Cambridge English

Principles of Good Practice

Research and innovation in language learning and assessment

May 2016

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

2015

Our mission

To be the experts in language assessment, delivering excellence and innovation

Our vision

To make our flexible range of assessment tools an indispensable component of language learning, teaching and testing throughout the world

Contents

Introduction

The *Principles of Good Practice* outlines the Cambridge English approach to language learning, assessment, test development and quality management. We describe some of the key concepts involved in language testing and follow this with examples of how we put these principles into practice. Our aim is to provide an accessible and concise overview of a complex area. Further key resources are signposted throughout and readers are invited to explore the issues in more depth.

Language ability is being used increasingly as one of the key criteria for life-changing decisions such as immigration, education and employment. If we look at English language testing, the stakes associated with passing a test have dramatically increased in the last two decades. With this comes a great responsibility for assessment providers to develop tests that are fair, accurate and valid.

Tests need to be fit for purpose, offering users a range of solutions that meet diverse needs. To achieve fitness for purpose, we aim to maximise the appropriate balance of the following qualities: Validity, Reliability, Impact, Practicality and Quality, or VRIPQ.

This approach underpins everything we do and is based on the most up-to-date thinking on educational assessment and on our own extensive research and validation activities.

The *Principles of Good Practice* is organised around four guiding principles which shape the Cambridge English approach to language assessment:

Fitness for Purpose

Communication and Collaboration

Quality and Accountability

Validity and Validation

All four guiding principles derive from our role within the University of Cambridge and the Cambridge Assessment Group. They reflect the University's goal *'to contribute to society through the pursuit of education, learning and research at the highest levels of excellence'*. They also embody the principles set out in Cambridge Assessment's document *The Cambridge Approach* and in the ALTE *Code of Practice*.

I hope that by reading the *Principles of Good Practice*, you will see how we have consistently developed our tests to ensure they are practical to deliver and have a positive impact on individuals and society as a whole.

Nick Saville, Phd

Director, Research and Thought Leadership
Cambridge English Language Assessment

Further reading

Cambridge English website

www.cambridgeenglish.org gives detailed information on our products and services. Regularly updated test data is available from www.cambridgeenglish.org/principles.

Studies in Language Testing (SiLT)

A series of academic books published by Cambridge English and Cambridge University Press. See www.cambridgeenglish.org/silt for a full listing and abstracts.

Hawkey, R and Milanovic, M (2013) *Cambridge English exams – the first hundred years. A history of English language assessment from the University of Cambridge 1913–2013.*

Research Notes

A quarterly publication providing up-to-date coverage of the Cambridge English research programme. See www.cambridgeenglish.org/researchnotes for a complete archive.

IELTS website

www.ielts.org includes regularly updated test data and *IELTS* research reports.

Association of Language Testers in Europe (ALTE) website

www.alte.org includes the ALTE *Code of Practice* (1994) and ALTE *Principles of Good Practice* (2001).

Cambridge Assessment website

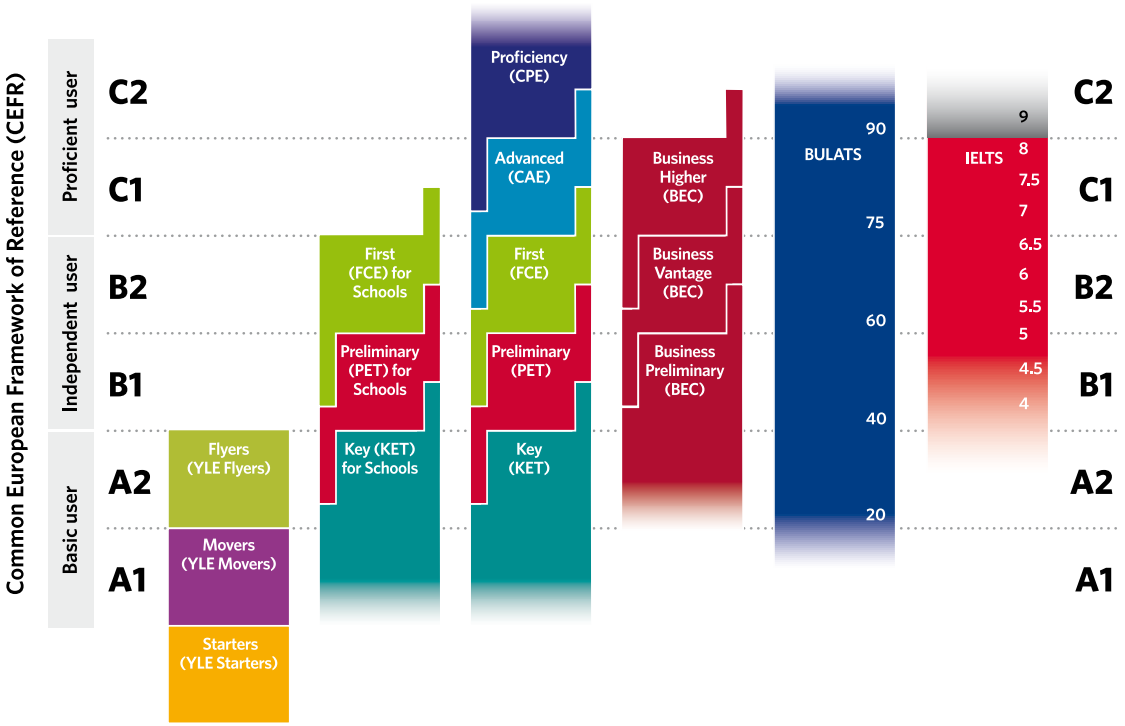
www.cambridgeassessment.org.uk includes *The Cambridge Approach: Principles for designing, administering and delivering assessment* (Cambridge 2009).

Section 1: Fitness for Purpose

Ever since our first exam was introduced in 1913, our approach has always been to develop tests that meet specific needs. Since then, we have continually extended the range to include exams at a wide variety of levels and for purposes as diverse as higher education and migration; business, legal and financial communication; and motivating and rewarding young learners. These are complemented by a range of qualifications for teachers, and by a very wide spectrum of supporting services, all designed to support effective learning and use of English. Most of our exams can be taken either as paper-based or computer-based versions.

Cambridge English

A range of exams to meet different needs



For more information on the Common European Framework of Reference for Languages (CEFR) see pages 25-26 and page 32

Tailored solutions

Cambridge English Language Assessment works with many national governments and other organisations to develop learning and testing solutions that meet their precise needs. Where these needs cannot be met using our existing services, we develop tailored solutions. For example:

In France, we have worked with the Ministry of National Education to develop and implement a bespoke B1 level test for students in the international sections of state secondary schools. Tens of thousands of students have taken the test.

We worked with the Colombian Ministry of Education and National Awarding Body in a long-term programme to raise English language standards, enhance teaching standards and improve the testing of English. The 4-year programme included benchmark testing, test design, development and deployment, and item-writing skills development to enable the localisation of test production. The Cambridge English tests are taken by over half a million students annually.

Other projects where we have provided specially developed solutions include the Beijing Speaks English project in China, SEPA Inglés in Mexico, a blended learning course for Italian universities, and a testing service for Manpower.

- Gomez Montez, I, Marino, J, Pike, N and Moss, H (2010) Colombia national bilingual project, *Research Notes* 40, 17–22.
- Harvey, A, Balch, A and Salamoura, A (2010) The adoption of international certification in the French state school sector, *Research Notes* 40, 7–9.

Learning Oriented Assessment (LOA)

Cambridge English Language Assessment approaches LOA from an assessment specialist perspective, taking a systemic view where assessment operates on multiple levels and takes many forms. It encompasses both the macro level of framing educational goals and evaluating outcomes, and the micro level of individual learning interactions which take place in the classroom or outside it – that is, both formal and informal assessment. The term LOA is chosen to emphasise that all levels of assessment can and should contribute to both the effectiveness of learning and the reliable evaluation of outcomes.

Learning Oriented Assessment provides a clear structure for integrating in-course tests, public examinations and less qualitative observations of learners. It helps plan course objectives and to ensure that lessons and study outside the classroom directly contribute to the achievement of each learner's personal objectives.

- Jones, N and Saville, N (2016) *Learning Oriented Assessment: A Systemic Approach*, Studies in Language Testing volume 45, Cambridge: UCLES/Cambridge University Press.

In summary, we develop high-quality assessments which are appropriate to their context and intended uses. This principle of fitness for purpose is essential in ensuring that our assessments are valid and that all test takers are treated fairly.

1.1 Fairness

A fair test should not discriminate against sub-groups of candidates or give an advantage to other groups. It should also be fair to those who rely on the results (such as employers and universities) by performing consistently and by accurately assessing the ability being tested. Below are some examples of the ways in which we work to ensure fairness and equality of experience for all of our stakeholders.

Equal opportunities: throughout our test development, administration and validation processes we take into account the diversity of our test takers and the need to treat them all fairly, irrespective of their background (see boxed section on next page and Section 4.4).

Security: it is essential that test materials are kept securely. If candidates are able to see questions in advance of taking a test then their answers will be based on memory rather than language ability and the validity of their results will be undermined. For this reason, we make sure that test materials are held in password-protected item banks (see page 25) and that items from 'live' tests are not published.

One way that candidates may try to breach regulations is by having someone else take the test for them. To combat this, we have introduced test-day photos for key exams alongside standard ID checks.

Special Circumstances

We provide Special Arrangements for test takers who have a permanent or temporary disability, for example we can offer some question papers in Braille or large print.

Where individuals have been affected by adverse circumstances such as illness immediately before or during an examination Special Consideration can be applied.

