

**27 and 28 September 2021**

# **Education assessment research seminar**

## **Online dissemination of work by leading UK education assessment researchers**

Following a second summer in the UK where public examinations were not taken and grades were instead based on teacher judgements, there continues to be great public interest in issues relating to assessment and learning. This seminar will provide research insights into the effect of the pandemic on this generation of students and their teachers, and some of the latest thinking about the future of assessment in the UK.

**Register to attend.**

## **Monday 27 September**

### **Session 1 – the future, part 1: stakeholder perspectives (Chair: Gill Elliott)**

**09:00 – 10:30**

#### **1. Effective teaching and learning during the pandemic**

Alison Rodrigues (Head of Research Hub) and Lynda Bramwell (School Liaison Manager)  
Cambridge Assessment International Education

This project is one part of a larger piece of work looking at the impact of Covid-19. We looked at all aspects of the school experience during the pandemic. A mixed methods approach was used for the research. Cambridge schools from a variety of education sectors, across the globe, participated in the research. This presentation focuses on Cambridge International schools' interviews and describes what schools said about changes to their teaching and their lesson content, where they found challenges, how they overcame them and finally what of these on reflection they want to take back.

#### **2. How well do we understand wellbeing? Teachers' experiences in an extraordinary educational era**

Chris Jellis (Research Associate)  
Cambridge CEM

We explored teachers' experiences and concerns during and after England's second national school closure in early 2021. Our aim was to improve understanding of how teachers had been impacted in these unprecedented times, and of their support needs.

Our findings were that teachers' longer-term working conditions impacted their wellbeing far more than teaching through lockdown did. Interestingly, some of the longest-serving teachers were among those finding that time pressure and administration affected their wellbeing.

### **3. The use of remote invigilation: awarding organisation views on its introduction and impact**

Diana Tonin (Research Associate) and Stuart Cadwallader (Associate Director – Research)  
Ofqual

The use of Remote invigilation (RI) has rapidly proliferated in response to the pandemic. RI systems are designed to maintain the integrity of an assessment while providing learners with flexibility, allowing them to be assessed at their home or place of work. However, the adoption of RI presents both challenges and opportunities for assessment. This presentation will report on seven case study focus group interviews with staff from a variety of awarding organisations. We explore how RI has been deployed and experienced under the following five themes: diversity of RI systems, infrastructural barriers, malpractice, learner experience, and developing capability.

### **4. The future of qualifications and assessment in England: stakeholder consultation outcomes for bottom-up reform**

Hayley Dalton (Head of Vocational Research)  
Pearson

As part of a 'post-pandemic' review of the qualification and assessment system, a wide-reaching consultation took place in early 2021, involving over 6000 participants, probing views on the 14-19 phase education phases in England.

This paper discusses the extent to which the data collected from this broad stakeholder group show how qualifications and assessments are valued in England. The research offers perspectives on the extent to which bottom-up approaches to education reform can or should be implemented on a national level.

### **5. Student voice: shaping the future**

Corina Balaban (Researcher), Katy Finch (Researcher, Research and Regulation)  
and Jeanne Marie Ryan (Researcher)  
AQA

This presentation outlines some recent AQA projects focusing on student perceptions and experiences of assessment. These include:

- exploring the perspectives of AQA's Student Advisory Group on the role of technology in assessment
- collaborating with Doublestruck/Exampro on evaluating an on-screen assessment for English GCSE

- asking students how Covid-19 affected their practical science lessons, as well as their confidence and skills as they transition to HE
- exploring student perspectives on Equality, Diversity and Inclusion (EDI) in the GCSE and A-level history curriculum.

We recognise students as key stakeholders and these projects highlight the importance of student voice for AQA.

