



Cambridge  
Assessment



Trusted experts in  
international *assessment*  
*and learning*

Cambridge Assessment research papers  
at AEA-Europe 2019



Cambridge Assessment  
Network



Postgraduate Advanced  
Certificate in  
*Educational Assessment*

## *New for 2020:* Cambridge PGCA is now worth 90 credits at Master's Level

From 2020 the Cambridge PGCA is evolving into the *Postgraduate Advanced Certificate in Educational Assessment* – a 15-month, part-time course now worth 90 credits at Master's level (Level 7).

The qualification continues to be practice based and is designed to directly impact your work as you learn to apply various research methodologies to your professional context.

You'll learn through a mix of online learning and four Saturday Day Schools in Cambridge led by experts from Cambridge Assessment and the University of Cambridge Faculty of Education.

On successful completion of the course you'll be awarded a Postgraduate Advanced Certificate in Educational Studies (PACES).

---

*“It was a very valuable learning experience that I will be returning to repeatedly over the next 12 months, re-reading, re-thinking and adapting practice.”*

---

**Find out more: [www.cambridgeassessment.org.uk/pgca](http://www.cambridgeassessment.org.uk/pgca)**

## About us

Cambridge Assessment is a leading expert in international education and assessment. We exist to help people learn, achieve and prove their true potential in over 170 countries across the world. A part of the University of Cambridge, we have an unrivalled depth of experience in national education systems, international education and English language learning. Accepted and trusted by thousands of organisations worldwide, our programmes and qualifications are recognised as the global mark of excellence by teachers, schools, organisations and governments. A not-for-profit, we are driven by our commitment to deliver on the University's mission to contribute to society through the pursuit of education, learning and research to the highest international levels of excellence.

At Cambridge Assessment, the reliability of our assessments stems from evidence-based and research-led approaches to all products, services and new developments. We have the largest research capability of its kind in Europe, with more than 80 researchers across the Group.

Our researchers conduct and publish authoritative research in order to validate, improve and develop our assessments and services, and to influence thinking and policy on educational assessment nationally and internationally.

We deliver our qualifications and services through:



Helping people learn English and prove their skills to the world.  
[cambridgeenglish.org](http://cambridgeenglish.org)



Preparing school students for life, helping them develop an informed curiosity and a lasting passion for learning.  
[cambridgeinternational.org](http://cambridgeinternational.org)



A leading UK awarding body, providing qualifications which engage people of all ages and abilities at school, college, in work or through part-time learning programmes.  
[ocr.org.uk](http://ocr.org.uk)



Providing professional development programmes and expert training for assessment professionals and organisations in the UK and internationally.  
[cambridgeassessment.org.uk/the-network](http://cambridgeassessment.org.uk/the-network)

## WORKSHOP 1

### Is assessment 'fair'?

Wednesday 13 November, 09:00–16:30, Castelo 9

Authors and presenters: Isabel Nisbet, Faculty of Education, University of Cambridge and Stuart Shaw, Cambridge Assessment International Education



Fairness in assessment is both complex and contentious. Assessment experts may disagree on whether scores from a particular testing programme are fair. However, most assessment experts agree that fairness is a fundamental aspect of validity. As a consequence, fairness has been elevated to a greater position of prominence in the assessment literature, so much so, that it is now considered one of the three primary measurement standards that must be met to legitimise a proposed test (the other two being validity and reliability). In this workshop, we will distinguish between different uses of 'fair' which have relevance to assessment. Then, we will identify some of the 'lenses' used to examine fairness in assessment and suggest a framework of questions which can be applied to lenses (measurement, legal, social justice and philosophy). Each approach will be subject to a common set of questions which will investigate whether there is an established consensus on fairness, whether that consensus should be questioned, what comprises an area of dispute, and what the implications are for other lenses. The research will culminate in a fairness agenda for the 21st century.

# From opinion to evidence: Transforming organisational culture in two Awarding Organisations

Thursday 14 November, 11:30–12:45

Authors: Alison Rodrigues, Cambridge Assessment International Education  
and Sarah Hughes, OCR

Presented by: Alison Rodrigues



Research evidence is just one factor influencing decision making. Other factors include fit with culture, practicality, financial considerations, colleagues' beliefs, dominance of personalities, professional wisdom, and policy and practice. We believe that research evidence should be a key driver for decision making. To this end a research-use initiative was launched based on a monitoring and evaluation framework and applied in two Awarding Organisations. Dimensions of the framework are: strategy and direction (what is the vision and mission?), research management (what processes and protocols are in place?), outputs (what mechanisms for sharing research are in place and are they appropriate?), uptake (are people accessing and sharing research?), impact (has research had any longer term impact?) and context (how have other factors affected impact?). Research use in the two Awarding Organisations is compared in terms of the aspects of the framework. Differences in research use across the two Awarding Organisations can be described in terms of contextual factors such as: organisational culture including leadership; regulatory pressures; level of embeddedness of the research function; whether research is expected to justify or inform decision making; how agile the research process is; intended audience; and appetite for research evidence.

