min score of 7 in each component; TOEFL iBT: min score of 24 in each of the 4 sections (NB TOEFL to be phased out by 2016)
	College of Physicians and Surgeons of Manitoba	http://www.cpsm.mb.ca	no information found
	College of Physicians and Surgeons of New Brunswick	http://www.cpsnb.org	no information found
	College of Physicians and Surgeons of Newfoundland and Labrador	http://www.cpsnl.ca	IELTS Academic: min score of 7 in each component; TOEFL iBT: min total score of 92, with min score of 24 in Speaking and 20 in each of listening, reading and writing
	College of Physicians and Surgeons of Nova Scotia	http://www.cpsns.ns.ca	IELTS Academic: min score of 7 in each component; TOEFL iBT: min score of 24 in each skill component

CONFIDENTIAL

	Professional Licensing, Dept of Health and Social Services, Govt of the NW Territories	http://www.hss.gov.nt.ca	If medical training was not in English, proof of fluency in speaking and understanding the English language will be required (submit results of TOEFL)
	Professional Licensing, Dept of Health and Social Services, Govt of Nunavut	http://www.nunavut-physicians.gov.nu.ca/licensing.shtml	no information found
	College of Physicians and Surgeons of Ontario	http://www.cpso.on.ca/	no information found
	College of Physicians and Surgeons of Prince Edward Island	http://www.cpspei.ca/	IELTS Academic: min score of 7 in each component; TOEFL iBT: min score of 24 in each skill component
	College of Physicians and Surgeons of Saskatchewan	http://www.quadrant.net/cpss/	IELTS Academic: min score of 7 in each component; TOEFL iBT: min score of 24 in each skill component (TOEFL iBT likely to be discontinued)
	Yukon Medical Council	http://www.yukonmedicalcouncil.ca/	reasonably able to speak, read and write in either French or English'
	College of Registered Nurses of British Columbia	http://www.crnbc.ca	(applicants need to apply through the National Nursing Assessment Service - NNAS) Proof of Language Proficiency documents: If your first language is not English, the following tests and scores are required: IELTS (academic) overall minimum score 7.0 with minimum 7.5 (listening), 7.0 (speaking, writing) 6.5 (reading) OR CELBAN (Canadian English Language Benchmark Assessment for Nurses) with minimum scores of 10 (listening) 8 (speaking and reading) 7 (writing) . Scores must be achieved at one sitting and within 6 months of the date when your file is sent to the Regulatory Body of application.
	College of Nurses of Ontario	http://www.cno.org	Changed June 2014 - NNAS application is mandatory first step in registration, therefore presumably same criteria as above

CONFIDENTIAL

	National Association of Pharmacy Regulatory Authorities	http://www.napra.ca/	(June 2014) IELTS (academic) minimum overall band 7.0, with minimum 6.0 in each of the four components; CanTEST 4.5 minimum in each of the four skills; Michigan English Assessment Battery (MELAB) - minimum total score 85 (written composition 82; oral component 3+; no minimum for other elements); TOEFL - internet-based test: minimum total score 97 (with minimum 27 speaking and 25 writing: no minimum reading or listening); TOEFL - paper-based test: minimum total score 580 (Written English 5; Spoken English 50; TOEFL computer-based test (CBT) not accepted.
	College of Pharmacists of British Columbia	http://www.bcpharmacists.org	IELTS (academic) overall 6.5 (with minimum 6 for listening & reading; minimum 5.5 for writing and speaking); CanTEST 4.5 minimum in each of four skills ; TOEFL - internet-based test: minimum total score 92 (with minimum 25 speaking and 22 writing: no minimum reading or listening); TOEFL - paper-based test: minimum total score 566 (Written English 4.7; Spoken English 50; Michigan English Assessment Battery (MELAB) - minimum total score 82 (written composition 79; oral component 3+; no minimum for other elements)
	Ontario College of Pharmacists	http://www.ocpinfo.com	IELTS (academic) overall band 7.0 (minimum 6 in each of the four components); TOEFL - internet-based test: minimum total score 97 (with minimum 27 speaking and 25 writing: no minimum reading or listening); TOEFL - paper-based test: minimum total score 580 (Written English 5; Spoken English 50); Michigan English Assessment Battery (MELAB) - minimum total score 85 (written composition 82; oral component 3+; no minimum for other elements); Canadian Test of English for Scholars and Trainees (CANtest) - 4.5 in each of the four skills.
CARIBBEAN		-	
	Caribbean Association of Medical Councils	http://www.camcouncils.org/node/30	no information found on ELP requirements - though the CAMC examination appears to be in English and claims to test 'written and factual knowledge' as well as 'clinical communication and problem-solving'
IRELAND			
	Irish Nursing Board / An Bord Altranais	http://www.nursingboard.ie	IELTS (academic) minimum overall score 7.0, with minimum 7.0 in speaking and writing, and minimum 6.5 in listening and reading, taken in the last two years.
	Irish Medical Council	http://www.medicalcouncil.ie	evidence of 'effective communication skills' can be demonstrated by the following: IELTS (academic) minimum overall score 7.0, with a minimum 6.5 in all four skills, taken in the last two years; OR Cambridge ESOL Certificate of Advanced English (CAE) - CEFR Level C1 or C2; OR proof that medical training and internship were completed through English in a country where English is the language spoken by the vast majority of the population (e.g. Australia, New Zealand, USA).