## **Session 2 – comparative judgement (Chair: Tom Benton)**

**11:00 – 12:30**

### **1. Comparative judgement for moderation: a feasibility study**

Lucy Chambers (Senior Research Officer) and Carmen Vidal Rodeiro (Senior Research Officer)  
Cambridge University Press & Assessment

This study looked at the practical feasibility of using comparative judgement for the moderation of non-exam assessment. This work follows on from a simulation study where we explored its theoretical feasibility. We ran a small experimental moderation task using authentic portfolios of work, supplemented by observations, interviews and a survey. We investigated aspects such as whether moderators can view/navigate the portfolios sufficiently to enable them to make the comparative judgements, on what basis they make their decisions, whether moderators can be confident making CJ judgements on large samples of candidates' work (e.g. portfolios), and the time taken to moderate.

### **2. Awarding using comparative judgement: do judges attend to construct-irrelevant features?**

Lucy Chambers (Senior Research Officer), co-authored with Euan Cunningham (Research Officer)  
Cambridge University Press & Assessment

This presentation will report on a study where we investigated the comparative judgement decision making of judges in an awarding context. In particular, whether judges were attending to construct-irrelevant features when making their decisions. This was an experimental study where we manipulated four script features (handwriting, appearance, spelling punctuation and grammar and missing vs. incorrect responses) and presented both original and modified versions to judges as part of an awarding exercise. This was supplemented with judge observations where judges were asked to think aloud whilst doing some judging and a survey.

### **3. Robustness of script evidence in comparative judgement awarding activities**

Jo Williamson (Senior Research Officer)  
Cambridge University Press & Assessment

Comparative judgement (CJ) can be used as a source of evidence for awarding in high-stakes assessments; a key example is informing where grade boundaries should be set in order to maintain standards. This research aimed to increase our understanding of the potential risks in this use of CJ.

Specifically, the research investigated how robust the outcomes of CJ-based awarding activities are to mismarking in the script evidence. We used a series of simulations to explore different kinds of mismarking, and the extent of their impact.

#### **4. The classification accuracy and consistency of comparative judgement of writing compared to rubric-based teacher assessment**

Chris Wheadon (Founder and Chief Executive)  
No More Marking

This talk will present the results of a NESTA funded study into the classification accuracy and consistency of a rubric-based approach to the grading of writing with a comparative judgement approach. The writing was gathered from 11-year-olds in low stakes settings in England. We present evidence that the comparative judgement approach has twice the classification accuracy as the rubric-based approach and is perfectly viable in terms of its efficiency. We discuss the limitations of the comparisons and consider what a national system for assessing writing based on a comparative judgement approach could look like.

## **Tuesday 28 September**

### **Session 3 – accessibility and inclusivity (Chair: Irenka Suto)**

**09:15 – 10:30**

#### **1. Equality of access to access arrangements and their impact on students' performance**

Carmen Vidal Rodeiro (Senior Research Officer)  
Cambridge University Press & Assessment

Access arrangements allow students with specific needs to access an assessment and remove unnecessary barriers without affecting the construct(s) the assessment sets out to measure. Evaluating access to access arrangements and how effective they are is important to ensure that the diverse learning needs of students are addressed and that the performance outcomes are a true reflection of students' knowledge and skills.

Using data on approved access arrangements and results from high-stakes assessments taken by students with and without arrangements in 2018/19, this research investigated the equality of access to access arrangements and their impact on students' performance in England.

## **2. Is the inclusivity paradox of technology in education holding back onscreen assessment?**

Hayley White (Assessment Director) and Mark Campbell (Head of Online Assessment Transformation)  
Pearson

Across primary, secondary and higher education, the education community was forced to seek new solutions to teaching, learning and assessment given the challenges presented by COVID-19. With that in mind, how do we seize the opportunity to advance the innovation and implementation in onscreen assessment?

The traditional context of high-stakes assessment in England dictates that in almost every instance examinations are paper-based. We are eager to explore the potential for maximising technology in assessment, whilst being mindful of the variation in provision across the education experience, both within and outside of school. It is, therefore, necessary to be cognisant of practical considerations needed to ensure comparability across national assessments and with paper assessments, as well as a viable solution for students who do not have digital access outside school.

Can we build accessible onscreen assessments that can be adopted successfully by schools? And, in doing so, explore new opportunities that enable a real paradigm shift with much broader benefit for teachers, learners and parents?

