

# Getting the right result

Building quality into the assessment process



Find out more at:  
[cerp.org.uk](http://cerp.org.uk)

Centre for Education Research  
and Practice (CERP)

# How does AQA deliver high-quality assessment?

Each summer around 25,000 examiners mark 7.5 million of our GCSE and A-level examination papers. Ensuring that students receive the right results is of paramount importance, and the process begins months – or even years – before the candidates enter the exam hall. Research evidence shapes every step of the assessment process: AQA has an on-site team of academics based at its Centre for Education Research and Practice (CERP) and their work informs question paper and mark scheme design to build in quality from the earliest stages. All the research papers referenced in this booklet are available to download from [cerp.org.uk](http://cerp.org.uk).

## The basics

The exam specification – previously called the syllabus – is based on criteria provided by the Department for Education and accredited by Ofqual. It is written by our team of subject experts and forms the basis for our question papers and mark schemes. Using our own data, we have shown that the design of mark schemes can impact on how effectively we can identify good

responses (*Effective discrimination in mark schemes*, Anne Pinot de Moira). When there are major changes to GCSEs and A-levels, each qualification needs a new specification and work starts on this over a year before it is accredited. At each stage of development, we share our proposals with teachers and wider stakeholders, and feed their comments back into the development process. Once we have decided on the best option to take forward, we begin to create the detailed content for the specification.

## Fair assessments

We draw on a wide range of research to ensure that our question papers test the full spectrum of candidates' abilities. This enables examiners to establish the relative



## Skills that are hard to assess

Employers and businesses increasingly expect young people to leave school with skills such as working collaboratively in groups and problem solving. These skills, though valued by the education world, are difficult to measure: options including peer and self-assessment raise validity issues.

We have investigated innovative approaches to assessments that are designed to elicit evidence on how well students can solve problems and apply knowledge. For example, we have examined the reliability and validity of approaches such as 'adaptive comparative judgement' to

the assessment of complex performances. We are continuing to explore such approaches in relation to assessing the Extended Project Qualification (EPQ), a qualification valued by many universities as an indication of students' aptitude for the kinds of independent research and analysis skills that are needed for success in higher education. (*Using adaptive comparative judgement to obtain a highly reliable rank order in summative assessment*, Claire Whitehouse and Alastair Pollitt; *Testing the validity of judgements about geography essays using the adaptive comparative judgement method*, Claire Whitehouse.)

difficulty of questions across an individual paper, as well as all the other assessments that make up the qualification. It's important that qualifications allow candidates to perform well in some questions and less well in others, and still achieve a total mark that accurately represents their performance. AQA's research team has been investigating this practice since the late 1980s (*Setting common examination papers that differentiate*, Frances Good) and our findings have informed both the inception and development of GCSEs.

## Valid assessments

We closely monitor how questions are written to be certain that they test what we mean to test. Evidence has shown that questions that use real-life contexts (for example adding up the total cost of items in a shopping basket) must be phrased in a particular way; otherwise there is a risk that we are testing the candidate's understanding of the context, rather than the intended skill. Culturally specific or unusual concepts must be avoided, as should topics that might alienate audiences, for



example questions that reference music could be inaccessible for some deaf students.

### Relevant assessments

In the English examination system, we assess students' abilities in a variety of ways. We ask them to write essays, draw graphs, perform recitals and construct arguments. When we want to examine their understanding of a foreign language we require them to actually speak that language. We expect our artists to apply their technical skills via canvas, screen or otherwise, in tangible settings. We believe that this adds weight to our qualifications.

### The human element

These principles underpin a strong and vigorous examination system – although working within this framework presents some

challenges. We need assessments to be consistent and reliable, and so we create strict marking schemes for subjective tasks and we continuously monitor markers' compliance. However, there must be scope for the examiner to exercise some professional judgement in deciding on the mark that best reflects the quality of the work. Some of our studies have shown that markers with experience mark more reliably, but for some types of marking, specific subject expertise is not always a requirement (*The effect of marker background and training on the quality of marking in GCSE English*, Michelle Meadows and Lucy Billington).