The *Notice to Candidates*, given to all candidates, clearly states our rules and regulations and makes it clear that breaching these rules, for example through copying from another candidate, will lead to disqualification. Breaches may be reported by centres, by examiners, or alternatively may be brought to light by our routine statistical analysis. Suspect results will be withheld until further investigation has been carried out.

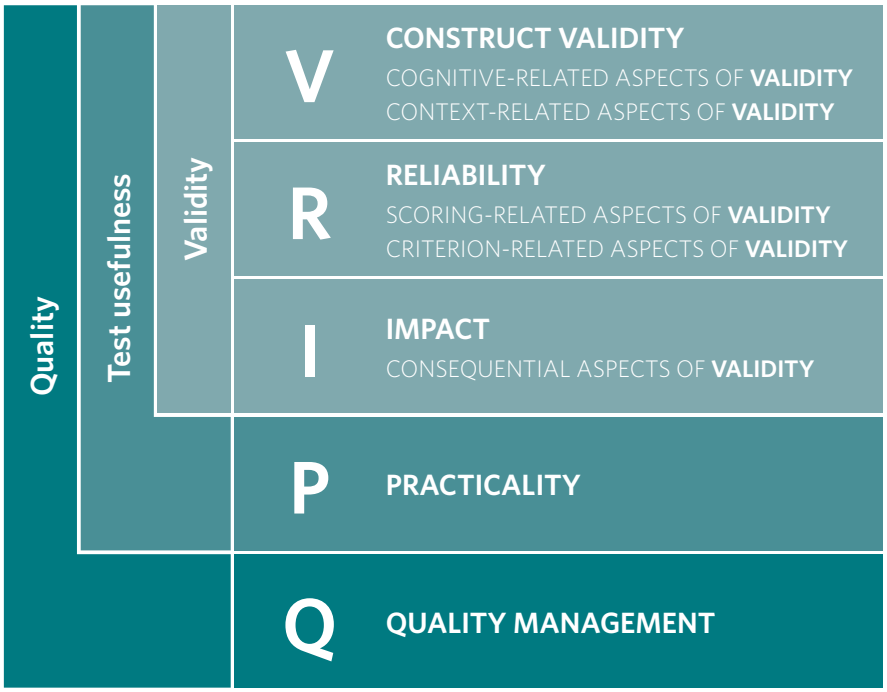
Elliott, M (2013) Test taker characteristics, in Geranpayeh, A and Taylor, L (Eds) *Examining Listening: Research and Practice in Second Language Listening* (Studies in Language Testing, volume 35), Cambridge: Cambridge University Press, 36–76.

Data protection: all personal data collected is stored in accordance with the data protection laws of England and Wales. Stakeholders are informed of the use we may make of data in the *Regulations* and *Summary Regulations for Candidates* documents.

Results: candidates first receive a Statement of Results, either online or from the centre where they took the exam, followed by their certificates. We explain the results and certificates clearly so that appropriate decisions can be made by candidates and test users. Certificates include user-oriented explanations such as Can Do statements that explain the kinds of things learners can do at each level. So, for example, at the level of our *Cambridge English: First (FCE)* exam one of these statements is: 'Can follow a talk on a familiar topic'.

Secure online verification services allow recognising organisations to confirm the authenticity of candidates' qualifications quickly and easily, increasing their confidence in the exams and simplifying their recruitment procedures. Where a candidate or teacher feels that a particular result is not in line with expectations, they may make use of our enquiries and appeals processes. We also keep historical results data securely and permanently so that candidates can contact us at any time to get a past result.

1.2 The Development of VRIPQ



The diagram above illustrates our approach to achieving fitness for purpose, and shows how developments in language testing theory and in quality management have been taken on board.

Traditionally, theoretical frameworks in language testing tended not to include practicality and quality management, focusing instead on validity and reliability, which were often seen as polarised. This view changed when Messick (1989) argued for a 'unitary approach' to validity in which the interacting nature of different types of validity is stressed. This is represented in the diagram above by the 'validity' side bar.

In 1996 Bachman and Palmer explicitly considered practicality within their 'test usefulness' concept. This coincided with the inclusion of practicality in the VRIP approach developed by Cambridge in the 1990s.

Since then there have been two key developments within our approach. The first has been the strengthening of quality management principles within VRIP to make VRIPQ. You can find out more about this in Section 3 of the *Principles of Good Practice*.

The second has been the development of the socio-cognitive approach in collaboration with Professor Cyril Weir (2005). This approach is used to set out the constructs of our tests in the *Studies in Language Testing* 'construct' volumes (see page 31) and is shown in the diagram above by the inclusion of five 'aspects' of validity. You can find out more about these aspects in Section 4 of this booklet.

In the Appendix you can see the framework that underlies the VRIPQ structure.

Further reading

Fitness for Purpose

The Cambridge English approach to ensuring fitness for purpose is explained in the following volumes in the *Studies in Language Testing* (SiLT) series:

Davies, A (Ed.) (2008) *Assessing Academic English: Testing English proficiency, 1950–1989 – the IELTS solution* (Studies in Language Testing, volume 23), Cambridge: Cambridge University Press.

Hawkey, R (2009) *Examining FCE and CAE: key issues and recurring themes in developing the First Certificate in English and Certificate in Advanced English exams* (Studies in Language Testing, volume 28), Cambridge: Cambridge University Press.

Jones, N and Saville, N (2016) *Learning Oriented Assessment: A Systemic Approach* (Studies in Language Testing, volume 45), Cambridge: Cambridge University Press.

Moeller, A J, Creswell, J W and Saville, N (Eds) (2016) *Second Language Assessment and Mixed Methods Research* (Studies in Language Testing, volume 43), Cambridge: Cambridge University Press.

O'Sullivan, B (Ed.) (2006) *Issues in Testing Business English: the revision of the Cambridge Business English Certificates* (Studies in Language Testing, volume 17), Cambridge: Cambridge University Press.

Weir, C and Milanovic, M (Eds) (2003) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002* (Studies in Language Testing, volume 15), Cambridge: Cambridge University Press.

Fairness

See the following pages on our website:

www.cambridgeenglish.org/exams-results

www.cambridgeenglish.org/special-circumstances

The following article is also relevant:

Kunnan, J (2008) Towards a model of test evaluation: using the Test Fairness and the Test Context Frameworks, in Taylor, L and Weir, C (Eds) *Multilingualism and Assessment: achieving transparency, assuring quality, sustaining diversity. Proceedings of the ALTE Berlin Conference May 2005* (Studies in Language Testing, volume 27).

Section 2: Communication and Collaboration

Cambridge English Language Assessment works with a wide range of individuals and organisations in developing, administering and validating our assessments. Our stakeholders include test takers and test users such as employers, universities, immigration departments and many other organisations, as well as our network of centres, educational publishers, materials writers and others.

2.1 Communication and Stakeholder Support

Language assessment is a complex process and stakeholders require clear, accurate information covering every stage of the process from descriptions of the exams themselves to administrative instructions and information on interpreting candidates' results.

At the same time, the organisation needs to be highly receptive to the needs, opinions and knowledge of our stakeholders. We aim to provide effective channels for two-way communication, listening to our customers and to other experts in language assessment and adjacent fields, and actively seeking the views of all our stakeholders.

This two-way communication is the responsibility of every part of the organisation. We also have a market research process to proactively collect and analyse the views of stakeholders.

Information and support services

As well as our academic and research publications (see page 4), a very wide range of information and support is provided for customers and stakeholders, including:

- information and administrative websites including dedicated websites for many of our key markets
- online candidate and teacher resources
- Handbooks for Teachers, including full descriptions of the test materials, and marking criteria and sample papers
- coursebooks, practice tests, supplementary materials and online courses.

We work with Cambridge University Press to provide support materials for the Cambridge English exams, along with a wide range of other books and materials managed by Cambridge Exams Publishing, a dedicated team of staff drawn from both organisations.

In addition to printed and online resources, it is essential that customers have access to expert advice and support, helping them to understand the examinations and overcome any concerns. The front line of these services is provided by our network of local offices around the world and by the Customer Services Helpdesk, as well as by dedicated operational support staff who are fully trained to deal with technical and administrative enquiries.

Our staff and consultants attend over 750 conferences, seminars and forums each year to take part in academic discussion, promote the exams, provide information to stakeholders, and help teachers and other professionals to develop their skills.

Developing assessment literacy

We are committed to a programme of assessment literacy, which aims to familiarise stakeholders with key concepts in language assessment to enable them to take better-informed decisions and to improve the quality of testing in different contexts. We also provide support for ALTE's growing programme of courses in assessment, which are available at three levels and take place several times a year in different European countries.

2.2 Collaborative Arrangements

We work closely with many organisations throughout the world. This helps us to deliver our products and services, and to contribute to the development of standards in language teaching, learning and assessment. Our collaborative relationships include:

- formal institutional collaborations such as *IELTS* (see boxed section below) and ALTE (see page 20)
- Alliance Française, Goethe-Institut and Universidad de Salamanca – who work with us to make *BULATS* available in French, German and Spanish
- long-term relationships with professional bodies such as Evaluation and Accreditation of Quality in Language Services (EAQUALS), English UK and the International Association of Teachers of English as a Foreign Language (IATEFL), as well as many national and regional organisations in key markets
- advisory and observer status through ALTE with international organisations including the Council of Europe and the United Nations
- long-term working relationships with leading publishers, language school chains and research bodies
- specific project groups such as English Profile, or SurveyLang, the consortium which produced the First European Survey on Language Competences in 2012.

We also have long-standing working relationships with many other organisations including universities, government departments, major commercial organisations, institutions of the European Union and many others.

Staff in all parts of the organisation develop and manage relationships with stakeholders. Our Legal Affairs team plays a key role in ensuring that appropriate arrangements are in place to enable effective collaboration, while minimising any risks that may arise.