## **3. Working definitions of error used within Cambridge Assessment**

Nicky Rushton (Senior Research Officer), co-authored with Sylvia Vitello (Senior Research Officer)  
Cambridge University Press & Assessment

Some problems in assessment materials clearly constitute an error but others, such as missing 'Oxford commas', fall into a grey area. Clarifying how assessment professionals define errors in assessment materials, either consciously or unconsciously, is important because personal conceptualisations can impact upon the rigour and practicalities of construction and checking processes and affect what is recorded in error logs. In this presentation I will discuss the findings from an interview study to investigate our colleagues' working definitions of error. I will also consider the subtle differences between problems in papers that affect whether they are defined as errors or issues.

## **4. From flying a plane to creating exam papers: how the SHELLO model can help us minimise errors in assessment materials.**

Sylvia Vitello (Senior Research Officer), co-authored with Nicky Rushton (Senior Research Officer)  
Cambridge University Press & Assessment

We need to strive to build a comprehensive understanding of how and why errors occur in our exam materials so we can identify effective solutions to minimise the risks. This must involve understanding both error introduction (why errors occur in the first place) and error detection (why errors go undetected). The SHELLO model from the aviation industry distinguishes six categories of factors that contribute to aviation accidents: Software; Hardware; Environment; Liveware; Liveware-Liveware; and Organisation. In this talk, I will discuss how this model can be used as an organising framework to understand the risk factors for error within exam materials.

## **Session 4 – marking and teacher assessment (Chair: Tom Bramley)**

**11:00 – 12:30**

### **1. How good can marking be? Exploring the marking accuracy that can be achieved for an examined GCSE English Language component**

Stephen Holmes (Research Fellow) and Beth Black (former Director of Research and Analysis)  
Ofqual

The marking literature has established that marking consistency is lower for subjects with extended response items. This study tries to determine the extent to which marking in GCSE English language could be improved, using three stages of intervention. First, mark schemes were redesigned on the basis of research literature and expert input. Second, marker recruitment was highly selective. Finally, markers were given extensive, in-depth and interactive face-to-face training. Marking consistency improved by around 5-15% for most of the questions. Therefore, whilst live marking consistency is good, there is some room for improvement, although there may be resource implications.

### **2. Comparing levels-only marking and comparative judgement**

Tom Benton (Principal Research Officer)  
Cambridge University Press & Assessment

In contrast to the long mark scales used in traditional essay marking, the levels-only approach simply requires markers to assign essays to one of a small number of levels. In addition, requirements for markers to annotate scripts or provide summative comments are removed. These changes make marking faster and more intuitive thus allowing multiple marking of each essay. We will reveal the results from an experimental study into applying this approach. Two ways of using comparative judgement rather than marking to score scripts (rank ordering and pairwise comparisons) were also explored. The efficiency of the various approaches will be compared.

### **3. Automarking of short free-text responses in science**

Gareth Wadge (Data Scientist), Tom Sutch (Senior Data Scientist) and Nick Raikes (Assistant Director of Research)  
Cambridge University Press & Assessment

We are trialling automatic marking of free-text, factual answers of up to a few sentences in length, in an online formative test currently marked by teachers. We are using two approaches: pattern matching based on the mark scheme, and machine learning. All items in the test are taken from past IGCSE papers, and we have a large number of marked candidate responses from the exams with which to train and test the machine learning models. In this presentation we present preliminary findings from the trial and discuss potential future applications of this technology.

#### **4. Scotland's alternative certification model 2020/21 as intended and realised**

Beth Black (Director of Policy, Analysis and Standards) and Martyn Ware (Head of Policy and Research)  
SQA

This session will present an overview of key elements of the design of the Alternative Certification Model (ACM) operated in Scotland in 2020/21 and early analysis of feedback from stakeholders on how this design was actually implemented in schools and colleges. It will suggest how any gap between the ACM as intended and implemented can inform future iterations should they be necessary.

