



Qualifications and
Curriculum Authority

QCA's review of standards

Description of the programme

March 2006

Ref: QCA/06/2374

Qualifications and Curriculum Authority. (2006). *QCA's review of standards: Description of the programme*. London: Qualifications and Curriculum Authority.

Contents

1. Aims	3
2. Methodology	5
Stage 1: examination demand	5
Stage 2: standards of performance	7
3. Limitations of the study	11
Changes in syllabus and examination content	11
Individual opinion	11
Limitations in the evidence	11
Difficulty of the judgements	12
Appendix A: Factual analysis of syllabuses, question papers and mark schemes	14
Appendix B: Differences in demand between qualifications (syllabuses, question papers and mark schemes)	20
Appendix C: 1999–2004 level/subject script review	29

1. Aims

The overall aim of the programme is to investigate standards in A level and GCSE examinations in order to determine any action needed to safeguard standards and to inform future development in the examinations.

The typical focus of a standards review is the maintenance of examination standards over time; that is, whether the level of performance required for the award of a grade in one year is the same as the level of performance required for the award of the same grade five or 10 years later. Although this sounds like a fairly straightforward question, it can actually be a very thorny one. This is because the programmes that students study (framed by examination specifications) naturally evolve over time. Unfortunately, the more a specification changes from one point in time to the next, the less definitive any investigation into standards can be. In fact, it may even get to a point where it becomes impossible to say anything definitive about the maintenance of examination standards over time. This is because – even within the same subject area – the kind of content to be studied and the kind of skills to be developed can change quite radically. When the kind of knowledge, skill and understanding to be acquired by a GCSE or A level student changes radically over time, it can become meaningless to compare the level of knowledge, skill and understanding required for the award of each grade.

Investigations into the maintenance of examination standards over time can give useful insights. However, it is rarely possible to draw definitive conclusions from their results. For this reason, only conclusions that appear to be particularly well grounded are reported. And even these should be interpreted with some caution.

Each review aims to find out if:

- the demand of syllabuses and their assessment instruments (for example question papers, mark schemes) has changed over time
- the level of performance required of candidates at key grade boundaries has changed over time.

The programme was organised to run in five-year cycles and to cover every major subject within the first cycle of reviews.¹ Reviews in the first cycle investigated standards in a subject in four

¹ For several reasons the programme has undergone various changes. In particular, a report in 2002 by an international panel of experts, *Maintaining GCE A level standards* (QCA, 2002), recommended that there should be more frequent reviews of subjects
© Qualifications and Curriculum Authority 2006

sample years over a 20-year period. This report is part of the second cycle of reviews. These follow on from work conducted in the first cycle and investigate standards since the date of the last review: they therefore normally cover the past five years.

such as mathematics and English. In addition, in a number of cases where it was felt to be sensible, both the GCSE and the A level in a given subject have been reviewed at the same time.

2. Methodology

On completion of the first cycle of reviews, the programme and methodology used for the studies were reviewed and evaluated. It was found that the first cycle of reviews had yielded significant and helpful information about syllabuses and their assessment instruments and the extent to which these reflect educational objectives. However, the information about standards of candidate performance was less easily evaluated. Consequently, a slightly revised methodology for investigating standards of performance was introduced for the second cycle of reviews.

The prime sources of evidence in the reviews remain essentially qualitative. Reviews in both cycles rely heavily on the judgements of subject specialists with a good understanding of the examination system.

Reviews are organised in two stages:

- Stage 1 – investigating and comparing examination demand
- Stage 2 – investigating and comparing standards of performance.

Stage 1: examination demand

Aim

The aim of the first stage is to compare the demand of current syllabuses and their assessment instruments across all awarding bodies with those of the reference year, focusing particularly on issues raised in the earlier review.

Evidence base

The awarding bodies supply all relevant examination materials from their syllabus with the largest candidate entry from the current year. This includes the specification, the question papers, related mark schemes and examiners' reports. It also includes any subject-specific materials routinely issued to candidates and centres, such as vocabulary lists and formula sheets.