### On your marks...

The senior examiners who design our papers and marking guidance are highly experienced subject

experts. We recognise the need to build capacity in the system, so that we can rely on a good supply of high-quality examiners. The reduction in the use of coursework, the move from modular exams to assessment at the end of a two-year course and the focus on academic subjects has led to an increase in the volume of written exams, and therefore the need for qualified examiners.

### Marking to the right standard

We ensure that we have the right number of examiners and that everyone is trained to mark answers accurately and consistently. This is called standardisation. Our aim is to ensure that professional judgement is applied on the basis of a common understanding of 'what good looks like'. Modern technology and on-screen marking mean we can monitor marking quality in real time, and take

action to fix any problems quickly. If there are any concerns about the way an examiner is marking, we'll reassign their work – including what they have already marked – to another examiner. We have very strict quality control procedures for marking, and a dedicated unit of experts who analyse live marking data and identify any concerns for immediate investigation.

### Awarding grades

When all the scripts have been marked, the grade boundaries are set. This process is known as 'awarding'. Exam boards aim to create papers to the same level of difficulty every time, but in practice there is some variation from year to year. As it would be unfair for students to get a lower grade just because they sat a more difficult paper, grade boundaries are set for each individual exam. This is an enormous undertaking and ensures

### Attracting and developing examiners

AQA has joined forces with the other major exam boards and representatives of schools and colleges to discuss how to build long-term examiner capacity. We understand the need for examining to be an attractive and valuable activity for teachers and their schools. We believe that attracting more teachers to become examiners

will have the added benefit of creating greater understanding of best practice in assessment among teachers.

The working group will identify barriers to examiner recruitment and retention, and help institutions to support and encourage teachers to join and remain in the examining profession.

## Marking on-screen

We use e-marking, which means exam papers are split into single questions or sections and sent electronically to examiners, who then mark the answers on-screen. Our research shows e-marking brings a number of benefits, for example:

- splitting an exam paper into different sections means marking can be distributed to examiners with specific skills and expertise
- the system monitors how well an examiner is marking and provides real-time feedback so any issues can be addressed quickly
- e-marking removes the need to store huge numbers of exam papers

*(Why item mark? The advantages and disadvantages of e-marking, Anne Pinot de Moira)*

that standards are maintained from one year to the next. It's applied to controlled assessments, coursework and written exams. Our statisticians and awarding experts work together with the examiners to set boundaries,

which are then approved by Ofqual. The minimum marks for each grade are confirmed and applied to the marks each student achieved to produce their final grade.

## Reviews and re-marks

Once results have been issued, we provide post-results services. Schools can request copies of marked exam papers and if necessary a senior examiner can review the original marking. We receive requests for re-marks for just 2.5% of the scripts we mark – and the average change is 1.67 marks. However, we recognise that these statistics are irrelevant for an individual student whose mark was not correct. That is why we are constantly working to improve the methods and metrics we use to identify any potential issues with marking at the earliest possible sign of a problem.

## Looking to the future

We continually look for ways to improve our assessment – from question paper or mark scheme design, IT systems and processes, to simply identifying an issue with an individual marker very early on. After each exam series we produce reports that show how well question papers have worked, based on data generated during the marking process. The data from these reports

is combined with feedback from the qualifications developer, customers and senior examiners, and informs our continuous improvement processes.

We design specific research projects, carried out by the CERP team, so that we can be confident that any changes we make are evidence-based and appropriate. This knowledge is shared with the Department for Education and Ofqual, and often informs policy direction. For example, our analysis and advice on marking reliability and the accuracy of classifying candidates in terms of grades was used to shape the design of the new national grading structures (*Setting the grade standards in the first year of the new GCSEs*, William Pointer).

### Find out more

For further information on exam standards, marking quality and assessment design, visit [cerp.org.uk](http://cerp.org.uk)

## At a glance



7.5 million

The number of exam scripts marked each year



25,000

The number of academics, teachers, lecturers and subject experts who help set and mark our exams



18

The number of months it takes to create an exam paper



12

The number of weeks in the summer marking period

