



CAMBRIDGE ASSESSMENT

CAMBRIDGE ASSESSMENT RESEARCH
BERA Annual Conference | 2009



Assessment is the key to identifying and measuring achievement – it shows when education works. Effective assessment leads to better opportunities in life.

Enriching Education

Cambridge Assessment provides fair, valid and reliable assessments that encourage personal development by recognising achievement.

We play a leading role in researching, developing and delivering educational assessment to eight million learners in over 150 countries every year through our three major exam boards: Cambridge ESOL, CIE and OCR. Cambridge Assessment is a department of the University of Cambridge and a not-for-profit organisation.

Established in 1858, we aim to promote educational excellence and high quality learning through the use of assessment. Although there have been many changes to the education system over the years, the sense of mission that sparked the creation of the University of Cambridge Local Examinations Syndicate (the original name of Cambridge Assessment) remains at the heart of everything we do today. We strive for continuous improvement of assessment systems and methodologies around the world to guarantee learners everywhere access to the benefits of their education.

www.cambridgeassessment.org.uk

Cambridge Assessment Research at the BERA Annual Conference 2009

This booklet features information about the Cambridge Assessment research that is being presented at this year's BERA conference.

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 50 researchers who pioneer the latest techniques and evaluate current assessments.

Externally funded research is also undertaken, including for the regulators in the United Kingdom and for many education ministries. The results of our research are widely published in well-respected major refereed journals such as *Review of Educational Research* and *Assessment in Education*, as well as being presented at seminars and conferences. We also have our own publications, *Research Matters* and *Research Notes*.



Publications available at www.cambridgeassessment.org.uk



Sylvia Green, Director – Research Division

Cambridge Assessment has supported the BERA Annual Conference for many years. Like BERA, we believe that educational research plays a vital role in the continuous improvement of education and assessment policies and practices. Members of our Research Division contribute to the comprehensive conference programme by presenting papers that cover a wide range of assessment issues. We look forward to seeing you either at one of our presentations or at our exhibition stand.

Presentations

Objective questions in GCSE science: Exploring question difficulty, item functioning and the effect of reading difficulties

Victoria Crisp

Time: Thursday 3 September from 2:30pm to 4:00pm

Session: Main Conference Parallel Session 3

Reference: 0159



Victoria Crisp, presenter

Breaking from a strong tradition of constructed response examinations, one revised science qualification in the UK now involves some examinations with only objective questions. It was considered of interest to investigate characteristics of these newer GCSE papers such as difficulty and their contribution to validity. This study also explored the potential to use access arrangements data to investigate how students with certain needs may be affected differently to other students by features of exam questions.

Item level performance data for the entire candidature of two GCSE science examination papers were obtained. Traditional statistical measures of difficulty (facility values) and discrimination (correlations of item score with total mark) were calculated for each item. Rasch analysis was also conducted to provide estimates of difficulty independent from student ability and information on item functioning. For one of the papers, a 'Reader' group of students was identified including all students who had access to a reader in their exam. A 'Norm' group of the same size was selected randomly from students without a reader. Measures of difficulty and functioning were compared between groups. For a number of interesting items a sample of student responses was analysed.

A number of factors potentially making questions easier (e.g. absence of technical terms) or more difficult (e.g. incorrect response option makes an accurate statement) were identified. Factors potentially contributing to problems with item functioning were also identified (e.g. objective questions that facilitate guessing). The analyses also suggested a number of question features that may have particularly influenced those requiring reading support (e.g. better performances on questions with little technical language), some of which are unsurprising. The findings have implications for question writing practice.

Standard-maintaining by expert judgement: using the rank-ordering method for determining the pass mark on multiple-choice tests

Milja Curcin, Beth Black and Tom Bramley

Time: Thursday 3 September from 2:30pm to 4:00pm

Session: Main Conference Parallel Session 3

Reference: 0176



Milja Curcin, presenter

The Angoff method for determining pass marks on multiple-choice tests is widely used in North America, Australia, and in the UK. It involves experts judging the difficulty of the test items for 'minimally competent' candidates.

However, as a standard *setting* method, the Angoff method has no explicit mechanism for standard *maintaining*, i.e. keeping the pass mark at the same standard session on session. Therefore, there is a need to explore judgemental methods of standard-maintaining for multiple-choice tests in situations where the requirements for statistical equating and linking are not met.

This study involved piloting an adapted rank-ordering method, which allowed direct comparison of items from a previous session with those from the current live session of a test. Each judge was given several packs of four items (two from each session). Their task was to place the four items in rank order of perceived difficulty.

By fitting a Rasch model which estimates relative difficulty for each item based on the judges' rank orders, we obtained a common scale of 'perceived difficulty' on which to compare the two tests. Knowing the pass mark for the previous test, we could map it to the pass mark on the live test which would be achieved by a candidate of inferred equivalent ability. This would allow standards to be maintained session on session. This exercise was carried out twice in two different OCR vocational qualifications in order to investigate its consistency across contexts and over time.