## SESSION C – TEST DEVELOPMENT I

### Spoilt for choice? Is it a good idea to let students choose which questions they answer in an exam?

Thursday 14 November, 13:45–14:15, Castelo 8

Authors: Tom Bramley and Victoria Crisp,  
Cambridge Assessment Research Division

Presented by: Tom Bramley



For many years, question choice has been used in some UK public examinations, with students free to choose which questions they answer from a selection (within certain parameters). In this paper we distinguish some different scenarios in which choice (or 'optionality') arises and explore the arguments for and against using optional questions. In particular we discuss i) whether having optional questions makes exams fairer or more valid; and ii) whether it is possible to discover if optional

questions are of different difficulty and hence make statistical adjustments to students' scores that can allow for this.

# Re-heated meals: Revisiting the teaching, learning and assessment of practical cookery in schools

Thursday 14 November, 14:15–14:45, Castelo 10

Authors: Gill Elliott and Jo Ireland,  
Cambridge Assessment Research Division

Presented by: Gill Elliott



The place of practical cookery within school subjects in England has, in recent years, been debated as part of concerns about the nation's health and obesity. Cookery has been a school subject for over a century, but has only ever held a minority place in the curriculum.

In 2017 we surveyed teachers of practical cookery in schools, in a repeat of a survey first carried out in 2007.

We asked them about the ingredients used and the skills taught in practical cookery lessons at school and also about the issues they faced delivering practical cookery teaching and assessment through the school food curriculum.

We have found that the nature of the products being taught in schools has changed, with less emphasis on sugary baked items than previously, which is consistent with the development of healthy eating initiatives and awareness. However, many of the issues surrounding the teaching of cookery skills in schools identified in 2007, such as insufficient equipment, lesson time and parental support, remain unchanged. In this presentation we will discuss the implications of this research and the role of practical cookery teaching and assessment in schools in the future.

## SESSION I – COMPARATIVE JUDGEMENT I

# A framework for describing comparability between alternative assessments

Thursday 14 November, 15:45–16:15, Castelo 6–7

Authors: Victoria Crisp, Cambridge Assessment Research Division,  
Stuart Shaw, Cambridge Assessment International Education  
and Sarah Hughes, OCR

Presented by: Stuart Shaw



The credibility of an Awarding Organisation is reliant upon the claims it makes about its assessments (including comparability claims) and on the evidence it can provide in order to support such claims. For example, for syllabuses with options, such as the choice to conduct coursework or take an alternative to coursework exam, there is a qualification claim that overall candidates' results are comparable regardless of the choice made. This presentation describes a study which sought to design a structure that can be used to

evaluate comparability between alternative assessments. The study was undertaken in two phases. The first phase of the research focused on the development of a framework for evaluating comparability against a set of four standards as well as a separate recording form for capturing declared comparability intentions and how well these have been achieved. In the second phase of the study, the framework was piloted using two assessment contexts: on-screen and paper-based tests; and an Alternative to Practical and a Practical test. Outcomes from the pilot, using two experts engaged with the framework and form, are summarised in terms of the comprehensibility, usefulness and frequency of application of the comparability framework.

# Student engagement with on-screen assessments: A systematic literature review

Thursday 14 November, 15:45–16:15, Castelo 4–5

Author and presenter: Carla Pastorino, Cambridge Assessment  
International Education



Assessment delivery models that include at least one on-screen component have become increasingly common. One often-cited, learner-centred reason for adopting on-screen assessments (OSAs) relates to their potential for being more engaging for candidates who are assumed to participate in daily digital activities. However, it is possible that what is known about digital activities for entertainment may not apply for assessment tasks. The question of whether OSAs are engaging

for candidates and, critically, whether they are more engaging than their paper-based counterparts, remains open.