IELTS

The worldwide success of *IELTS* is driven by the three-way *IELTS* agreement between Cambridge English Language Assessment, the British Council and IDP: IELTS Australia. This relationship goes back over 25 years and allows us to use the different expertise, networks and resources of the three organisations to deliver a world-leading assessment system. Cambridge is responsible for developing and setting the tests, while the British Council and IDP: IELTS Australia run the centre network and administer the tests, and are responsible for examiner management. The partners work together on business development, communications and research projects.

Recognising organisations

A large proportion of candidates take exams to help them to achieve specific educational or professional objectives or to enable them to migrate for study or work. Cambridge English qualifications are recognised by more than 20,000 universities, employers and government departments around the world, making them extremely valuable to candidates. We have a Recognition team in our Business and Marketing Group which works with current and potential recognising organisations to increase the recognition and use of the exams and to enhance the services we provide them.

UK immigration

Language qualifications are increasingly used as part of the regulation of international migration, and Cambridge English Language Assessment and our ALTE partners play an active role in discussion of this issue, advocating appropriate, fair and secure approaches to language testing to ensure that the use of the tests is not a source of discrimination or open to abuse.

We work closely with the UK Home Office to ensure the availability of tests which fully meet the English language requirements for those who wish to work, study or settle in the UK.

Publishers and media

Many publishers in the UK and overseas produce resources for candidates and teachers who are preparing for our exams. We collaborate with some of these publishers to help them produce materials which are accurate, relevant and useful, although we do not normally endorse or recommend specific titles produced by third parties.

Standards of language learning, recognition of qualifications and other subjects related to our work are frequent topics of discussion in national and local media around the world. We regularly provide journalists with comment and information on issues of national and local interest.

Cambridge Assessment and the University of Cambridge

Cambridge English Language Assessment is a department of the University of Cambridge and a part of the Cambridge Assessment Group – Europe's largest assessment agency.

We worked closely with other parts of Cambridge Assessment to develop the common standards known as '*The Cambridge Approach*' and to deliver joint projects, events and services in several countries. We also share services such as information systems, distribution and printing, enabling us to benefit from additional resources and economies of scale.

As part of the University of Cambridge, we have close working relationships with many departments including Cambridge University Press and the Department of Theoretical and Applied Linguistics.

Section 3: Quality and Accountability

Cambridge English Language Assessment is certified to the ISO 9001 international standard for quality management systems. We comply with regulations set by Ofqual, the UK regulator of assessments, and take part in the audits of the Association of Language Testers in Europe (ALTE). These external checks provide a sound basis for accountability and give stakeholders confidence in the quality of our assessment systems.

Good governance

Our *Quality Policy* and the *Principles of Good Practice* are reviewed every year as part of the business planning cycle. Our Business Management team manages the business planning process and key business decisions are signed off at a regular meeting chaired by the Cambridge English Chief Executive.

There are six groups within Cambridge English Language Assessment: Business and Marketing; Digital and New Product Development; Global Network; Network Services, Operations, Validation and Assessment; Partnerships, Projects and Policy; and Research and Thought Leadership.

3.1 Defining Quality

Quality is about achieving fitness for purpose and doing it consistently. Quality is therefore about all of the policies, processes and procedures that enable an organisation to do this. It involves a drive for continual improvement, and is achieved by focusing on good planning, good record keeping and on cutting error and waste wherever possible. Quality activities are often classified as quality control or quality assurance activities. Quality control activities focus on checking and testing a product or service to make sure it meets quality requirements. Quality assurance activities focus on managing, monitoring and measuring processes so we can be confident that quality requirements will be met.

Scope of our management system: Our systems and processes for designing, developing and delivering examinations and assessment services are certified as meeting the internationally recognised ISO 9001 standard for quality management. The scope of our management system: the design, development and provision of language examinations, assessment services and teaching qualifications. The management control of country offices and regional activities.

Some of the quality assurance activities which ensure consistency and control across the organisation include:

Document control: when communicating information through documents such as Standard Operating Procedures (SOPs) and Work Instructions, staff need to be confident that they are referring to the right document and that it is as up to date as possible. In order to help in this aim a Central Document Register lists those documents which are used across the organisation's groups, and a series of Local Document Registers list documents which are needed for carrying out activities in particular groups or units.

Records management: this helps us to make the best use of electronic and physical space and to follow legal requirements in relation to data protection. We have retention schedules that show how long different record types should be kept.

Risk management: part of good governance involves carrying out appropriate planning for the future. Risk analysis should therefore be carried out, both for new projects and processes or

where current processes are affected by external or internal change. Any high-level risks are managed on an Action Plan overseen by the Chief Executive and all senior staff.

Non-conformances and corrective actions: we try to make sure that errors do not happen. But if they do, we learn from them by putting corrective actions in place to stop them happening again.

Internal audits: our internal auditors are staff members who want to contribute to the success of the organisation and gain an insight into areas outside their own job role. They receive training and then carry out audits of other parts of the organisation. In essence auditors are trained to ask four key questions: 'What do we say we do?'; 'Do we do what we say?'; 'Is it effective?'; 'Can we do better?'.

The Cambridge English Quality Management team can give further advice and guidance on these quality assurance activities. Most quality control and quality assurance activities, however, are planned, designed and carried out by teams across the organisation and are not 'owned' by a separate quality department. Throughout the rest of this document there are boxed examples of these activities.

Dealing with enquiries and complaints

Our Helpdesk staff ensure that customer enquiries and complaints are dealt with clearly and in a professional, timely manner. Each enquiry is assigned an owner and tracked via a unique reference number during its lifecycle. Service Level Agreements are closely monitored to ensure that our customers receive the right level of support, and responses to standard queries are quality checked, centrally stored and reviewed on a regular basis to ensure that they remain current. Helpdesk staff receive professional training in the effective use of written correspondence and telephone manner, as well as in the procedures and systems involved in the exam administration cycle.

3.2 Delivering Quality

We have over 400 staff based in Cambridge and more than 20 offices around the world.

Quality is the responsibility of all staff

Every member of staff has personal performance objectives which are implemented as part of Cambridge Assessment's Performance Management Scheme (PMS). Individuals link their personal objectives to those of their team or work group, and to those of the organisation as a whole. Professional development and training requirements are identified and supported through this system.

All staff are required to:

- understand how their role contributes to the organisation as a whole
- ensure that key elements of the continual improvement cycle are carried out.

Where they work on processes that contribute to the validation of a particular assessment, staff have a responsibility to:

- ensure that they have understood the concepts involved, as set out in Section 4 of this document
- document the evidence needed in carrying out validation and in presenting the validity argument (see Appendix).

Staff work in collaboration with approximately 20,000 examiners (see also pages 27–28), item writers and consultants. These assessment professionals are selected and managed according to defined quality assurance processes including procedures for recruitment, training and standardisation and for performance monitoring and evaluation. Writing and Speaking Examiners in the UK are now recruited online through our Examiner Management System and Speaking Examiners globally will also be able to apply using the system shortly. This is done to ensure the quality of our assessments and the consistency of standards over time and across the world.

We also work with a network of over 2,700 examination centres, and around 40,000 registered preparation centres. The Centre Quality Assurance team has the main responsibility for ensuring the quality of the work of centres.

Quality assuring the work of our examination centres

Centres are selected according to clear, transparent procedures which include criteria covering security, expertise, financial stability and customer service. Performance of all centres is regularly monitored.

Each centre has a named responsible person called a Centre Examinations Manager (CEM). Centre staff, including invigilators, are given support and training to ensure that the exams are run in accordance with our regulations. Support includes online and face-to-face training as well as comprehensive documentation and guidance. The *Handbook for centres*, which is updated yearly, includes detailed information on running a centre and running an examination.

Wilson, J (2009) Support and training for Cambridge ESOL exam centres, *Research Notes* 38, 2–4.

Co-ordinated communications activities including the monthly *Centre News* and regular meetings with centre representatives ensure that centres are kept up to date with developments and that we receive feedback from them.

Quality control – centre inspections

We have around 140 inspectors worldwide who carry out regular inspections of our centres to ensure they are meeting our exacting quality standards. These inspections cover security of materials, examination room set-up, conduct and supervision of the exam, and arrangements for computer-based tests and Speaking tests. Inspectors produce a full report on each visit, identifying any areas for improvement. In addition, some centres complete a self-audit process, and support visits are provided to newly approved centres in their first year of running the Cambridge English exams.

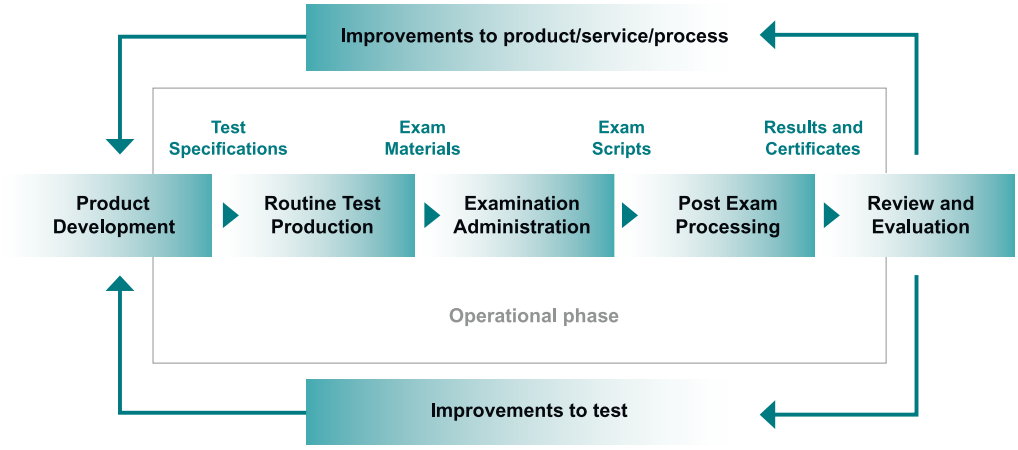
McCauley, C and Collett, D (2010) Working with examination centres to encourage continuous high standards: The centre inspections programme, *Research Notes* 39, 20–23.