CONFIDENTIAL

	Prehospital Emergency Care Council	http://www.phecit.ie	(nb. a letter in 2008 from the Director of the Council recognises the need for English language competence and communication skills but indicates the difficulty of systematic testing in this area, given EU Member State rules in force at the time.) No evidence of any current ELP requirements.
SOUTH AFRICA		-	
	Health Professionals Council of South Africa	http://www.hpcsa.co.za/	overall Band score 6 in IELTS
HONG KONG			
	The Medical Council of Hong Kong	http://www.mchk.org.hk/	Part II: The Proficiency Test in Medical English It is a written English paper designed to ensure that the candidates' command of English is adequate for professional purpose. The test is tentatively scheduled for 3 March 2014. *The Medical Council decided that, with effect from 1 January 2011, all candidates are required to take Part II: Proficiency Test in Medical English.
SINGAPORE			
	Singapore Medical Council	http://www.smc.gov.sg/	a. IELTS - at least 7 for ALL components (Listening, Reading, Writing AND Speaking) b. TOEFL - at least 100 marks for internet-based test c. OET - at least Grade B
	Singapore Pharmacy Council	http://www.spc.gov.sg/	IELTS: 7 overall and in all skills TOEFL iBT: total: 94 OET: B in all skills or above
	Allied Health Professions Council - Occupational Therapy and Physiotherapy	http://www.healthprofessionals.gov.sg	IELTS: at least 7 in all skills TOEFL iBT: total 100 OET: B in all skills or higher
	Allied Health Professions Council - Speech Therapy	http://www.healthprofessionals.gov.sg	IELTS: 8 overall and at least 7.5 in all skills TOEFL iBT: total: 100 OET: at least 3 Grade As and 1 Grade B
IMG : significant source countries			

CONFIDENTIAL

Bangladesh	Bangladesh Medical and Dental Council	http://www.bmdc.org.bd	no information found
Egypt	Egyptian Medical Syndicate	http://www.ems.org.eg	no information found (website in Arabic)
Egypt	Ministry of Health and Population	http://www.mohp.gov.eg	no information found (website in Arabic)
India	Medical Council of India	http://www.mciindia.org	no information found
India	Dental Council of India	http://www.dciindia.org.in	no information found
Iraq	*[no professional body identified]*		no information found
Nigeria	Medical and Dental Council of Nigeria	http://www.mdcnigeria.org/	no information found
Nigeria	Nigerian Medical Association	http://www.nationalnma.org	no information found
Sudan	Sudan Medical Council	http://www.smcreg.gov.sd	no information found
Sri Lanka	Sri Lanka Medical Council	http://www.srilankamedicalcouncil.org	no information found
Pakistan	Pakistan Medical and Dental Council	http://www.pmdc.org.pk	IELTS score 5.5 or TOEFL score 500 required for admission of overseas students to MBBS/BDS courses. No obvious statements about English language requirements for practising doctors.
EEA : significant source countries	NB For most of the countries listed below, it proved extremely difficult to reliably identify relevant professional organisations for these countries so as to determine policy and practice regarding certification of language proficiency at the national level. The task was further complicated by the websites being in the language of the country concerned, thus requiring translation capabilities and resources which had not been factored into this study. A 2009 study comparing 10 regulatory authorities worldwide, commissioned by the GMC from RAND (www.gmc-uk.org/about/research/25057.asp), provides some useful information and where possible this has been included in this table and on the report. However, the 2009 report notes the 'decentralised polycentric systems' regarding professional registration for many countries, including Spain, Germany and Italy (similar to those in India and Nigeria). In some countries doctors are simply required to sign a declaration stating that they possess a level of language that would allow them to practise medicine in that country, or that language requirement is tested during the recruitment process.		
Bulgaria			no information found
Czech Republic			no information found
France			no information found