To answer this question, in this paper we describe the first stage of an ongoing research project, a systematic literature review. Its objective was to gain a better understanding of what is known about candidate engagement with OSAs and to provide the parameters for the design of subsequent behavioural experiments (phase 2). Initial results revealed that while the topic of engagement with on-screen materials has been abundantly explored with regard to learning, more and more nuanced investigations are required to conclude whether OSAs are more or less engaging. We also discuss the various ways in which engagement is defined in this context and the characteristics of OSAs that have been studied in relation to engagement.

## SESSION J – ON-SCREEN ASSESSMENT

# On-screen assessments for young learners: Considerations for on-screen item type design and usage

Thursday 14 November, 16:15–16:45, Castelo 4–5

Author and presenter: Sanjay Mistry, Cambridge Assessment  
International Education



Understanding how younger learners interact with on-screen assessments is essential to designing valid and fair on-screen assessments for this age group. A study was conducted to investigate how young learners (age 6–12) interact with different on-screen item types, in terms of cognitive and motor skills and to inform design and functionality considerations. The methodology was based on two phases of observational research of on-screen tests, using past paper-based assessment content. Phase 1 consisted of remote observations

conducted by teachers in India, Indonesia and UAE. Phase 2 involved a validation of the findings generated in phase 1 through face-to-face observations of learners conducted in India and Indonesia. Key findings indicated significant interactions between learner age, learner region and their ability to complete on-screen item types, raising several key issues pertaining to how socio-cultural differences in learners' exposure to technology and developmental differences in cognitive abilities impact the validity of using on-screen assessments across differing contexts and age groups. Further research will include item prototyping with young learners and exploring in detail on-screen item type design and functionality.

# Tests as texts: Investigating test questions from a sociolinguistic perspective

Thursday 14 November, 16:15–16:45, Castelo 10

Author and presenter: Filio Constantinou, Cambridge Assessment Research Division



Assessment has the potential to transform teaching and learning. For this potential to be realised, the tools via which assessment is performed need to be understood in depth. One of the most commonly used assessment tools in education is the written test. To date, written tests have been investigated mainly as measurement tools or as socio-political constructs. However, they are neither merely measurement tools nor merely socio-political constructs. In the first instance, they are linguistic entities, or texts. In an attempt to

illuminate this less recognised facet of tests, this study investigated written tests from a sociolinguistic perspective. The study was informed by sociolinguistic theory that suggests that the linguistic features of a text are not arbitrary but are dictated by the situational context of communication (e.g. who is writing, for whom, for what purpose). Drawing on this theory, this research sought to understand the linguistic design of written questions both at a structural and a functional level. Specifically, it aimed to (a) identify the most common linguistic features of written questions and, (b) explain their prevalence by reference to the situational context. This presentation will report the findings of the study and discuss their implications.

## SESSION J – ON-SCREEN ASSESSMENT

### The use of touchscreen vs. standard devices for marking high-stakes exams

Thursday 14 November, 16:45–17:15, Castelo 4–5

Authors and presenters: Sarah Hughes and Martina Kupalja, OCR



In addition to other upgraded and some new features, the new RM Assessor3 (RMA3) marking software enables the use of standard (non-touchscreen) as well as touchscreen devices for the purpose of marking exams. OCR is looking to use RMA3 for some exams in 2019 and the old version of the software will be completely replaced by RMA3 from June 2020. Before moving to marking using RMA3, OCR wants to investigate whether marking on touchscreen devices affects the quality of marking and users' experience of marking (compared to marking on standard devices when using the same version of the RM software). The main aim of the study was to compare the quality of marking using RMA3 on touchscreen and RMA3 standard devices. An additional aim was to record examiners' experience of marking using these two modes of marking (touchscreen and standard). Findings will be presented in relation to the quality of marking at an exam and question level, and markers' experience of marking using standard and touchscreen devices will be discussed.



## Assessment of problem-solving skills

Friday 15 November, 09:30–10:00, Castelo 10

Authors: Martina Kuvalja, OCR, Stuart Shaw and Giota Petkaki,  
Cambridge Assessment International Education and Sarah Matthey,  
Cambridge Assessment Research Division

Presented by: Martina Kuvalja and Stuart Shaw



It is crucial to provide an accessible description of the theoretical construct(s) which underlie assessments. This is especially important for exams that attempt to elicit complex, higher-order constructs that are 'hard-to-measure' (Stecher & Hamilton, 2014). If these construct(s) are not well defined and understood, then it will be difficult to support the claims we wish to make about the usefulness of the assessments, including claims that they do not suffer from factors such as construct under-representation and construct-irrelevant variance. This work focused on one such skill, problem-solving, and aimed to identify problem-solving processes and behaviours described in the literature and to explore how these are usually assessed. Two specific problem-solving contexts are investigated: domain-general (cross-curricular problem-solving for which a specific curricular knowledge is not required) and domain-specific (specific to a certain domain/subject and requires a certain level of subject knowledge).