3.3 A Process Approach

Defining processes is a crucial part of a quality management approach. When processes are defined and agreed, quality control and quality assurance procedures can be carried out. These procedures ensure that the processes are being followed and that they are effective and efficient.

The Cambridge English Continual Improvement Cycle shown below is fundamental to our approach to quality management and validation. It shows an iterative cycle, following the Plan, Do, Check, Act model on which the ISO 9001 standard is based.

The Cambridge English Continual Improvement Cycle



The boxes across the centre of the diagram above – from Product Development to Review and Evaluation – are our core processes. These processes (see Section 3.4) contribute directly to the creation of exam materials, and to the marking and grading of candidate results.

We are always looking for ways we can improve our service to our customers. In particular, we know that if we can make things easier for our centres, we can improve the accessibility and availability of our exams. A recent development which supports this aim has been the introduction of web-based administration for centres. We are also involved in major projects to improve the flexibility of our processing systems so that we can deliver a more innovative range of products.

Quality assurance – delivering projects

All development and revision processes, whether for tests or for other products and services, are managed using project management methodologies and tools. Projects have defined stages including proposal, initiation and implementation phases. Guidance on running projects is available from the Business Change team. This team also supports staff in other areas when they work on projects that require acceptance testing of software developments.

3.4 A Model for Test Development and Validation

The Cambridge English Continual Improvement Cycle is also a model for test development and validation (Weir and Milanovic 2003). The rationale for this approach is based on the need to establish fitness for purpose.

Plan: Product Development

The process begins when there is a perceived need for a new or revised test, and can be broken down into three phases: planning, design and development. The first task in the planning phase is to define the intended context and use of the prospective test by identifying stakeholders and their needs. These needs are then linked to the requirements of test usefulness (see Section 4) and attention is paid to both theoretical and practical issues. The key output of the product development stage is a set of test specifications. This is a document or documents defining the test, its validity argument (see Section 4) and its operational requirements. The specifications act like a 'blueprint' for the operational production of tests. Most developments include extensive trialling of materials which are analysed and reviewed before the final specifications are produced.

Do: the Operational Phase

When a test goes 'live' it moves into the operational phase; an iterative process that is repeated for each test version or session. For all tests there are three stages in this process: routine production of test versions (see the box on page 18 for further information); examination administration; and post-exam processing.

Essentially, examination administration means making sure that all necessary arrangements are in place so that candidates can take the exam of their choice. Key tasks include quality assurance of the work of centres (see page 15), delivery of exam materials and administrative documentation to centres, and the recruitment, training and allocation of examiners (see pages 15, 27–28).

The main stages of post-exam processing are marking, grading (see page 28) and the reporting of results (see page 7). Data on test takers, test materials, and marking and grading procedures is captured and analysed for all exam sessions.

Review: Review and Evaluation

All assessments and related services are reviewed regularly. Review takes place during the routine monitoring of operational processes, as well as periodically to a timescale which is defined for each product or service. In some cases review may lead to the decision to withdraw a test, and in this situation stakeholders would be consulted and informed in advance. Improvements can always be made. Where they are small in nature they may be implemented in an ongoing manner, whereas any major revisions will be carried out as a project. In essence this involves looping back to the 'plan' phase of the cycle.

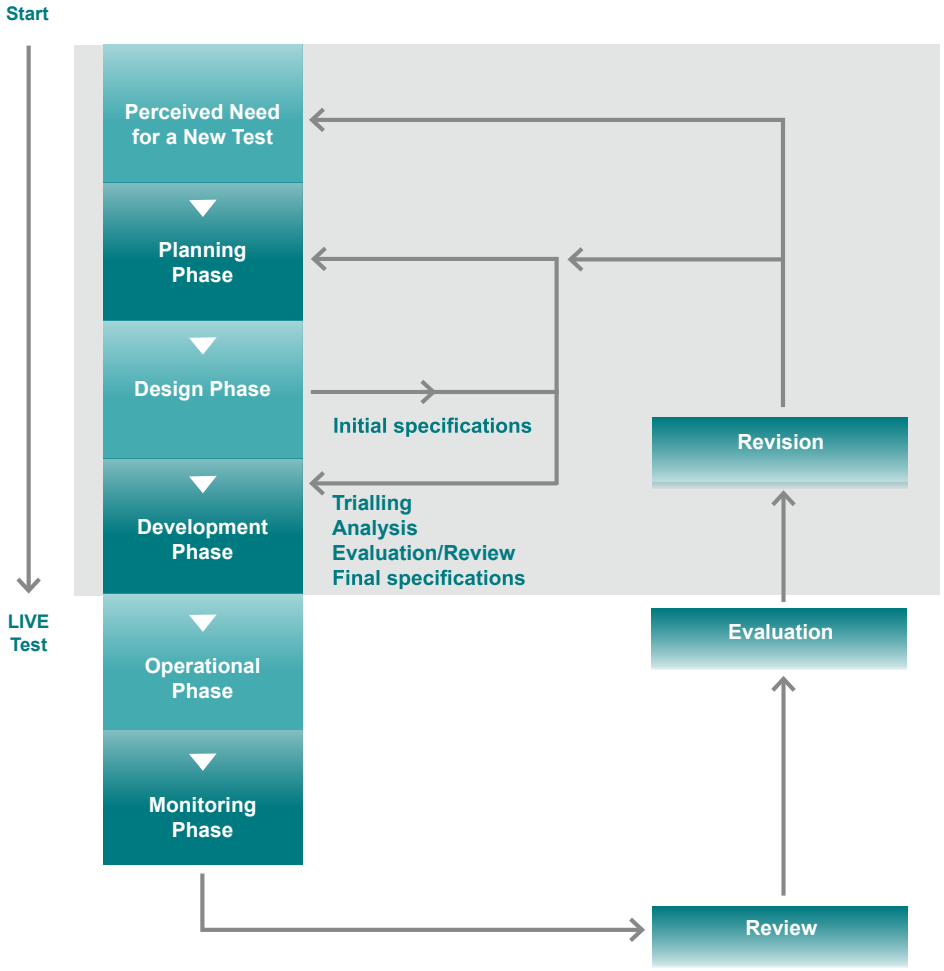
The Question Paper Production (QPP) process

This process is managed by staff in our Network Services, Operations, Validation and Assessment Group. It begins with the commissioning of draft materials and ends in the printing of the final question papers, or in the case of computer-based tests the publishing of the test version to the server. The process for each test component (e.g. the Listening paper) is managed by an Assessment Manager who works with an external expert called a Chair to manage a team of item writers. There is a yearly planning and review cycle and the process contains numerous quality checks.

Each paper has item writer guidelines that specify the requirements of each task type and list quality criteria such as text content and source. Questions that do not meet these criteria are rejected or rewritten. Those that are accepted are taken through a thorough editing process by experienced consultants. Key to the QPP process is pretesting (see page 25), where material is tested before it is used. Performance data is obtained for each task, including how difficult the sample of candidates found the questions and how well the questions discriminated between stronger and weaker candidates. These statistics, along with the expert judgement of a pretest-review panel, enable further amendments to be made. Materials which meet requirements go into a bank of items (see page 25) ready for test construction.

For more information on this process see Green, A and Jay, D (2005) Quality Assurance and Quality Control: reviewing and pretesting examination material at Cambridge ESOL, *Research Notes* 21, 5–7.

A model of the test development process



Further reading

Quality management and assessment

Association of Language Testers in Europe (2011) *Manual for Language Test Development and Examining, Strasbourg: Language Policy Division*, Council of Europe.

Saville, N (2012) Quality Management in Test Production and Administration, in Fulcher, G and Davidson, F (Eds) *The Routledge Handbook of Language Testing*, Routledge.

The theme of *Research Notes* 39 (2010) is quality assurance and its impact on language assessment and teaching, including the following article:

Rose, D (2010) Setting the standard: Quality Management for language test providers, 2–7.

An overview of quality management in assessment is given in Ramaswamy, R and Wild, C (Eds) (2007) *Improving Testing: Process Tools and Techniques to Assure Quality*, London: Routledge.

Weyant, K and Chisolm, A (2014) Safeguarding fairness principles through the test development process: A tale of two organisations, *Research Notes* 55, 3–6.

General quality management

Influential books on quality include:

Deming, W (1986) *Out of the Crisis*, Cambridge: Cambridge University Press.

Juran, J (1999) *Juran's quality handbook*, New York: McGraw Hill.

The ISO 9001 standards (London, British Standards Institution) are set out in:

British Standards Institution (2005) ISO 9000:2015
Quality Management Systems – fundamentals and vocabulary.

British Standards Institution (2008) ISO 9001:2015
Quality Management System Requirements.

British Standards Institution (2009) ISO 9004:2009
Managing for the sustained success of an organization – A quality management approach.

A large amount of useful information on quality management is also available on the website of the Chartered Quality Institute (CQI) www.thecqi.org

Section 4: Validity and Validation

In order to achieve fitness for purpose, we design and deliver assessments which appropriately balance the essential features of assessment, taking into account contemporary views of validity and validation.

Validity is generally defined as the extent to which an assessment can be shown to produce scores and/or outcomes which are an accurate reflection of the test taker's true level of ability. It is concerned with the appropriateness and meaningfulness of the inferences made when using the test results within a particular social or educational context. Validation is therefore the process of accumulating evidence to support these interpretations.