We will discuss the validity of this method and compare it with the Angoff procedure. We will also discuss its potential as a standard maintaining technique in different examination contexts.

Aspects of AS and A-level Physics uptake

Tim Gill, Carmen Vidal Rodeiro and John F. Bell

Time: Wednesday 2 September from 3:00pm to 4:30pm

Session: Main Conference Parallel Session 1

Reference: 0350



Tim Gill, presenter

Concern is often expressed about the declining uptake of A-level Physics in England. However, such concerns do not always take account of important information. In particular some analyses have been based on A-level entries when the actual supply of A-level Physicists is based on passes.

This paper will draw together the findings from previous work undertaken by Cambridge Assessment into the uptake of different subjects at A-level and report on new work that addresses some issues arising.

We review the trends in Physics A-level uptake over the last twenty years in relation to other A-level subjects (for example, the Physics entry in terms of the size of the overall A-level entry) and consider the impact of broadening the sixth form curriculum.

We also consider the uptake of A-level by school type, ethnicity and social factors. We will address the claim that the independent sector is particularly successful at encouraging students to take A-level Physics. The apparent decline of entries in the state sector compared to the independent sector is placed in the context of the ability of the candidates taking the A-level.

Patterns of entry of GCSE science subjects and how they relate to A-level uptake will also be described. In particular, it will be demonstrated how the compensatory nature of Double Award Sciences has led to misleading views of its efficacy as preparation for Physics A-level.

The paper will also review the reasons why students choose Physics at A-level, including secondary analyses from a large scale survey conducted by Cambridge Assessment into the reasons for A-level choice.

How are archive scripts used in judgements about maintaining grading standards?

Jackie Greatorex

Time: Thursday 3 September from 2:30pm to 4:00pm
Session: Main Conference Parallel Session 3
Reference: 0182



Jackie Greatorex, presenter

Background

Generally GCE and GCSE Awarding Bodies use:

- ♦ *Awarding procedures* to determine grade boundaries (including *archive scripts* to remind examiners of the previous year's standard).
- ♦ *Comparability studies* to monitor standards over time or between Awarding Bodies.

Some authors suggested replacing aspects of marking and/or Awarding with *Thurstone pairs* and/or *rank ordering*. Both involve judging the quality of scripts and are used in some comparability studies. These ideas are still being explored, refined and debated.

At the International Association for Educational Assessment conference in 2008 Greatorex et al presented some findings from a wider project. The project data constituted over twenty verbal protocols of examiners judging script quality in experimental conditions which replicated Thurstone pairs, rank ordering, and part of the Awarding procedure. Greatorex et al reported that the questions that statistically discriminated between grade A and grade B performance were not necessarily the questions examiners attended to most in the *live scripts*. My BERA paper also draws from the wider project and focuses on *archive scripts*.

BERA paper

There are two aims (1) to compare between conditions in terms of the questions receiving most attention in archive scripts and (2) to identify how well these questions discriminated between the performance of candidates who actually received grades A and B. Data analysis is still ongoing. Interim results indicate that two questions statistically discriminated between question level marks of candidates who were awarded grades A and B and these two questions were not always the most referenced questions. Discussion will focus on how the findings relate to practice or potential practices.

An investigation into marker reliability and other qualitative aspects of on-screen essay marking

Martin Johnson

Time: Thursday 3 September from 9:00am to 10:30am
Session: Main Conference Parallel Session 2
Reference: 0205



Martin Johnson and Hannah Shiel, presenters

Literature suggests that readers' comprehension of texts might be weaker when extended texts are read on screen rather than on paper. This has important implications for assessment, implying a need to explore whether the mode in which an essay is accessed might influence assessors' judgements about the quality of the essay.

This project investigated whether examiners could mark digital images of a set of GCSE English Literature essays as reliably on screen as they could in the traditional paper mode, whilst also employing a variety of methods to capture some of the complex reading behaviours that pertain to the assessment of extended texts.

To investigate essay marking reliability, examiners' marks were statistically compared across both modes and with an independent reference mark for each essay. To consider whether mode affected the script features (or constructs) being attended to by the examiners, Kelly's Repertory Grid technique was used to elicit constructs and ratings from two senior examiners. These were then used to build a profile of each script, while marking reliability analyses were used to infer any potential relationship between construct recognition and mode.

Examiners' cognitive load whilst marking was measured by a Task Load Index which enabled a comparison of each marker's cognitive workload in each mode. This was complemented by a measure comparing examiners' spatial encoding abilities across modes. Finally, examiners' navigation flow and annotation practices were observed, coded and compared across a sample of scripts marked in both modes. These observations were then used to inform a series of semi-structured interviews with each examiner.