Personnel

QCA appoints an independent consultant to lead each review. This person is a subject expert with considerable knowledge of the examination system. In addition they appoint three other

consultants to carry out the review.² These too are subject experts who should understand the examination system. All appointments are made by a mixture of public advertisement and drawing on a list of experienced QCA consultants. The lead consultant is normally an existing QCA consultant. A member of QCA staff manages the work.

The process

Because there are several reviews being conducted in any one year, the basic process for each review is the same and the documentation broadly similar. However, the work begins with the lead consultant and then the other consultants being given the opportunity to customise the documents as necessary. For example, different subjects use different question types and it is important that the review focuses on those specific to the subject. This process helps to ensure that all consultants have a shared understanding of the documentation and the meaning of the various terms used in it.

The first aim of the review is to make sure that there is a clear, agreed, descriptive evidence base. This is achieved by completing a form for each syllabus in each year, which summarises the nature of the examination. Once the exact structure of this form has been agreed, it is completed by the QCA member of staff and provided to the consultants. Their first task is to check its accuracy and provide any other factual comment that they feel is important to understanding the demand of the examination. A generic version of the form is provided in Appendix A.

Once these factual descriptions are complete, they are a key reference point for the more evaluative processes which follow. These processes involve the consultants independently rating key aspects of the examination from a particular awarding body in the years being reviewed. These aspects include the range of skills being tested, the subject content being tested and all those factors that the consultants have agreed are significant in the demand an examination makes on the candidates. The ratings use a scale of 1 (very undemanding) to 5 (very demanding). In addition to giving the ratings, the consultants are required to provide an explanation for them and it is made clear that the explanation is the critical element in the process, not the rating itself. An example of the form used is provided in Appendix B.

As part of the process of completing the forms, reviewers are also asked to rate the question papers against a set of factors that are known to affect the demand of questions irrespective of the subject. These factors are:

- the complexity or number of processes required to answer the question

² Where a review is looking at both GCSE and A level, this number is increased to seven.

- the extent to which the candidate has to generate the answer from their own knowledge or the extent to which the resources are provided
- the level of abstract thinking involved in the question
- the extent to which the candidate has to devise a strategy for answering the questions.

These factors sometimes need interpreting for a given subject and the consultants discuss them to agree exactly how to apply them. For example, if a question in a language examination provides a passage upon which the questions are based, it is possible to argue that all the resources are provided; however, if the passage is considered to be much too difficult for the candidates, then this is not really true.

Once consultants have completed this evaluative process for each awarding body, they transfer their judgements onto a final form, which allows them to see all their judgements in a single place. Again it is stressed that the ratings are not that important (they are not after all different from those already made); what matters is the explanatory comment that accompanies them, for example to explain why the ratings for a particular awarding body are higher than those for another.

Each consultant sends the completed forms to the lead consultant, who collates the information and identifies key issues arising. This includes points where there seems to be disagreement as well as those where there is consensus. The consultants all then come together and explore the various questions, to ensure that there has been no misunderstanding and, as far as possible, that there is consensus about the various points. This meeting only takes place once the consultants have had the chance to see some candidates' work, so that any questions arising, for example from lack of clarity about marking criteria, can be explored.

The lead reviewer then provides a summary of the main findings against the various factors used in the review.

Stage 2: standards of performance

Aim

The aim of the second stage is to find out if the level of performance required of candidates at the key grade boundaries has changed over the period of study and whether it is comparable across awarding bodies in the current year of the examination. The review focuses on the performance of candidates at grades A and E at A level, and grades A, C and F for GCSE

examinations.³ Where appropriate at GCSE, candidate performance at grade C from different tiers within an awarding body is compared.