#### **5. A quantitative analysis of the factors affecting centre assessed grades**

Tim Stratton (Research Associate) co-authored with Nadir Zanini (Associate Director for Research, Data and Analytics)  
Ofqual

In summer 2020, GCSE and A level exams were cancelled in England. Instead, schools allocated Centre Assessment Grades (CAGs) based on teacher judgement. We utilised data on CAGs in 2020 and exam results from the previous two years, to identify if there were any changes in the patterns of relationships between grades awarded by teacher judgement in 2020 and other candidate, school and subject level features, when compared to those relationships in a 'normal year'. We discuss key findings from the analysis, as well as the generalisability of these results to other instances of teacher judgement.

## **Session 5 – The future, part 2 (Chair: Jackie Greateorex)**

**13:30 – 15:00**

### **1. More like Germany's? System and ideological tensions in the UK government attempt to make vocational education and training (VET) in England more like the German model**

Tony Leech (Research Assistant, Research and Technical Standards)  
OCR

Recent UK government policy around English post-16 vocational qualifications aims to reshape the system to more closely mirror the “world-class” German model, including by creating a clearer distinction between academic and technical routes. This presentation draws on a review of academic and grey literature analysing the recent history of English and German vocational policy. It explores how the success of the German model can be partly explained by the social and economic architectures underpinning it, which are not replicated in England, and draws out tensions between rhetoric and reality in the UK government interest in this model.

### **2. Next generation Higher National Qualifications in Scotland: continuity and change**

Eleanor Boyd (Policy and Research Manager) and Emma McLaughlin (New Product Development Manager)  
SQA

SQA is engaged in a major programme of work to fundamentally reshape our flagship Higher National Qualifications. Higher National are vocational qualifications specifically designed to meet the needs of employers. This session will outline key aspects of our work with stakeholders in Scotland which underpins and continues to shape this development activity. It will focus on the extensive research undertaken over the last three years to inform the design of and approach to assessment for the Next Generation of Higher National qualifications and how this research is being reflected in the design and development process now underway. It will also describe our approach to the evaluation activity that is threaded throughout the project to support the next phases of development.

### **3. Challenges for 2022 in awarding grades in Northern Ireland**

John Truman (Business Manager, Standards, Data Science and Assessment Development) and Pamela Larmour (Programme Manager, Data Science, Research and Statistics)  
CCEA

A return to public examinations is anticipated for 2022, however, measures must be in place to accommodate the considerable issues faced by students in recent times. With consideration of young people's mental health and well-being, and the significant disruption to their learning during the past two academic years, CCEA has planned to reduce assessment burden in 2022 by omitting some examined units in GCSE and GCE qualifications. As most CCEA GCSE and GCEs are modular, a unit of study will be omitted from the vast majority. To ensure comparability when awarding, CCEA will select the unit to be omitted. This presentation will outline the political and regulatory context for this decision and discuss the possible methods that will be used to ensure fairness for all students.



#### **4. The potential impact of unconditional university offers on A level attainment in England: evidence to inform the debate on proposed changes to university admissions**

Rachel Taylor (Associate Director for Research, Standards and Technical Issues)  
and Nadir Zanini (Associate Director for Research, Data and Analytics)  
Ofqual

This research considers the potential impact of unconditional University offers on students' A level attainment. Drawing on data from 2017 and 2018, we used regression techniques to explore the relationship between attainment at A level and receiving an unconditional offer (and, crucially, which type of unconditional offer) once other factors are controlled for. To provide further context to our findings, we also considered the factors influencing whether students receive an unconditional offer and whether an offer is unconditional. The findings are discussed within the context of the current debate on proposed changes to the University admissions system in England.

#### **5. Using educational research evidence in an agile product development**

Sarah Hughes (Research and Thought Leadership Lead)  
Cambridge University Press & Assessment

I provide research evidence to colleagues who design digital assessments using agile working methods. On starting this role, I asked myself: Where does academic research fit into this? How do I make sure the right research questions are being addressed? How do I ensure that research outcomes reach designers in time to influence their decisions?; How can this be done without jeopardising the quality of the research? I describe the streamlined process for carrying out literature reviews and share examples of research questions and outcomes and the related assessment design decisions