What was this student doing?: Evidencing validity in A-level assessments

Stuart Shaw and Victoria Crisp

Time: Thursday 3 September from 4:30pm to 6:00pm
Session: Main Conference Parallel Session 4
Reference: 0160



Stuart Shaw and Victoria Crisp, *presenters*

Validity is about the extent to which the inferences made from an assessment's outcomes are appropriate. A claim of validity is generally agreed to require evidence of a number of factors. Whilst a number of possible frameworks for evaluating validity have been proposed, there have been few attempts to apply such frameworks in the UK.

As part of the piloting of a multi-faceted methodology for providing comprehensive validity evidence, this paper reports some of the evidence garnered to address one of the validation questions within the framework used: *'Do performances on exam tasks reflect relevant qualities/intended thought processes?'*

Eleven questions from the examinations of an international A-level geography qualification were selected. For each exam question, six geography experts were presented with the question and its mark scheme and asked to identify the processes that they would expect students to use to answer each sub-question well. The experts were then shown responses to the question from three students (one strong, one average and one weak response) and were asked to identify the processes that they thought the students had actually used to arrive at these answers. Finally, the experts were asked to reflect on the match between the expected and apparent processes.

The experts' views on the anticipated and perceived processes were analysed, looking for commonalities. Additionally, expected and apparent processes were compared, with reference to the experts' reflections.

The paper will report on stronger and weaker matches between expected and apparent processes and what these suggest with respect to this aspect of validity.

Must examiners meet in order to standardise their marking? An experiment with new and experienced examiners of GCE AS Psychology.

Nicholas Raikes, Jane Fidler and Tim Gill

Time: Thursday 3 September from 9:00am to 10:30am
Session: Main Conference Parallel Session 2
Reference: 0672



Nicholas Raikes, *presenter*

When high stakes examinations are marked by a panel of examiners, the examiners must be standardised so that candidates are not advantaged or disadvantaged according to which examiner marks their work.

It is common practice for Awarding Bodies' standardisation processes to include a 'Standardisation' or 'Co-ordination' meeting, where all examiners meet to be briefed by the Principal Examiner and to discuss the application of the mark scheme in relation to specific examples of candidates' work. Research into the effectiveness of standardisation meetings has cast doubt on their usefulness, however, at least for experienced examiners.

In the present study we address the following research questions:

1. What is the effect on marking accuracy of including a face-to-face meeting as part of an examiner standardisation process?
2. How does the effect on marking accuracy of a face-to-face meeting vary with the type of question being marked (short-answer or essay) and the level of experience of the examiners?
3. To what extent do examiners carry forward standardisation on one set of questions to a different but very similar set of questions?

Detailed results and discussion will be included in the paper presented at the conference.

The findings of the study will help stakeholders in public examinations decide whether examiners must meet in order for them to be standardised, and whether this varies according to the experience of the examiners and the type of questions.



CAMBRIDGE ASSESSMENT
NETWORK

CAMBRIDGE ASSESSMENT NETWORK

Continuing development for assessment professionals

The Cambridge Assessment Network enables professionals in assessment to share and develop knowledge and expertise – a centre of excellence in assessment. Our aim is to build an international community committed to high quality assessment that enhances learning.

We offer formal and informal professional development activities for those working in assessment:

- ◆ Seminars, training sessions and other events covering key issues in assessment
- ◆ A certificated course in the subject, accredited by the University of Cambridge
- ◆ Access to formal and informal sources of expertise
- ◆ A wide range of networking opportunities.

Our programme is supported by AssessNet, a virtual learning environment, through which we can deliver bespoke online courses to members of the profession.

www.assessnet.org.uk

Forthcoming events/courses

Certificate in the Principles and Practice of Assessment

This innovative programme is offered by the University of Cambridge Institute of Continuing Education together with Cambridge Assessment. Led by specialists in assessment, the programme provides an introduction to educational assessment, using topical and relevant examples.

Cambridge Assessment Conference – Issues of control and innovation: the role of the state in assessment systems. 19 October, Robinson College, Cambridge

The keynote speakers will be Professor Alison Wolf, King's College London, and Professor Robin Alexander, University of Cambridge. Experts including: Professor Mary James, Faculty of Education, University of Cambridge; Isabel Nesbit, Ofqual; and Dr John Allan, SQA; will lead a series of seminars. For further information please visit www.assessnet.org.uk/annualconference.



CAMBRIDGE ASSESSMENT

Cambridge Assessment
1 Hills Road
Cambridge CB1 2EU
United Kingdom
tel +44 (0) 1223 553311
fax +44 (0) 1223 460278
www.cambridgeassessment.org.uk



UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge. Cambridge Assessment is a not-for-profit organisation.



Cover: Corbis Images