Evidence base

The awarding bodies provide 15 examples of candidates' work at the defined boundaries for the syllabus under review. For each candidate, the work to be provided is the complete examination work of the candidate in the subject, including all examination papers, coursework and any oral examinations. This is difficult for awarding bodies to provide, since it involves identifying suitable candidates (as well as being close to the subject grade boundary, it is important that their performance on separate parts of the examination is reasonably balanced) and then obtaining all their work that is held at the school or college as well as their examination papers.

The fact that consultants need to judge a whole candidate's work also makes theirs a very demanding task. However, there are two key reasons why the exercise has to be conducted at the level of the whole subject. The first is philosophical: it is the whole subject result which is the real currency in a qualification and thus this has to be the point of comparison, especially in terms of standards over time. The second is more practical. The structures of the different awarding bodies' qualifications often differ considerably, especially over time. It would therefore not be at all clear which separate elements it would be sensible to compare; nor would it be possible to see how to aggregate the outcomes of such comparisons. But the argument must be that the outcome at the whole subject level should be comparable, containing all compulsory elements and overall meeting the subject criteria.

Personnel

The nature and intensity of the task is such that a significantly larger team of consultants is required than for stage 1. The syllabus reviewers are all involved, but in addition there are nominees from the various awarding bodies and relevant subject associations as well as other independent consultants as necessary. Typically a review of a single subject at a single level (ie GCSE or A level) will involve about 12 consultants.

The inclusion of awarding body nominees is helpful at this stage because they can often throw light on some of the questions that the review of the examinations has already posed. They are also used to the kind of intense review and judgement of candidates' work that this part of the

³ A review of a subject at A level also considers the standards required at AS for the most recent year, but there is no attempt made to compare the current AS with AS examinations that existed prior to 2000. This is because they were explicitly of a different standard.

exercise involves.⁴ However, the exercise is designed to try to avoid any bias, and outcomes are checked to ensure that this has been successful.

The process

The review takes place at a residential meeting held over two-and-a-half days, involving a dozen separate sessions. This is more about the efficient use of time than for any other reason, since reviewers are expected to work independently on their allocated set of tasks.

In each session, each consultant is given two batches of scripts for a particular grade. These batches are normally from either the same awarding body in different years or different awarding bodies in the current year.⁵ They are asked to work their way through each batch, comparing pairs of candidates' work, one from each batch, and deciding which one represented a better performance at the grade boundary. Thus comparisons are either over time within an awarding body or across awarding bodies within the same year. In a given session, each consultant will typically make around 10 separate comparisons. A sampling framework ensures adequate coverage of the range of material available.

They record each comparison made, together with their decision, on a form provided for the purpose. A copy of the relevant form is provided in Appendix C. They are also invited to provide a comment on the factor or factors that have influenced that decision. This part of the form is optional, and reviewers are encouraged to base any comments they make either on the assessment objectives for the subject or any aspects of the subject which they feel the assessment objectives do not properly recognise.

At various stages of the process, the consultants are asked to summarise the quality of work they have seen at a particular grade boundary by drafting a description of the performances at that boundary. For A level reviews, the published performance descriptions form the basis of the exercise. However, it is recognised that those performance descriptions were drafted without access to the kind of whole-candidate performance available to this review. Any differences between the performance descriptions are not taken to mean that awarding bodies had applied the wrong standards when grading, but may rather indicate a necessary adjustment to the published descriptions.

The meeting ends with a plenary session, where there is an opportunity to discuss any preliminary trends in the pattern of judgements and to finalise the performance descriptions.

⁴ They also have value in communicating the outcomes to their awarding bodies, in that they can explain the process from a participant's point of view.

⁵ For some GCSE reviews, they could also be for the same awarding body at grade C across different tiers.

The judgements are then carefully analysed and the outcomes tested for statistical significance. At the same time, the qualitative comments on the forms are scrutinised to identify areas of consistency to support the statistical findings.

For the consultants involved in the syllabus review, the exercise also provides useful evidence to help them finalise their conclusions from their scrutiny of the examination papers.