Different assessment models for assessing problem-solving skills are presented and analysed through examples from PISA and Cambridge Assessment International Education assessments. Validity issues associated with each model are discussed and the recommendations for assessment design are made in order to improve the authenticity of assessment tasks and, therefore, to minimise threats to construct validity.

SESSION DD – COMPARATIVE JUDGEMENT II

## Moderation of non-exam assessments: A novel approach using comparative judgement

Friday 15 November, 14:30–15:00, Castelo 10

Authors: Lucy Chambers, Sylvia Vitello and Carmen Vidal Rodeiro,  
Cambridge Assessment Research Division

Presented by: Sylvia Vitello and Carmen Vidal Rodeiro



In England, many high-stakes qualifications include non-exam assessments that are marked by the teachers rather than external examiners. Awarding bodies then apply a moderation process to bring the marking of these assessments to an agreed standard. Current practice requires moderation to be conducted at centre level, with one moderator per centre who builds up a holistic view of the centre's approach to marking. As each centre is only viewed by one moderator, this raises challenges with regard to holding the standard across centres – this is currently overcome using standardisation and monitoring procedures.



In recent years, technological advances have allowed electronic submissions of candidates' work (e.g., portfolios). This opens the door for novel ways of moderating that can move beyond the allocation of centres to individual moderators towards a scenario in which candidates' work is distributed across multiple moderators (without being bound by centre). Such new methods could ensure that the marking standard is consistently applied across centres.

This research investigated, using simulation, whether comparative judgement (a technique whereby a series of two pieces of work are compared side by side to generate a rank order of work) could offer a feasible, and potentially more efficient, alternative to the current moderation process.

## The CEFR as an assessment tool for learner linguistic and content competence: Assisting learners in understanding the language proficiency needed for specific content goals in the CLIL classroom

Friday 15 November, 15:00–15:30, Castelo 1–2

Author and presenter: Stuart Shaw, Cambridge Assessment International Education



The construction of an academic language proficiency scale whose model of reference is the Common European Framework of Reference for Languages (CEFR) has clear implications for Content and Language Integrated Learning (CLIL) pedagogy. CLIL introduces a cognitive dimension not explicitly treated in the CEFR – ‘using language to learn’. However, a descriptor scale for academic language proficiency is complex and multidimensional, to the extent that a functional description of academic language use inevitably introduces a range of factors: cognitive stage, general language proficiency, the processes and skills involved in mastering the specific curricular objectives of each subject area, as well as the processes and skills involved in learning in general. Neither can it be assumed that these processes and skills are the same across countries or cultures. An example of how an academic language scale may be employed in the CLIL classroom is in the application of learning outcomes. Both the content subject and the language used as the medium of instruction are similarly involved in defining the learning outcomes. The clarity of content and academic learning outcomes can be enhanced with references to academic CEFR descriptors. By way of illustration, a history lesson plan focusing on mediation activities is described.

## SYMPOSIUM

# The rare but persistent problem of errors in examination papers and other assessment instruments

Friday 15 November, 16:00–17:00, Castelo 9

Convenor: Irenka Suto, Cambridge Assessment Research Division

Discussant: Paul Newton, Ofqual



This symposium is about understanding why errors occasionally occur in examination papers and other assessment instruments, and why error detection can be slow despite the numerous checks included in most construction processes. We draw upon research on error reduction in complex sectors such as medicine, manufacturing, the nuclear industry and aviation. In recent decades, greater understanding of how and why errors occur in these domains has been credited with significant improvements in safety and quality, saving countless lives.



All three papers assume that most assessment instrument construction processes form a complex system, since many of the numerous latent conditions that influence human performance are difficult to identify and measure. We share the theoretical position that system-level failure engenders human failure, which in turn gives rise to manifested errors such as those that appear in assessment instruments. Aspects of causation and of pre-emptive action to minimise errors are discussed throughout the symposium. Together we argue that the educational assessment community could benefit greatly by adopting principles of best practice developed in other industries.