CONFIDENTIAL

Germany			Rand 2009 reported: applicant needs to have German language skills but no test of German language skills for applicants. In 2013, however, a new dual level, 4 skills test of medical German was introduced targeting international doctors in a German-speaking hospital: telc Deutsch B2/C1 - Medizin. A second test is designed to target nursing staff: telc Deutsch B1/B2 - Pflege. Go to: www.telc.net/medizin-pflege.
Greece			Rand 2009 reported: applicants require proof of knowledge of Greek language and medical terminology
Hungary			no information found
Ireland			see above RAND 2009 reference to Ireland
Italy			see above RAND 2009 reference to Italy
Lithuania			no information found
Malta	The Medical Council of Malta	https://ehealth.gov.mt/HealthPortal/others/regulatory_councils/medical_council/medicalcouncilregisters.aspx	no obvious mention of ELP requirements (although statement made in 2003 that the "Medical Council Malta may require a Medical or Dental Professional whose primary degree is not from Malta or from an EU-Member State 'to sit for and pass a professional and linguistic proficiency test' in order to be granted registration in Malta" - but no indication of the nature of such tests.
Netherlands			no information found
Pakistan			see entry for Pakistan above
Poland			no information found
Portugal			no information found
Romania		-	no information found
Slovakia			no information found
Spain			see reference to Spain above

CONFIDENTIAL

Appendix 10 – List of Approved English Language Test Providers Criteria for Inclusion (2010[7])



LIST OF APPROVED ENGLISH LANGUAGE TEST PROVIDERS
CRITERIA FOR INCLUSION

December 2010



[7]This criteria of inclusion was published in December 2010. This document and the list of approved tests have now been superseded.

When you apply for inclusion on the UK Border Agency list of approved English language test providers, you must:

- provide evidence that you have an established trading presence in the UK
- provide details of the organisation that administers the test and awards the resulting qualification, including:
  - the organisation's name;
  - the nature of the organisation;
  - details of registration/accreditation with other bodies or government departments or agencies, enabling the organisation to award qualifications (for example, Awarding Organisation status with Ofqual);
  - evidence that the organisation is currently on the register of sponsors under the points-based system – or, if it is not, a minimum of 4 documents to establish the organisation as bona fide (for example, annual reports).
- provide contact details for the test provider, including:
  - name of contact in the organisation;
  - address;
  - telephone number;
  - fax number;
  - email;
  - website address.
- declare any other business interests; including any business functions that may present potential conflicts of interest with your inclusion on the list.

You must also meet the criteria set out in each of the four tables overleaf.

We will only accept applications from test providers as single company entities. We will not accept applications from two or more test providers acting as a single entity.

We reserve the right to refuse an application where we believe that an unacceptable conflict of interest exists.

1. TEST CRITERIA

Requirements	Further information
The UK Border Agency currently tests the English language ability of applicants in the following immigration routes: • employment routes (e.g. Tiers 1 and 2) • study routes (e.g. Tier 4); and • spouses and partners of those already in the UK For each test submitted, the test provider should indicate which immigration route(s) it wishes that test to be considered for.	For example, you may wish a test to be used for all immigration routes, or alternatively may wish that test to be used only for study routes.
The UK Border Agency needs to ensure that this exercise is robust against any future changes in government policy. For each test submitted for consideration, the test provider must give details of test validity and the test score equivalent to the following CEFR levels: • A1 • A2 • B1 • B2 • C1 • C2 We note that some tests will only test at one or a number of these levels. Where this is the case please provide scores for the CEFR level(s) tested.	For employment and study routes, the test(s) must score the candidate on all four facets (reading, writing, listening and spoken English) and the applicant must achieve at least the required CEFR level in every area. For example, a candidate requiring a C1 test must achieve C1 level in every facet.
The UK Border Agency receives applications from migrants worldwide. The test provider must provide evidence that they have established test centres in at least 40 countries for the PBS routes. For spouses and partners, the test provider must have international coverage, with concentration especially where the majority of applications are made: Pakistan, Bangladesh and India.* * If, following this exercise, the required coverage is not in place for the spouses and partner route, the current list of providers at this level will be retained and those successful as part of this exercise will be added to it. A further exercise will then be undertaken for spouses and partners.	Evidence must include the country in which each test centre is located, the full address and contact details for each centre, the frequency of testing at each centre and the capacity each centre has per test.
The UK Border Agency needs to ensure that test providers have a track record of delivering English language tests and that their testing processes are well established. The test provider must provide evidence that at least 2,400 candidates have undertaken the test provider's tests in the last 12 months.	
The test provider must provide evidence that they have arrangements in place for candidates with special needs.	You must provide evidence detailing the arrangements you have in place.

2. TEST DEVELOPMENT

Requirements	Further information
The test provider must undertake original and ongoing in-depth research into the design and structure of their test(s).	You must provide: • proof of ongoing research into testing and question design; • details of how your organisation ensures that its tests produce consistent, fair and unbiased results regardless of the nationality of the applicant; • evidence of constant monitoring of both long- and short-term performance, including trends in results and testing, supported by investigations into any anomalies and how this informs future test structure and design; and • proof that research is undertaken by qualified and experienced research staff specialising in language and linguistics teaching/education.
The test provider must have tests and questions drafted, designed and maintained by suitable, trained question writers, specialising in English language testing.	You must provide evidence detailing the experience/qualifications the trained question writers hold.
The test provider must have clearly defined guidelines that they require question writers to follow.	You must provide evidence of your organisation's guidelines.
The test provider must have a satisfactory quality assurance process that potential test questions pass through before publication.	You must provide evidence of your organisation's processes.
The test provider must hold a large bank of questions/papers to avoid applicants taking the same questions/paper twice. They must continually refresh this bank of questions.	You must detail your organisation's process and the number of papers/questions held at any one time.
The test provider must have arrangements for independent validation and/or peer review of both testing and assessment methodologies and of mapping undertaken against the Common European Framework for reference (CEFR).	You must provide evidence of: • the credentials and contact details of the validating body / peer reviewer; and • the mapping process that was undertaken.
The test provider must audit the test development processes and file these audit records available for inspection.	You must provide evidence of your organisation's processes.

4

3. TEST DELIVERY

Requirements	Further information
The test provider must deliver tests in a secure testing environment.	You must provide evidence of: • how your organisation assesses that the premises in which tests will be delivered are secure and appropriate for testing, and the criteria used to make this assessment; • regular monitoring of test centre compliance, with processes in place for handling poor performance (we have strong preference that this monitoring be undertaken by independent, qualified auditors) – monitoring must take place once over the period of licence; • the checks and procedures that your organisation undertakes to prevent identity fraud/substitution; • test centres with any other business interests, including any business functions that may present potential conflicts of interest.
The test provider must have secure processes in place for the control and storage of test registration/test papers/test material to prevent unauthorised access.	You must provide evidence of your organisation's processes.
The test provider must have secure processes in place for the movement and handling of test papers/test material on test days to prevent unauthorised access.	You must provide evidence of your organisation's processes.
The test provider must have security arrangements in place where computer-based assessment is used, in particular to ensure the prevention of access to other computer based tools such as the internet.	You must detail your organisation's security arrangements.

The test provider must have appropriate standardised invigilation arrangements in place.	You must provide evidence of: • the criteria that candidates must satisfy in order to become test invigilators, and how invigilators are assessed; • checks that all test invigilators are fully trained and have a competent level of spoken English; • clearly defined guidelines that invigilators must apply during testing; • the minimum ratio of invigilators to students that your organisation accepts for testing; • measures in place to prevent conflicts of interest between invigilators and test candidates; and • the process for monitoring the performance of invigilators.
The test provider must have appropriate standardised test marking arrangements in place.	You must provide evidence of: • the criteria that candidates must satisfy in order to become test markers, and how test markers/examiners are assessed; • checks that all test markers/examiners are fully trained and have a competent level of English; • a clearly defined marking scheme for assessing level of attainment; • measures in place to prevent conflicts of interest between test markers/examiners and test candidates; and • the process for monitoring the performance of test markers/examiners.
The test provider must audit their test delivery processes and file the records of these audits available for inspection.	

4. SECURITY AND VERIFICATION OF TEST RESULTS

Requirements	Further information
The test provider must have security features on the test award to counteract any attempts to produce fraudulent evidence.	You must provide evidence of how test results are made available to the applicant. You must provide an original copy of the award where this is issued to a candidate.
The test provider must make test results available for verification by the UK Border Agency either via a website verification service or a telephone or e-mail facility where verification can be obtained within a 24 hour period.	We have a strong preference that verification be available online and contain a photograph of the applicant. You must demonstrate how your organisation will work with the UK Border Agency to overcome technical and other procedural issues in relation to verification.

6