3. Limitations of the study

Comparing examination standards is a complex task, heavily dependent on the evidence available and the ability of reviewers to make judgements on it. This is particularly true when the judgements are being made about standards over time. Several limitations need to be kept in mind when considering any findings and conclusions.

Changes in syllabus and examination content

Syllabuses and examination papers may have changed over the period of the review. Where these changes are fundamental, reviewers find it difficult to make valid judgements about relative standards because they are not comparing like with like. In particular, where the way a subject is conceived by the subject community changes significantly, all that can usefully be done is to identify the change and check that the current interpretation is still fit for purpose.

Individual opinion

Each individual attaches different value to each part of a subject. The use of structured forms and the discussions that take place in finalising these forms both help to improve consistency of judgement. But it is inevitable that reviewers will still apply their own values: indeed, the use of several consultants both recognises this and seeks to exploit it. However, there are likely to remain differences of opinion about the same syllabus or piece of candidate's work. What is important to remember is that this only serves to illustrate that there are legitimate variations in how a subject is defined, even among specialists.

Limitations in the evidence

Consultants have a wide range of the materials needed to analyse standards of performance. However, it is rarely perfectly suited to the task. In particular, it is often possible to supplement the materials for the current syllabuses in response to specific questions about the nature of, for example, the support materials available to teachers. However, this is seldom true for the earlier examinations. In general, judgements are based on only what is common to all syllabuses in the review. However, current examinations normally provide much more by way of support than used to be the case and this aspect is not taken into account in these reviews. It should be noted that while such support does much to improve teaching and learning for an examination, it does not inherently change the difficulty of the tasks involved.

The materials available for the script review are also sometimes less than ideal for the purpose. This ranges from coursework not being available for individual candidates, which makes it hard to judge the overall performance, to cases when it has been impossible to supply the whole work of a single candidate; instead the work is a composite candidate, scoring the right marks in the right combination for the grade but without showing how an individual candidate performs. Reviewers regularly comment on how difficult it is to judge such composite candidates.

In addition, there are two further aspects of the sample of work seen that need to be taken into account.

- First, the actual sample of work seen is relatively small, although as large as can be managed. Moreover, it is not always possible for the awarding bodies to provide even 15 candidates that match the specification. In such circumstances, the decision has to be taken whether to loosen the specification or to use fewer samples of work. Either reduces confidence in any outcomes.
- The second arises from the fact that the work used is from candidates whose performance across all the elements of the examination is reasonably balanced.⁶ This is because there is substantial evidence that people find it easier to make the kind of judgements required when this is the case. However, it must be noted that such candidates are far from typical. In fact at A level typical performance across six units seems to involve variation of three or more grades. Critically, consultants may find it easier to make their judgements with balanced performance, although there is no guarantee that these would be the same decisions if candidates did significantly better in some aspects of the examination and worse in others.

Difficulty of the judgements

It is important to remember that although the people making the judgements are subject experts and familiar with the examination system, they are still being called on to make highly complex decisions. For example, in judging the demand of an examination they have to consider breadth and depth and then to decide how far any differences in the two factors balance each other out.

This is even more true of the judgements about performance, where consultants are called on to make relatively quick judgements about the overall standard of work of two candidates answering different examination questions on different papers designed to test different syllabuses.

⁶ Balance is only considered at the level of performance in separate elements in the examination. On any particular paper, a candidate's performance may very well be uneven.

For all these reasons, and in particular because these judgements are so difficult, the reports only identify aspects of syllabuses where there is strong consensus among the consultants that they result in real differences in demand. Similarly, although the outcomes of the review of candidate performance do not report the statistical information, a difference in performance, whether over time or across awarding bodies, is only identified if it is statistically highly significant.