The papers comprising this symposium are:

1. 'To err is human' but it's time to go deeper. An analysis of human and system level challenges in the construction of assessment instruments. Suto, I., and Ireland, J. (Cambridge Assessment).
2. On the psychology of error: a process analysis method for understanding error detection during the construction of assessment instruments. Williamson, J., Ireland, J., Macinska, S., and Suto, I. (Cambridge Assessment).
3. How and why do errors occur? Insights from people directly involved in assessment instrument construction. Vitello, S., and Rushton, N. (Cambridge Assessment).

## Getting out of their heads – using concept maps to elicit teachers' assessment literacy

Friday 15 November, 17:00–18:20, Castelo 1–2

Authors: Martin Johnson and Victoria Coleman,  
Cambridge Assessment Research Division

Presented by: Martin Johnson



Although it is a key component of teacher professional competency, there are concerns in the UK that teachers have only limited assessment literacy (AL). Teacher AL is a difficult concept to define, and evaluating it represents a challenge. Many evaluations have considered it in a narrow sense, but it is more than simply the acquisition of assessment knowledge and related skills, since it implicates a teacher's beliefs and feelings about assessment that have been acquired over time.

We used a novel concept-mapping approach to elicit AL with a group of teachers who were also examiners. We wanted to see how formal examining affected their learning about assessment and helped to transform their understandings of assessment, and how this influenced their teaching. In this presentation we will outline the method in broad terms and discuss how it gains insight into embedded professional knowledge in ways that other methods find difficult.

SESSION W – POLICY

## The 'grey history' of assessment: Understanding the origins of England's new model of assessment of practical work in Science

**Saturday 16 November, 09:00–09:30, Castelo 6–7**

**Author and presenter: Tim Oates, Cambridge Assessment Research Division**



England recently has introduced into its high-stakes assessment a new model of assessment; one in which marks from practical work no longer contribute to the grades in the qualifications. This model caused considerable controversy, and was adopted by the national regulator in the midst of highly adverse reaction. Various organisations predicted a collapse of practical work in schools. However, initial piloting work suggested the opposite was occurring in the limited trial centres. On national roll-out, similar benefits seem to be occurring across the system. Critics continue to feel that practical work should contribute to grades, despite the positive findings of the initial and continuing evaluation studies.

The presentation will trace the history of the development of the new model and examine emerging evaluation research on its success and impact. In doing so it will reveal an aspect of 'grey' history not present in the official record. It will examine issues of context and background; why a radical and seemingly unpopular model was conceptualised and introduced. The analysis of the development and introduction of the new model gives insights into public policy-making and the technical issue of assessment design. We believe key elements support international comparative work and national policy formation across many nations.

## Assessment literacy – how does being an examiner enhance teachers' understanding of assessment?

Saturday 16 November, 09:30–10:00, Castelo 4–5

Authors: Martin Johnson and Victoria Coleman,  
Cambridge Assessment Research Division

Presented by: Victoria Coleman



Concerns have been raised that many teachers do not have sufficient assessment literacy (AL), and this has implications for teacher professionalism and classroom practice. AL is an important component of teacher professionalism. It encompasses the basic understandings, skills and applications that underpin a teacher's ability to use and understand assessment. AL also encompasses a teacher's beliefs and feelings about assessment. This means that the relationship between AL and assessment practice is complex and multidirectional.

Thinking about the transformation of teachers' AL, it is useful to use the metaphor of an 'assessment career'. AL is changeable over time and is influenced by both personal and professional experience. This makes it of interest to explore whether and how teachers' participation in formal examining influences their AL.

To explore the influence of examining on their AL we used concept maps and interviews with a sample of Science and English teacher-examiners. These were then used to develop a survey to explore the influence of examining on the development of AL amongst a wider sample of international teacher-examiners. The outcomes of our study will investigate the contribution that professional examining work has on transforming teachers' AL and any impact on their teaching practices.

At Cambridge Assessment, the reliability of our assessments stems from evidence-based and research-led approaches to all products, services and new developments. We have the largest research capability of its kind in Europe, with more than 80 researchers across the Group.

Our researchers conduct and publish authoritative research in order to validate, improve and develop our assessments and services, and to influence thinking and policy on educational assessment nationally and internationally.

[cambridgeassessment.org.uk/our-research](https://cambridgeassessment.org.uk/our-research)



Cambridge Assessment  
Triangle Building  
Shaftesbury Road  
Cambridge, CB2 8EA  
tel +44 (0) 1223 553311  
[cambridgeassessment.org.uk](https://cambridgeassessment.org.uk)

Find us on Twitter, LinkedIn,  
Facebook, Instagram and YouTube.