In adopting this approach, we are drawing on internationally recognised standards such as the AERA/APA/NCME *Standards for Educational and Psychological Testing* (1999). We are also complying with obligations which we have undertaken as part of the Association of Language Testers in Europe.

Association of Language Testers in Europe (ALTE)

Cambridge English Language Assessment is a founder member of this organisation which was established in 1990 and now has 33 member organisations throughout Europe. ALTE is a collaborative association that works to increase standards and coherence in language qualifications throughout Europe and provides a forum for discussion and collaboration through its regular conferences and meetings. ALTE has developed its own quality management system including procedures for auditing member organisations. The ALTE *Code of Practice* (1994) and ALTE *Principles of Good Practice* (2001) are available from www.alte.org.

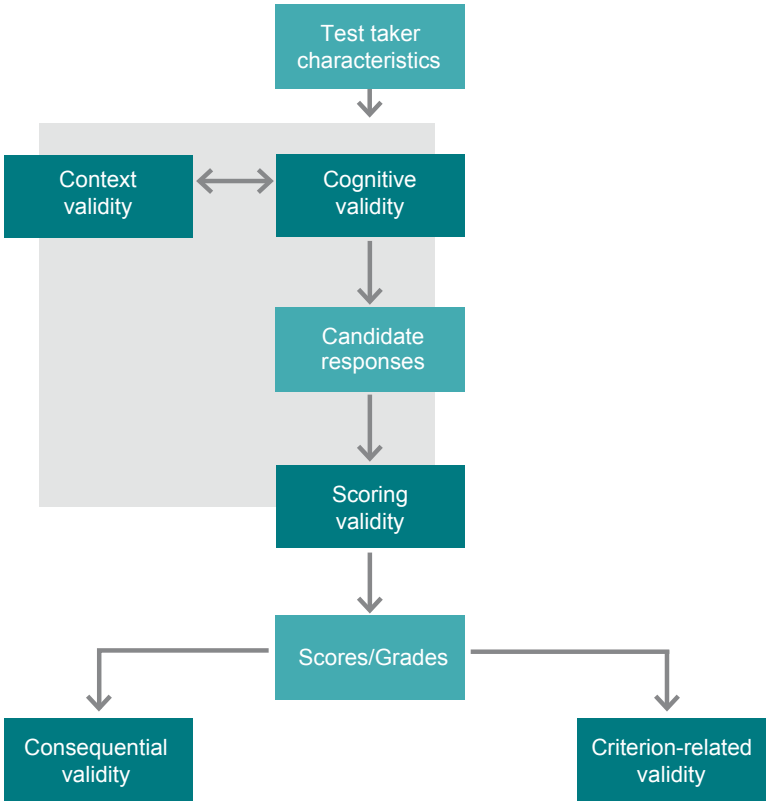
Members of ALTE play a key role in the development of *BULATS* and in the SurveyLang project, which has recently completed its survey on the language competences of secondary school students in several European countries for the European Commission.

For information on the ALTE audit system see Saville, N (2010) Auditing the quality profile: from code of practice to standards, *Research Notes* 39, 24–28.

As a point of principle, it should be noted that the individual qualities of validity, reliability, impact and practicality cannot be evaluated independently, rather their relative importance must be determined in order to maximise the overall 'usefulness' or 'fitness for purpose' of the exam. This is consistent with the 'unitary' approach to validity as described by Messick (1989) and Kane (2006).

Alongside VRIPQ, we also work with the socio-cognitive model. This is a theoretical framework of language learning and assessment that is used for more in-depth analysis of the validity of our exams. This model has been developed in collaboration with Professor Cyril Weir (Weir 2005) and draws on the work of Messick, Kane, Bachman and others. The approach is described as socio-cognitive in that the use of language is seen as a social phenomenon (the social dimension) and the language abilities to be tested are seen as mental constructs of the test taker (the cognitive dimension).

The diagram on the next page illustrates the principal direction of hypothesised relationships between elements of the socio-cognitive framework. It shows that, while all aspects of validity need to be considered at test development stages, some types of validity evidence cannot be collected until after the test event.



Key – test specifications account for all of these aspects of validity

4.1 Building a Validity Argument

The conceptualisation of validity described above requires a validity argument to be presented by the examination provider. A validity argument is a well-reasoned rationale in which an examination provider presents an overall evaluation of the intended interpretations and uses of the test which is being validated. This is consistent with the definition of validation as:

'the ongoing process of demonstrating that a particular interpretation of test scores is justified …' (Bachman and Palmer 1996:22).

In building and presenting a validity argument for our exams, responsible staff members carry out work to:

- set out our claims relating to the usefulness of the test for its intended purpose
- explain why each claim is appropriate by giving reasons and justifications
- provide adequate evidence to support the claims and the reasoning behind them.

We build up this evidence over time, based on the processes described in our test development and validation model (see Section 3.4). In other words, we begin to gather the evidence at the design and development stages and continue to do so for as long as the test remains operational.

Research and Thought Leadership and Validation

A key element of our validity argument is empirical evidence provided by our Research and Thought Leadership and Validation teams. The teams have expertise in fields including applied linguistics, educational measurement and statistical analysis. The role of the team is to provide rigorous quality assurance for Cambridge English examinations at every stage of the assessment process. This is done by carrying out both routine operational analysis and research projects. The latter may be proposed by staff from across the organisation, and often involves collaboration with external experts.

In the following sections we give an overview of some of the key concepts of validity and validation with some (brief) examples of Cambridge English practice along with references to more in-depth discussion and descriptions.

4.2 Validity

a. Construct-related aspects of validity

The construct of a test is the theory that the test is based on. For language tests therefore, this is the theory of language ability and for teaching qualifications it is the theory of teaching knowledge and practice. Construct validation is therefore about investigating whether the performance of a test is consistent with the predictions made from these theories (Bachman 1990:255). Construct validation activities can include those where experts analyse the content of tasks, those that involve study of the cognitive processes of candidates and those involving statistical analysis of underlying factors shared by a series of tasks or items. One statistical method used is Structural Equation Modelling.

Cambridge English test construct model

Our approach to language testing draws on the work of Canale and Swain (1980) and Bachman (1990) who have proposed models of communicative language ability, and on the approach taken by the Council of Europe in the Common European Framework of Reference (2001) and its earlier specifications: *Waystage 1990* (Van Ek and Trim 1998a) and *Threshold 1990* (Van Ek and Trim 1998b).

This means we see language proficiency in terms of language users' overall communicative ability subdivided into skills and sub-skills. Our position is that since each skill can be developed to different degrees or at different rates, it can be separately recognised and measured. In general terms we test the four main skills in separate test components of Listening, Reading, Speaking and Writing. In some contexts we also test a separate fifth component (language knowledge). We believe that each skill-focused component provides a unique contribution to the building of a profile of communicative language ability for an individual candidate.

- For an example of an exploration of construct see: Geranpayeh, A (2007) Using Structural Equation Modelling to facilitate the revision of high stakes testing: the case of CAE, *Research Notes* 30, 8–12.
- Zeronis, R (2015) Continuity and innovation: Updating FCE and CAE, *Research Notes 62*, 3–5.

We are also keen to increase our understanding of learner English at different levels of proficiency, from beginners to highly proficient users of the language. Current work on reference level descriptions as part of the English Profile Programme (EPP) is evidence of our concern with this area.

b. Cognitive- and context-related aspects of validity

Cognitive-related validity is concerned with the extent to which the cognitive processes employed by candidates are the same as those that will be needed in real-world contexts beyond the test. These real-world contexts are known as the Target Language Use (TLU) domain (Bachman and Palmer 1996). Context-related validity is concerned with the conditions under which the test is performed and so includes aspects such as the tasks, the rubric and the topic as well as the administration conditions.

Validation of these aspects should therefore include investigation of the degree to which the sample of items, tasks or questions on an examination are representative of the TLU domain in terms of relevance and coverage.

It should be noted that a feature common to all Cambridge English examinations, irrespective of which skill is being tested, is the inclusion of a variety of task and response types. This is supported by numerous researchers who have made the case that multiple-task tests allow for a wider range of language to be elicited and so provide more evidence of the underlying abilities tested, i.e. the construct, and contribute to the exam's fairness (e.g. Chalhoub-Deville 2001).

Test takers

In designing a test for a particular context and purpose, we profile the intended test takers in terms of their characteristics: demographic features (such as age, gender and language background), existing knowledge and prior learning experiences. We also continue to collect this information in an ongoing manner during operational phases on Candidate Information Sheets (CIS) to make sure that the test is still fit for purpose.

- Elliott, M (2013) Test taker characteristics, in Geranpayeh, A and Taylor, L (Eds) *Examining Listening: Research and Practice in Second Language Listening* (Studies in Language Testing, volume 35), Cambridge: Cambridge University Press, 36–76.
- Khalifa, H and Weir, C (2009) Test Taker Profile of Main Suite Examinations, in *Examining Reading: Research and practice in assessing second language reading* (Studies in Language Testing, volume 29), Cambridge: Cambridge University Press, 27–29.
- Vidakovic, I (2009) Profile of Skills for Life candidature in 2007–8, *Research Notes* 36, 28–30.

The authenticity of test content is a further important consideration. In designing authentic tasks, the relationship between the input and the expected response is a key factor. For more information on this issue, see the discussion of authenticity in Bachman (1990:301).

Test content

At the test design phase, we specify the domain of content that the examination is intended to represent. For *Cambridge English: Key (KET)* and *Cambridge English: Preliminary (PET)*, specification of this domain was linked to the development of the *Waystage 1990* (Van Ek and Trim 1998a) and *Threshold 1990* (Van Ek and Trim 1998b) specifications, respectively Levels A2 and B1 in the CEFR. The Handbooks for Teachers for these tests therefore contain lists of functions, notions, communicative tasks, grammatical areas and lexis.

Over the last 10 or so years however, the collection of ongoing data indicated that the candidature for these exams contained a growing and significant proportion of younger candidates, and a review was put in motion. This involved extensive consultation, involvement of suitable content experts and a review of the literature around the cognitive development of young learners. The findings of the review led to the creation of *Cambridge English: Key for Schools* and *Cambridge English: Preliminary for Schools*, which are aimed at a younger candidature. Since that point *Cambridge English: First for Schools* has also been created.

- Hackett, E (2009) Adapting testing materials for younger learners: developing KET and PET for Schools exams, *Research Notes* 36, 12–14.
- Moeller, A J, Creswell, J W and Saville, N (Eds) (2016) *Second Language Assessment and Mixed Methods Research*, (Studies in Language Testing, volume 43), Cambridge: Cambridge University Press.
- Papp, S (2009) Development of Can-do statements for KET and PET for schools, *Research Notes* 36, 8–12.

4.3 Reliability

Reliability concerns the extent to which test results are stable, consistent and free from errors of measurement.

a. Criterion-related aspects of validity

Criterion-referenced tests are ones in which candidates are assessed against specific target knowledge skills and abilities. Test scores are then an indication of what a candidate can and cannot do.

Criterion-related validity, therefore, aims to demonstrate that examination marks are systematically related to some other indicator of what is being measured. This may be to a framework of observable behaviours as described above or alternatively to another external indicator such as another examination. The aim is to build up an understanding of the comparability of the measure or measures.

Criterion-related validity is also concerned with the setting and maintaining of standards. Wherever multiple versions of a test are produced it is important to be able to show that these different versions are comparable and that the standard (e.g. in terms of proficiency level) has been maintained. Item banking and the use of common scales are two key ways that testing organisations can ensure comparability.

Item banking and the Cambridge English Scale

Our Local Item Banking System (LIBS) is an electronic system that allows us to build up and store detailed information about all tasks and items used within our tests. As described on page 18, we also have rigorous quality assurance procedures for production of our test items, including a stage called pretesting. At this stage objectively marked items (Listening, Reading and Use of English) are trialled on learners preparing for a particular exam. The response data is captured and a statistical approach called Rasch analysis (a branch of Item Response Theory) is used to estimate the difficulty of items in a process known as calibration.

Anchor tests are tests which contain items of known difficulty that are administered with pretests. The tests are designed to link versions at different levels with common items. This design enables cross-level linking to be built up over time. In this way a single measurement scale, called the Cambridge English Common Scale, has been constructed for all objectively marked papers.

Having a common scale for objective test items allows us to construct tests of known difficulties from these calibrated items. It also allows us to monitor stability across a number of test versions and sessions, to check for equivalence across those versions and to make comparisons across different suites of exams.

- For more information on LIBS see Marshall, H (2006) The Cambridge ESOL Item Banking System, *Research Notes* 23, 3–5.
- Jones, N and Saville, N (2016) *Learning Oriented Assessment: A Systemic Approach* (Studies in Language Testing, volume 45), Cambridge: Cambridge University Press.
- See Taylor, L (2004) Issues of Test Comparability, *Research Notes* 15, 2–5 for a discussion of the definitions of 'comparable' and 'equivalent'.

Additional experimental studies that have also provided useful evidence for constructing the Common Scale include a project using anchor tests administered alongside live exams, studies on candidates taking two exams, and re-calibrations of items used in computer-adaptive tests.

Common Scales for Writing and Speaking

Our Writing and Speaking tests use assessment scales that are linked to one another, ensuring that evaluation happens within a coherent common framework and allowing users to interpret performance across levels. The assessment scale for each exam can be found in the relevant Handbook for Teachers, and details of the development of the scales can be found in Lim (2012) and Galaczi, ffrench, Hubbard and Green (2011).

- Also see Hawkey, R and Barker, F (2004) Developing a common scale for the assessment of writing, *Assessing Writing 9* (3), 122–159.
- Galaczi, E and ffrench, A (2007) Developing revised assessment scales for Main Suite and BEC Speaking tests, *Research Notes* 30, 28–31.
- Jones, N and Saville, N (2016) *Learning Oriented Assessment: A Systemic Approach* (Studies in Language Testing, volume 45), Cambridge: Cambridge University Press.

Where appropriate, we also compare and align our examinations to external tests, frameworks and benchmarks. An early study that influenced the direction of our validation strategy is the

3-year Cambridge–TOEFL comparability study begun in 1987 and published as the first *Studies in Language Testing* volume.

As can be seen from the diagram on page 5, our exams are aligned to the Council of Europe's Common European Framework of Reference (CEFR). The CEFR defines a 6-level language proficiency scale which has been widely adopted in Europe and throughout the world to describe the level of courses, textbooks and qualifications, and to specify the objectives of teaching programmes and requirements for university entrance and employment. It has become essential for an exam provider to state how its exams relate to the CEFR and to provide evidence for this linking.

The concept of a set of 'European levels' was based in part on existing Cambridge English exam levels at the time, such as the *First Certificate in English (FCE)* and the *Certificate of Proficiency in English (CPE)*. Likewise the Cambridge exam levels have developed in response to Council of Europe initiatives (the *Waystage 1990* (Van Ek and Trim 1998a) and *Threshold 1990* (Van Ek and Trim 1998b) learning objectives, which have become CEFR Levels A2 and B1). Thus the historical link between the Cambridge levels and the CEFR is a close one, and we were also closely involved in the development phase of the CEFR.

We contributed to the piloting of the preliminary version of the Manual for *Relating language examinations to the Common European Framework of Reference for Languages* (Council of Europe 2009), and provided exemplars for the final version. Examples of how we have worked with the Manual procedures, both for maintaining alignment (*Cambridge English: First*) and for establishing alignment (Asset Languages), can be found in Martyniuk (2010).

b. Scoring-related aspects of validity

When a group of learners takes a test their scores will vary depending on their ability. Reliability in this sense is defined as the proportion of score variability caused by the ability measured, and not by other factors. The variability caused by other factors is called error. Sources of error include variation in test administration conditions, variation in the test takers such as motivation or tiredness, and variation in the examiners and in the tasks.

A key aim of an examination provider therefore, is to seek to minimise the effects of these sources of error. For example, by standardising the way marking or test administration is carried out.

Marking objectively scored tests (Reading, Listening and Use of English)

Objectively scored items are those that do not require a marker to make a subjective judgement; or in other words those where a marking key containing all possible answers can be given to the marker. In fact, questions requiring candidates to answer with a single letter can be completed on an Optical Mark Reader (OMR) sheet, scanned and marked by machine. 'Short answer' questions require the candidate to construct a response that may be a single word or a short sentence. The candidate also writes these answers on their OMR but they are currently marked by general markers rather than by machine.

- Jones, N (2016) 'No More Marking': An online tool for comparative judgement, *Research Notes* 63, 12–15.
- Khalifa, H and Weir, C (2009) General Marking: performance management, in *Examining Reading: Research and practice in assessing second language reading* (Studies in Language Testing, volume 29), Cambridge: Cambridge University Press, 276–280.

In terms of reporting reliabilities, a range of statistical evidence can be used, and only a brief overview can be given here. In general however, reliability for objectively scored tests is estimated and reported in terms of measures called reliability coefficients, one of which

is called Cronbach's Alpha. A perfectly reliable test would have a reliability coefficient of 1. Another key statistic in the reporting of reliability is the Standard Error of Measurement, which is concerned with the reliability of individual scores rather than the reliability of tests. It gives an indication of how dependable the test score of an individual is likely to be.

Calculating reliabilities on objectively scored tests

Cronbach's Alpha and the Standard Error of Measurement are routinely calculated. Whenever there are sufficient sub-group members to permit meaningful analysis, we also calculate the reliability of the sub-group scores. It should be noted that the task-based nature of most of our tests leads to the expectation that we will not achieve such high reliability indices as, for example, one could achieve on a longer discrete-point multiple-choice test. Despite this, reliabilities of .80 and above are regularly estimated.

- Beresford-Knox, N (2015) The role of Quality Management in ensuring accurate and meaningful test scores, *Research Notes 59*, 40–44.
- Cope, L (2009) CB BULATS: Examining the reliability of a computer-based test, *Research Notes* 38, 31–34.
- Geranpayeh, A (2004) Reliability in First Certificate in English objective papers, *Research Notes* 15, 21–23.

Writing tests

In our standard model for Writing tests, candidates complete two or more tasks designed to show a range of writing ability. Most exams are marked using an analytic mark scheme where separate marks are given for different aspects of writing ability, such as language use, organisation and communicative effect.

Writing examiners

Examiners must meet Minimum Professional Requirements (MPRs) before they are considered for the role. They then go through an induction, training and certification procedure before they are allowed to mark exams. During marking sessions, examiners are monitored by experienced Team Leaders and the whole marking process is in turn guided and monitored by a Principal Examiner. Most candidate scripts are now scanned, randomly allocated to examiners and marked on screen.

Where an examiner is found to have been marking erratically, and their scripts are re-marked by another examiner, they may be asked to retrain as appropriate. Examiners receive feedback about their performance and have to recertify regularly (ffrench, Bridges and Beresford-Knox 2012).

Speaking tests

In our standard model for Speaking tests, tests are taken by paired candidates face-to-face with two examiners, one of whom marks against a holistic scale and one of whom marks against an analytical scale. The latter scale covers grammar and vocabulary, discourse management, interactive communication and pronunciation.

The format of the tests, and the nature of the assessment criteria, reflect the broad, multi-faceted construct underlying these exams. This is not driven exclusively by lexicogrammatical accuracy, but includes a balance of important aspects of communicative competence such as the ability to produce coherent and relevant contributions, both in interactive and individual tasks.

Speaking examiners

New examiners go through a training process where they become familiar with the global issues underlying Cambridge English Speaking tests, including assessment, examiner roles, test format and the function of interlocutor frames. They then have to be certified, and thereafter go through a rigorous yearly standardisation process which focuses on the assessment process and includes a marks collection exercise.

After each test session we estimate reliability, accounting for the accuracy and consistency of the ratings made by the examiners (Galaczi 2005).

Grading

Grading is the process of setting cut-off scores for various grades. The Cambridge English approach to grading allows candidates' results to be compared from session to session and from year to year to ensure that grades in a particular examination reflect a constant standard. Reports and analyses which have been carried out on the score data, and in relation to various groups of candidates, are reviewed according to an established procedure. Grade boundaries are scrutinised and approved by senior management.

4.4 Impact

Assessment has important effects and consequences within the educational system and within society more widely. These effects are referred to as impact. Test takers in particular are affected because the results of tests are used to make important decisions about them which can affect their lives.

At Cambridge English Language Assessment we adopt the principle of impact by design. We strive to achieve positive impact in the contexts in which our assessments are used and we undertake to investigate this through our validation processes. In promoting positive effects on curricula and learning, we seek to design and develop test features that are consistent with those found in instructional programmes. To implement this principle and to integrate an action-oriented approach to investigating impact into the working practices of our organisation, we adhere to the following maxims of test impact:

Impact by design

Maxim 1 PLAN

Use a rational and explicit approach to test development

Maxim 2 SUPPORT

Support stakeholders in the testing process

Maxim 3 COMMUNICATE

Provide comprehensive, useful and transparent information

Maxim 4 MONITOR and EVALUATE

Collect all relevant data and analyse as required

These maxims are used to guide all aspects of the test design and delivery decision-making process. For example, we adhere to Maxim 1 by following a model for test development and validation which is based on a rational and explicit approach (see Section 3.4). Under Maxim 2, we recognise that stakeholders are also participants in the assessment process, and therefore, need support. This is an important aspect of the approach because examination systems only function effectively if all stakeholders collaborate to achieve the intended outcomes. In accordance with Maxim 3, we endeavour to provide accurate, comprehensive and comprehensible information about its tests (see Section 2). Various kinds of impact are anticipated at the design and development stages, and we put procedures into place when an examination becomes operational to collect relevant data and carry out routine analyses (Maxim 4). This information also allows us to monitor both the anticipated and unanticipated effects of our exams.

To learn more about the Cambridge English model for investigating impact, read:

- Saville, N (2012) Applying a model for investigating the impact of language assessment within educational contexts: The Cambridge ESOL approach, *Research Notes* 50, 4–8.

Investigating bias

We design our examinations to be fair and not biased in favour of one group of test takers over another. One way in which we do this is by having clear guidance for item writers in the item writer guidelines for each component. These include lists of suitable and unsuitable topics so as to avoid distressing or distracting certain groups of candidates. We make sure that item writers understand who the target users are, and that they consider aspects such as the level of cognitive processing that candidates can cope with and the cultural contexts they will be used to. We are also careful not to test general knowledge or technical material as this could disadvantage certain groups.

These issues are monitored as part of the operational testing cycle, and are investigated, as appropriate, through research studies using a type of analysis called Differential Item Functioning (DIF).

- Geranpayeh, A (2008) Using DIF to explore item difficulty in CAE listening, *Research Notes* 32, 16–23.

Impact studies

A range of impact studies have been carried out by staff, consultants and independent researchers. For example, studies have been carried out which track academic performance at university on the basis of the *IELTS* score at entry.
See *IELTS* research reports at www.ielts.org.

Research Notes 50 focuses on the impact of Cambridge English exams in a range of educational contexts:

- Ashton, K, Salamoura, A and Diaz, E (2012) The BEDA impact project: A preliminary investigation of a bilingual programme in Spain, *Research Notes* 50, 34–42.
- Chambers, L, Elliott, M and Jianguo, H (2012) The Hebei Impact Project: A study into the impact of Cambridge English exams in the state sector in Hebei province, China, *Research Notes* 50, 20–23.
- Gu, X, Khalifa, H, Yan, Q and Tian, J (2012) A small-scale pilot study investigating the impact of *Cambridge English: Young Learners* in China, *Research Notes* 50, 42–48.
- Gu, X and Saville, N (2012) Impact of *Cambridge English: Key for Schools* and *Cambridge English: Preliminary for Schools* – parents' perspectives in China, *Research Notes* 50, 48–56.
- Khalifa, H, Nguyen, T and Walker, C (2012) An investigation into the effect of intensive language provision and external assessment in primary education in Ho Chi Minh city, Vietnam, *Research Notes* 50, 8–19.
- Salamoura, A, Hamilton, M and Octor, V (2012) An initial investigation of the introduction of Cambridge English examinations in Mission laïque française schools, *Research Notes* 50, 24–33.

4.5 Practicality

Practicality is an integral part of the concept of test usefulness and affects many different aspects of an examination. It can be defined as the extent to which an examination is practicable in terms of the resources necessary to produce and administer it in its intended context and use. A practical examination is one that does not place an unreasonable demand on available resources.

We consult relevant stakeholders during test development and revision processes on aspects of practicality, such as test length. For example, while longer tests can increase validity because they capture more measurement data they may be impractical to administer. In addition an overly long exam could induce fatigue in candidates, which in turn could introduce error into the measurements.

We also work with our network of centres to make sure that the systems and processes we use are up-to-date and flexible enough to allow effective and efficient administration. In line with our educational mission we wish to maintain access for the widest proportion of candidates possible. This means we must pay attention to holding costs at a reasonable level and make sure that our tests can be administered sufficiently frequently.

Further study

You may wish to follow up on topics covered in Section 4 of this document by attending one of the following courses:

Assessment of Language Proficiency

This course is run at Cambridge English every year, co-ordinated by the Research and Validation Group.

Certificate in the Principles and Practice of Assessment

A certificated course run by the University of Cambridge Institute of Continuing Education with Cambridge Assessment.

Further reading

Construct

The Cambridge English approach is explained in detail in four volumes in the *Studies in Language Testing* series:

Geranpayeh, A and Taylor, L (Eds) (2013) *Examining Listening: Research and practice in assessing second language listening* (Studies in Language Testing, volume 35), Cambridge: Cambridge University Press.

Khalifa, H and Weir, C (2009) *Examining Reading: Research and practice in assessing second language reading* (Studies in Language Testing, volume 29), Cambridge: Cambridge University Press.

Shaw, S and Weir, C (2007) *Examining Writing: Research and practice in assessing second language writing* (Studies in Language Testing, volume 26), Cambridge: Cambridge University Press.

Taylor, L (Ed.) (2011) *Examining Speaking: Research and practice in assessing second language speaking (Studies in Language Testing,* volume 30), Cambridge: Cambridge University Press.

Weir, C J, Vidakovic, I and Galaczi, E D (2013) *Measured constructs: A history of the constructs underlying Cambridge English examinations 1913–2012* (Studies in Language Testing, volume 37), Cambridge: Cambridge University Press.

Jones, N (2012) Reliability and Dependability, in Fulcher, G and Davidson, F (Eds) *The Routledge Handbook of Language Testing,* Routledge.

Impact

Green, A (2005) Staying in Touch: tracking the career paths of CELTA graduates, *Research Notes* 19, 7–11.

Green, A (2007) *IELTS Washback in Context: Preparation for academic writing in higher education* (Studies in Language Testing, volume 25), Cambridge: Cambridge University Press.

Hawkey, R (2005) The CPE Textbook washback study, *Research Notes* 20, 19–20.

Hawkey, R (2006) *Impact Theory and Practice: Studies of the IELTS test and Progetto Lingue 2000* (Studies in Language Testing, volume 24).

Hawkey, R and Ellis, S (2016) Impacts of international language assessments on multilingualism: Evidence from an iterative impact study of Progetto Lingue 2000, in Docherty, C and Barker, F (Eds) *Language Assessment For Multilingualism: Proceedings of the ALTE Paris Conference, April 2014*, (Studies in Language Testing, volume 44), Cambridge: Cambridge University Press, 182–208.

The theme of *Research Notes* 34 and 35 is the educational impact of language assessment in a range of contexts, including the following articles:

Tsagari, D (2009) Revisiting the concept of test washback: investigating FCE in Greek language schools, *Research Notes* 35, 5–10.

Valazza, G (2008) Impact of TKT on language teachers and schools in Uruguay, *Research Notes* 34, 21–26.

Also see:

Saville, N (2010) Developing a model for investigating the impact of language assessment, *Research Notes* 42.

The CEFR

The *Common European Framework of Reference for Languages: Learning, teaching, assessment*, the *Manual for Language Test Development and Examining* and the Manual for *Relating language examinations to the Common European Framework of Reference for Languages* are published by the Council of Europe at www.coe.int.

The Cambridge English CEFR web pages give a comprehensive overview of the historical, conceptual and empirical links between Cambridge English and the CEFR: www.cambridgeenglish.org/cefr

Volume 33 in the *Studies in Language Testing* series looks at the CEFR:

Martyniuk, W (Ed.) *Aligning Tests with the CEFR.*

The theme of *Research Notes* 37 is the CEFR, including the following articles:

Khalifa, H and ffrench, A (2009) Aligning Cambridge ESOL exams to the CEFR: issues and practice, 10–14.

Milanovic, M (2009) Cambridge ESOL and the CEFR, 2–5.

North, B (2014) *The CEFR in Practice*, English Profile Studies volume 4, Cambridge: UCLES/Cambridge University Press.

English Profile Programme (EPP) research

See www.englishprofile.org

Ćatibušić, B and Little, D (2014) *Immigrant Pupils Learn English: A CEFR-Related Empirical Study of L2 Development,* English Profile Studies volume 3, Cambridge: UCLES/Cambridge University Press.

Green, A (2012) *Language Functions Revisited: Theoretical and Empirical Bases for Language Construct Definition Across the Ability Range*, English Profile Studies volume 2, Cambridge: UCLES/Cambridge University Press.

Harrison, J and Barker, F (Eds) (2015) *English Profile in Practice,* English Profile Studies volume 5, Cambridge: UCLES/Cambridge University Press.

Hawkins, J A and Filipović, L (2012) *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*, English Profile Studies volume 1, Cambridge: UCLES/Cambridge University Press.

Appendix: VRIPQ Framework

This is a summary of the framework that we use for reviewing and evaluating our exams. Within this framework we specify the evidence that adds to the overall validity argument. By following the page references here you can find further reading on these points.

VRIPQ	Framework clauses	page
Construct Validity Cognitive and context-related aspects	**A** Define the construct of an assessment and its components, making reference to appropriate theoretical model(s) and how these are operationalised.	22
	B Profile the intended test takers.	23
	C Specify the intended purpose, context of use and impact of the test.	17, 23, 28
	D Determine the proficiency level and Target Language Use domain(s) for an assessment, ensuring that a test includes adequate coverage.	23
	E Define the test and task characteristics and how these are related to the domain of knowledge or skill being assessed.	23
	F Implement procedures that ensure that bias in test items is minimised.	29
	G Provide appropriate information to stakeholders about the construct and content of an examination, including how to interpret results. Provide additional supporting evidence whenever substantial changes are made to an examination.	10
	H Implement administration and security procedures that ensure that test-taking conditions are equivalent, and that validity is not compromised.	6, 7, 15
	I Carry out analysis of examinations and their components to confirm that the intended skills and cognitive processes are being tested.	22
	J Carry out analysis of likely sources of construct-irrelevant variance.	22, 26
Reliability Criterion-related aspects	**A** Develop and validate rating scales for Speaking and Writing tests, and for other performance assessments.	25
	B Provide appropriate evidence to stakeholders where claims are made of comparability to other tests or criterion measures, and to predictions of future performance.	24, 25
	C Have a rationale and procedures to ensure that test materials are appropriately calibrated so that standards are set and maintained.	24
	D Implement procedures that ensure that expert judgement is involved in the quality and stability of test content, and that there is appropriate selection and training of these experts.	14, 18
Scoring-related aspects	**E** Investigate statistical performance of items and tasks to ascertain if they are performing as expected.	18, 26
	F For objectively scored tests: estimate the reliability of exams including sub-component scores and combinations of marks where appropriate and provide such information to users. Also provide Standard Errors of Measurement for mark regions within which decisions about individuals are made. Provide this information in such a way as to enable test users to judge whether the results are sufficiently reliable for their intended use.	27, 28

VRIPQ	Framework clauses	page
	G For performance assessments (inc. speaking and writing): estimate the degree of agreement between examiners.	27, 28
	H Have a rationale and procedures for selecting, training and monitoring general markers, also for allocating candidate performances to markers and adjusting any discrepancies in their marking.	26
	I Have a rationale and procedures for selecting, training and monitoring examiners for performance assessments (inc. speaking and writing); also for allocating candidate performances to examiners and for adjusting any discrepancies in their marking.	27, 28
Impact Consequential aspects	**A** Monitor who is taking the examination (i.e. profile the test takers).	29
	B Carry out Differential Item Functioning analyses to identify potential bias.	29
	C Monitor who is using the examination results and for what purpose.	29
	D Monitor who is teaching towards the examination and under what circumstances, and what kinds of courses and materials are being designed and used to prepare test takers.	29
	E Monitor what effect the examination has on public perceptions generally (e.g. regarding educational standards) and/or how the examination is viewed by those directly involved in educational processes (e.g. by students, examination takers, teachers, parents, etc.) and/or how the examination is viewed by members of society (e.g. by politicians, businesspeople, etc.).	28
Practicality	**A** Ensure that test materials can be produced in sufficient quantity and quality within the time frame required.	30
	B Ensure that the test can be effectively administered with the available resources.	30
	C Ensure that results can be released in the time frame required.	30
Quality Management	**A** Appropriately define core and support processes.	16
	B Maintain appropriate quality assurance processes across the organisation.	14, 15
	C Support staff with appropriate training and guidance to allow them to carry out their roles effectively and efficiently.	14
	D Maintain appropriate support systems for stakeholders such as helpdesk, web support and support products.	10, 14
	E Consult with appropriate stakeholders regarding the development of, and ongoing operational delivery of, products and services and monitor and measure customer satisfaction in appropriate ways.	10, 17, 18
	F Ensure that legal and statutory requirements are taken into account, including data protection.	6, 7, 13, 14

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999) *Standards for Educational and Psychological Testing* (2nd edition), Washington DC: American Educational Research Association.

Bachman, L (1990) *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press.

Bachman, L and Palmer, A (1996) *Language Testing in Practice*, Oxford: Oxford University Press.

Canale, M and Swain, M (1980) Theoretical bases of communicative approaches to second language teaching and testing, *Applied Linguistics* 1, 1–47.

Chalhoub-Deville, M (2001) Task-based assessments: Characteristics and validity evidence, in Bygate, M, Skehan, P and Swain, M (Eds) *Researching Pedagogic Tasks*, London: Longman (167–185).

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.

Council of Europe (2009) *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR), A Manual*, Strasbourg: Council of Europe.

Davies, A (Ed.) (2008) *Assessing Academic English: Testing English proficiency, 1950–1989 – the IELTS solution* (Studies in Language Testing, volume 23), Cambridge: Cambridge University Press.

ffrench, A, Bridges, G and Beresford-Knox, J (2012) Quality assurance: A Cambridge ESOL system for managing writing examiners, *Research Notes* 49, 11–17.

Galaczi, E D (2005) Upper Main Suite speaking assessment: towards an understanding of assessment criteria and oral examiner behaviour, *Research Notes* 20, 16–19.

Galaczi, E D, ffrench, A, Hubbard, C and Green, A (2011) Developing assessment scales for large-scale speaking tests: A multiple-method approach, *Assessment in Education: Principles, Policy & Practice* 18 (3), 217–237.

Geranpayeh, A and Taylor, L (Eds) (2013) *Examining Listening: Research and practice in assessing second language listening,* Cambridge: Cambridge University Press.

Green, A (2007) *IELTS Washback in Context: Preparation for academic writing in higher education* (Studies in Language Testing, volume 25), Cambridge: Cambridge University Press.

Hawkey, R (2006) *Impact Theory and Practice: Studies of the IELTS test and Progetto Lingue 2000* (Studies in Language Testing, volume 24), Cambridge: Cambridge University Press.

Hawkey, R (2009) *Examining FCE and CAE: key issues and recurring themes in developing the First Certificate in English and Certificate in Advanced English exams* (Studies in Language Testing, volume 28) Cambridge: Cambridge University Press.

Hawkey, R and Barker, F (2004) Developing a Common Scale for the Assessment of Writing, *Assessing Writing* 9 (2), 122–159.

Kane, M T (2006) Validation, in Linn, R L (Ed.) *Educational Measurement* (4th edition), Washington, DC: The National Council on Measurement in Education and the American Council on Education, 17–64.

Khalifa, H and Weir, C (2009) *Examining Reading: Research and practice in assessing second language reading* (Studies in Language Testing, volume 29), Cambridge: Cambridge University Press.

Lim, G S (2012) Developing and validating a mark scheme for writing, *Research Notes* 49, 6–10.

Martyniuk, W (Ed.) (2010) *Aligning Tests with the CEFR* (Studies in Language Testing, volume 33), Cambridge: Cambridge University Press.

Messick, S (1989) Validity, in Linn, R L (Ed.) *Educational Measurement* (3rd edition), New York: Macmillan, 13–103.

O'Sullivan, B (Ed.) (2006) *Issues in Testing Business English: the revision of the Cambridge Business English Certificates* (Studies in Language Testing, volume 17), Cambridge: Cambridge University Press.

Shaw, S and Weir, C (2007) *Examining Writing: Research and practice in assessing second language writing* (Studies in Language Testing, volume 26), Cambridge: Cambridge University Press.

Taylor, L (Ed.) (2011) *Examining Speaking: Research and practice in assessing second language speaking* (Studies in Language Testing, volume 30), Cambridge: Cambridge University Press.

van Ek, J A and Trim, J L M (1998a) *Waystage 1990*, Cambridge: Cambridge University Press.

van Ek, J A and Trim, J L M (1998b) *Threshold 1990*, Cambridge: Cambridge University Press.

Weir, C J (2005) *Language Testing and Validation: An Evidence-Based Approach*, Oxford: Palgrave.

Weir, C and Milanovic, M (Eds) (2003) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002* (Studies in Language Testing, volume 15), Cambridge: Cambridge University Press.

Find out more

Contact us:

Cambridge English
Language Assessment
1 Hills Road
Cambridge
CB1 2EU
United Kingdom

www.cambridgeenglish.org/helpdesk

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Cambridge English Language Assessment is part of the University of Cambridge. We develop and produce the most valuable range of qualifications for learners and teachers of English in the world. Over 5 million people in 130 countries take our exams every year. Around the world over 20,000 universities, employers, government ministries and other organisations rely on our exams and qualifications as proof of English language ability. Cambridge English exams are backed by the work of the largest dedicated research team of any English language test provider.

Cambridge English Language Assessment – a not-for-profit organisation.

A DIVISION OF CAMBRIDGE ASSESSMENT

ALTE

bsi. ISO 9001 Quality Management
FS 514675

7458293649

All details are correct at the time of going to print in May 2016