Appendix A: Factual analysis of syllabuses, question papers and mark schemes

Form A

Qualification:	Subject:
Awarding body:	Year:
Reviewer:	

General Information

Overall syllabus page length	
General philosophy/rationale	
Aims	

Specification of prior knowledge	
Additional syllabus support material	
Access to resources (eg 'open book' exam)	
Anticipated teaching time (if stated)	
Background knowledge (if stated)	

Assessment objectives

Summarise assessment objectives, giving weightings where possible.

Assessment objective	Weighting (%)	Assessment method (written paper, coursework, etc)	Comments

Scheme of assessment

Summarise the details of each assessment unit, giving weightings where possible.

For type of assessment the codes are:

E = externally set and marked

I = internally set and marked, and externally moderated

W = written exam

C = coursework

O = other (give details)

Unit Title	Weighting (%)	Duration	Type of assessment	Comments

Content and syllabus structure

4a Topics

Identify the topics within the syllabus. Comment on sub-topics and other issues that will help to evaluate the relative demands made by the content.

Topic	Units	Comments (How assessed, depth of treatment, number of sub-topics etc)

4b Options

Note the number of optional routes within the syllabus:

Note any optional routes within each question paper:

Summarise the pattern of option availability:

4c Further comments

Identify any additional issues (depth, breadth, etc). Assume the minimum coverage allowed by the structure of the examination, rather than that indicated as desirable by the syllabus. NB This may be affected by choice in the units taken /choice in the questions answered within a unit.

Question papers (including coursework)

For question type the codes are:

A = multiple choice/completion

B = short answer

C = structured

D = open-ended

E = essay

Unit Title	Duration	Question type	Comments (eg weighting, content coverage, etc)

Are mark allocations given?

Other general comments (eg layout, rubrics, stimuli, etc):

Mark schemes (including coursework assessment criteria)

Comment on the nature and extent of the mark schemes (eg general marking instructions, marks for use of language, etc).

Appendix B: Differences in demand between qualifications (syllabuses, question papers and mark schemes)

Form B

Qualification:	Subject:
Awarding body:	
Reviewer:	

1 General information

For each element, please consider factual differences identified through Form A, then comment on whether the differences identified lead to differences in demand, and whether differences in some areas are offset in others. Quote examples wherever possible.

General philosophy and syllabus aims

Anticipated teaching time

Access to resources

Specification of prior knowledge

Level of syllabus detail and support material

Please rate each element below, for each year, on the given scale. Add any comment you wish in the space below, especially where there are differences in rating.

2 Syllabus

2.1 Assessment objectives (taking into account any changes in assessment objectives and their respective weightings across time)

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

2.2 Scheme of assessment (taking account of any levels of choice, the balance of breadth and depth)

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

3 Content

3.1 Nature of topics

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

3.2 Range of topics

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

3.3 Number of topics

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

3.4 Balance of compulsory and optional topics (considering all possible routes through the specification)

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

4 Question papers and their associated mark schemes

4.1 CRAS analysis

Please complete a CRAS analysis as instructed in the CRAS analysis briefing note.

In all cases, please also consider the marking schemes/criteria when evaluating the demand.

CRAS ratings should be recorded on the CRAS question paper analysis form.

4.2 Nature of tasks

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

4.3 Time available per question

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

4.4 Layout and presentation of question papers

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

4.5 Coverage of assessment objectives (commenting on any changes in the clarity of the relationship between assessment objectives, question papers and mark schemes)

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

4.6 Accessibility of questions for lower attaining candidates

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

4.7 Appropriateness of questions/tasks to allow grade A candidates to demonstrate their knowledge/skills/understanding

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

4.8 Overall demand of question papers and mark schemes (with reference to CRAS and previous answers in this section):

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

5 Coursework: requirements (eg changes in what is assessed and how, and any changes to weighting and impact on demand)

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

6 Overall demand

Year	Very undemanding		About right		Very demanding
	1	2	3	4	5
1997 A level					
2005 AS					
2005 A level					

Justification for your judgements:

