

Post 16

2001

Five year review of standards

A level mathematics



Guarding standards

Introduction

Every summer, the publication of GCSE and A level examination results prompts public interest in the standards of those examinations.

In 1996, Lord Dearing in his *Review of Qualifications for 16–19 Year Olds* made several recommendations to ensure that ‘there is a basis and accepted procedure ... for monitoring and safeguarding standards over time’. In the same year, SCAA (one of QCA’s predecessors) and the Office for Standards in Education jointly investigated standards in English, mathematics and science (chemistry) in 16+ and 18+ public examinations over time.¹

The outcomes of this work were published in *Standards in Public Examinations 1975 to 1995*. One of the recommendations was that there should be:

‘... a rolling programme of reviews on a five-year cycle to ensure examination demands and grade standards are being maintained in all major subjects. Physics, history, French and German should be included in the programme at an early stage.’

The five-yearly review of standards programme is a response to these recommendations. It is run by QCA in collaboration with the regulatory authorities for Wales and Northern Ireland, ACCAC and CCEA, and is designed to investigate the standards in A level and GCSE examinations. It aims to find out if:

the demand of syllabuses and their assessment instruments has changed over the last 20 years (examination demand);

the level of performance required of candidates at grade boundaries has changed over the last 20 years (grade standard).

Organised to run in five-year cycles, the programme was structured to cover every major subject during its first cycle. Each year, up to 100 independent specialists review around 2,000 exam scripts, drawn from all the awarding bodies, together with their associated syllabuses, question papers and mark schemes.²

¹ 16+ examinations cover GCE O level and Certificate of Secondary Education (up to 1987), and GCSE (from 1988).

² For the purposes of this report, the general term *awarding bodies* is used to cover both the A level examination boards and the GCSE examining groups.

Methodology

Each study was organised in two stages:

- stage one – investigating changes in examination demand;
- stage two – investigating changes in standards of performance.

Each covered two sample years: the year of the study and 1995, the year used for the SCAA/Ofsted study.

Stage one: examination demand

Aim

The aim of this review was to establish whether the demand of syllabuses and their assessment instruments changed over the period of the review.

Evidence base

The awarding bodies were asked to supply, for each subject, copies of one major syllabus from the most recent year. They were also asked to provide the related question papers, mark schemes, examiners' reports, and details of the procedures in operation at the time of each examination. The materials used in the SCAA/Ofsted study were available for comparison.

The process

A coordinator and three reviewers – independent experts from a variety of backgrounds – were appointed for each subject. Each coordinator was given a framework and asked to use it to describe the main differences between the syllabuses from the different years. This description was given to the reviewers, who were asked to study the syllabuses, question papers and mark schemes and independently judge whether the differences between years affected the demand of the examination. After the material had been reviewed, the team for each subject area met and discussed any issues. The coordinator then reported on the findings and identified any conclusions.

Stage two: standards of performance

Aim

The aim of the second stage was to find out if the level of performance required of candidates at grade boundaries has changed over the period of the study. The review focused on the performance of candidates at grades A and E at A level, and grades A, C and, sometimes, F for 16+ examinations.

Evidence base

The awarding bodies were asked to provide 15 examples of candidates' work at the defined boundaries from the most recent year of examination. They were asked to

submit the complete examination work of candidates, including all examination papers, coursework and any oral examinations. The materials used in the SCAA/Ofsted study were available for comparison.

The process

A team of up to 12 reviewers was recruited for each subject. The reviewers came from a variety of backgrounds, including universities, selective and non-selective schools, maintained and independent schools, and further education institutions (including sixth form colleges). Some of them had backgrounds working for the various awarding bodies.

The coordinator from stage one was used again in this stage and the syllabus reviewers normally participated.

The review took place over two days. Before the meeting, each coordinator produced a general description of the standards expected for the grade boundaries in the study. Where these were available, published grade descriptions normally formed the basis of the performance descriptors. The coordinators were asked to take into account the fact that they would be looking at borderline performance rather than that comfortably in grade which is the intention of grade descriptions. The performance descriptors were discussed and agreed by the team at the start of the meeting.

Reviewers were each given a batch of scripts for a particular year, grade and awarding body. Working independently, they were asked to judge if the scripts matched the agreed grade description. They could categorise the work as:

- above the expected standard;
- slightly above the expected standard;
- at the expected standard;
- slightly below the expected standard;
- below the expected standard.

They were then given another batch of scripts of the same grade, either from another awarding body or of a different year from the same awarding body. They categorised these scripts and compared them with the first batch to identify any significant differences between candidates' performance. A sampling framework ensured adequate coverage of the sample. A copy of part of one framework is provided on page 4.

At the end of the two days, a plenary session was held and the reviewers discussed their findings and any significant issues. As with stage one, the coordinator reported on the findings and conclusions.

Limitations of the study

Comparing examination standards over time is a complex task, heavily dependent on the evidence available and the ability of reviewers to make valid judgements on it.

When considering the findings and conclusions, several limitations need to be kept in mind.

Changes in syllabus and examination content

Syllabuses and examination papers changed significantly over the period of the review. For example, in assessing GCSE science examinations, the three tiers of entry of 1995 had been reduced to two. Fundamental changes make it difficult for reviewers to make valid judgements about relative standards because they are not comparing like with like.

Individual opinion

Each individual places different values on each part of a subject. Agreed definitions of standards and frameworks show reviewers the standards they should work to, but it is difficult for them to avoid applying their own values. This can lead to differences in opinion about the same syllabus or piece of candidate's work.

Lack of evidence

While reviewers had syllabuses and examination papers (although not always mark schemes) for all the years in the study, they did not have all the evidence they needed to analyse standards of performance. This applies particularly to examination scripts. What was used in the SCAA/Ofsted study was work for separate components of the examination rather than the whole work of candidates. Coursework and any oral examinations were usually missing.

Table 1: Sampling framework for part of a typical A level study

DAY 1

8:30	BOARD A, GRADE A	BOARD A, GRADE E	BOARD F, GRADE A	BOARD F, GRADE E	BOARD C, GRADE A	BOARD C, GRADE E
10:00	1996 1-7	1996 1-7	1996 1-7	1996 7-1	1996 1-7	1996 15-8
10:10	BOARD A, GRADE A	BOARD A, GRADE E	BOARD F, GRADE E	BOARD F, GRADE A	BOARD C, GRADE A	BOARD C, GRADE E
11:30	1991 1-3	1991 1-3	1996 8-15	1996 7-1	1991 1-7	1991 15-8
11:50	BOARD A, GRADE E	BOARD A, GRADE A	BOARD C, GRADE E	BOARD C, GRADE A	BOARD E, GRADE A	BOARD D, GRADE A
1:05	1996 1-7	1996 15-8	1996 1-7	1996 8-15	1996 1-7	1996 15-8
2:15	BOARD A, GRADE E	BOARD A, GRADE A	BOARD A, GRADE E	BOARD B, GRADE E	BOARD E, GRADE E	BOARD D, GRADE E
3:30	1991 1-3	1991 3-1	1996 15-8	1996 15-8	1996 1-7	1996 15-8
3:30	BOARD B, GRADE A	BOARD D, GRADE E	BOARD B, GRADE A	BOARD D, GRADE E	BOARD D, GRADE A	BOARD E, GRADE A
4:45	1996 1-7	1996 1-7	1996 15-8	1991 4-1	1996 7-1	1996 8-15
5:05	BOARD B, GRADE E	BOARD D, GRADE E	BOARD B, GRADE E	BOARD D, GRADE E	BOARD D, GRADE E	BOARD E, GRADE A
6:20	1996 1-7	1991 1-4	1996 8-15	1986 4-1	1996 8-15	1991 1-3

DAY 2

8:30	BOARD C, GRADE E	BOARD E, GRADE E	BOARD E, GRADE A	EDEC , GRADE A 1996	BOARD F, GRADE A	BOARD A, GRADE E
9:45	1996 7-1	1996 15-8	1996 1-7	7-1	1996 8-15	1996 15-8
9:45	BOARD C, GRADE E	BOARD E, GRADE E	BOARD E, GRADE A	BOARD B, GRADE E	BOARD F, GRADE E	BOARD A, GRADE E
11:00	1991 1-7	1991 3-1	1991 3-1	1996 8-15	1996 8-15	1986 7-1
11:20	BOARD C, GRADE A	BOARD E, GRADE A	BOARD E, GRADE E	BOARD E, GRADE A	BOARD C, GRADE A	BOARD A, GRADE A
12:35	1996 7-1	1996 7-1	1996 8-15	1996 8-15	1996 15-8	1996 1-7
1:45	BOARD C, GRADE A	BOARD E, GRADE A	BOARD E, GRADE E	BOARD E, GRADE A	BOARD C, GRADE A	BOARD A, GRADE A
3:00	1991 7-1	1991 1-3	1991 1-3	1991 3-1	1991 15-8	1991 3-1

A level mathematics: review of standards 1995–8

Introduction

SCAA, QCA's predecessor body, together with Ofsted, conducted an enquiry into examination standards. The results of the work, published in 1996 as *Standards in Public Examinations, 1975–1995* (SCAA, 1996), included a series of recommendations concerning future examinations in each subject reviewed.

The subjects included A level mathematics where changes were already in train at the time of the report. The most significant of these were:

- a revision to the common subject core for first examination in 1996;
- a continuing move from linear to modular examinations.

Syllabuses in 1999 were therefore those approved under the revised subject core, and had been approved prior to the 1996 work. Implementation of the 1996 report's recommendations has been effected for the specifications accredited for Curriculum 2000. This review offers, however, a chance to evaluate whether the last set of changes anticipated some of the recommendations or made them more urgent.

Examination demands

Materials available

Reviewers considered the syllabus documents, the question papers and associated mark schemes, and the examiners' reports for syllabuses from each of the awarding bodies in 1995 and 1998.

For AQA/A, Edexcel, CCEA and WJEC, the year-on-year comparisons were relatively straightforward; the comparisons involving OCR and, in particular, AQA/N were less easy to make. For each of these awarding bodies in 1995 a linear syllabus had been used; a modular one was chosen for 1998. Any differences identified in these cases could be as much due to the linear/modular contrast as to any genuine change over time.

About 61,000 candidates took A level mathematics in 1998. Just over 50 per cent of those entered for the syllabuses used for that year in this study.

Sources of demand in mathematics

For mathematics, the demand of any particular examination was seen to be defined almost entirely by the form and extent of the assessment scheme, the details of the syllabus content, the questions actually set in the question papers (and tasks carried

out in coursework, where relevant), and the grade boundaries set.² Of remaining aspects of syllabuses, some such as the general philosophy and aims were judged to have no direct bearing on demand. Others such as assessment objectives may affect what topics need to be included in the syllabus and how they are to be examined, but the level of demand implied was felt to lie in the details of the implementation rather than in the objectives themselves.

In matters such as the amount of support material that awarding bodies make available, the prior knowledge that students embarking on a course are assumed to have, and to have the use of, calculators and formula books, there was no evidence of differences, either between awarding bodies or between years, that would indicate any noticeable effect on overall demand. However, not all of the relevant formula books were available for reviewers to compare.

Mark schemes represent an obvious possible source of difference of demand. In the event, reviewers judged that, although schemes were not always easy to interpret without seeing some examples of their implementation, the approaches adopted by the different awarding bodies in the two years were very similar, and that little or no significant difference in demand could be attributed to mark schemes, at any rate as far as written papers were concerned.

Schemes of assessment

The features of a scheme of assessment that reviewers saw as most affecting overall demand were the total length of examining time, the presence or absence of question choice, and the degree of modularity. An increase in examining time was seen as increasing the demand on candidates; providing a choice of questions within a paper was seen as reducing demand; and reducing the amount of content per examination unit was seen as reducing demand. Schemes that include a variety of different types of assessment (eg coursework, or a comprehension test, in addition to the normal timed, written papers) were seen as being, in principle, more demanding. However, the inclusion of coursework in a scheme was not considered in practice necessarily to increase demand.

Subject content

The content of the pure mathematics examination components under review was determined largely by the subject core for the relevant year. Reviewers considered that the changes made to the core between 1995 and 1998 had, if anything, reduced demand. Where the core had increased demand in at least some application areas, for example in greater emphasis on modelling and interpretation, reviewers found only marginal evidence for this in the relevant question papers and mark schemes.

Question papers

The individual questions set in examination papers varied considerably in terms of length and difficulty, as was to be expected. Longer questions were almost always quite highly structured, and reviewers detected little or no systematic difference

²These were not known to the reviewers at the time of the syllabus review: any differences in this aspect would be the subject of the script review.

between awarding bodies in this. There was little evidence of a change in the amount of structuring between 1995 and 1998.

Optional routes

The increased use of modular schemes had a major impact on the evaluation of optional routes through assessment schemes. Judgements were, moreover, based on only a partial evaluation of most of the modular syllabuses. Although the syllabus content for all the modules was available, the corresponding question papers and mark schemes were provided only for the most popular. Questions about the equivalence of the demand of different routes through a modular syllabus could not therefore be adequately addressed. Reviewers could not be confident that all allowable combinations of modules made equal demands, since different routes might involve different proportions of coursework; different proportions of 'pure' and 'applied' modules; and different balances between breadth and depth. There was doubt whether Discrete Mathematics modules made mathematical demands comparable to those made in other branches of the subject.

Grade thresholds

The examinations being studied were judged to have been broadly comparable. This meant that final judgement depended on the placing of the grade thresholds, which would only become clear during the next phase of the review. The findings concerning both between-awarding body and between-year comparisons should therefore be understood in this context.

Demand over time within awarding bodies

The 1998 AQA/A examination was judged less demanding than that in 1995. The total examination time for 1998 was less than in 1995; the total syllabus content was judged less demanding; and the Pure Mathematics questions, in particular, were thought to be a little less algebraic and more structured than in 1995. The introduction into the Statistics paper of some short questions, where the 1995 paper had contained long questions only, was also considered to have reduced demand. These were only partly offset by the disappearance of question choice between 1995 and 1998.

Reviewers identified no clear difference in overall demand in the Edexcel examination between the two years.

The AQA/N assessment material on Application was not available (it took the form of tests drawn from a confidential item bank), so reviewers were not able to provide a final judgement. For the Pure half of the assessment, the demand in the two years was judged very similar, with 1998 being, if anything, a little less demanding.

The 1998 CCEA examination was judged to be a little less demanding than that in 1995. The changes in the subject core, combined with the structural change in this scheme whereby the single 3-hour Paper 1 on Pure Mathematics in 1995 became the two separate 1½-hour modules A1 and A2 in 1998, were thought to have resulted in a less extensive range of Pure Mathematics being tested in a generally more straightforward way. There was also some reduction of Mechanics content. The content of the Statistics section of the syllabus was very comparable in the two

years, but the 1998 papers tended to focus on the less challenging aspects of the material.

The 1998 OCR examination was judged more demanding than that in 1995, by reason of the more extended and varied assessment (inclusion of coursework and a comprehension test) in 1998 and the fact that there had been a choice of questions in 1995. However, some of the 1995 questions (though optional) were judged harder than any of the 1998 ones; 1995 questions were marginally less structured; and the modular structure in 1998 reduced demand within individual examination units by focusing on only a subset of the total content.

Little overall difference was detected in the demands of the 1995 and 1998 WJEC schemes.

Comparability between awarding bodies in 1998

Reviewers judged the overall demand in the AQA/A and CCEA schemes to be less demanding than in the others. For AQA/A, the syllabus content seemed light; the total examination time was short; and the question papers were quite straightforward. In the case of CCEA, some of the shorter questions (in particular) in the papers appeared to place too much weight on the more elementary aspects of the syllabus; the syllabus content itself was not thought to be particularly out of line. Both OCR and Edexcel had demanding aspects, eg the content of the OCR Pure Mathematics modules was judged relatively extensive, and the total amount of examining, including coursework, meant that candidates had to do a great deal of work in total. However, OCR questions were perhaps a little more structured than those in Edexcel papers. The WJEC scheme represented a very reasonable average standard of demand for 1998, while no overall judgement on the AQA/N scheme was possible.

Recommendations from Standards in Public Examinations, 1975 –95

Algebraic manipulation

The revised core in 1996 had not increased algebraic requirements and there was no evidence of an increase in the requirements for algebraic manipulation. Even where there was enough algebra, the standard of manipulation expected was not necessarily demanding enough. Overall, there was no evidence of awarding bodies trying to increase the emphasis on algebra. Scrutiny of candidates' work at the script review corroborated this perception and suggested that actual performance in this area had, if anything, declined.

Balance between structured and unstructured questions

All reviewers considered that there is scope for all awarding bodies to reduce the amount of structuring in the question papers, and that there was no evidence that this was being undertaken. Indeed, the amount of structuring had, if anything, increased. Again, the script review suggested that the level of structuring in 1998 continued to make it difficult for candidates to demonstrate their ability to carry out a multi-stage solution.

Reasoning and problem-solving

There was no evidence of attempts to introduce more reasoning into written examinations. Reviewers felt that the continuing lack of unstructured questions was a

major issue. There were, however, some questions where the required method of solution was not quite obvious, and where some 'problem-solving' skills were therefore required. Some Mechanics questions were identified as quite demanding in this respect, but other parts of the examination made few demands of this type. It is possible that some aspects of reasoning and problem-solving skills might be identified in candidates' coursework.

Summary

The alterations to the Pure Mathematics core were thought to have resulted, on the whole, in reduced demand in 1998 compared to 1995, while increased emphasis on other aspects of mathematics, such as modelling and interpretation, may have resulted in some increase in demand. The demands of the various examination schemes were sufficiently comparable to be affected by the precise placing of grade boundaries.

In general, awarding bodies seemed to have made very little effort to address any of the recommendations in *Standards in Public Examinations, 1975–1995*. This was disappointing since several of these recommendations concerned the nature of question papers rather than content or the structure of the examination.

Performance at grade A and grade E

Materials available

Scripts were available at grade A and grade E for the 1995 and 1998 syllabuses. However, the evidence provided was flawed in some respects. In particular, for the 1998 OCR candidates, no coursework was available for reviewers to inspect. This means that less weight can be attached to conclusions involving this syllabus. Similarly, in the case of AQA/N and CCEA, scripts from one year only were available, and a number of the sets of scripts were scrutinised by one team member only. Reviewers also experienced significant difficulties in applying the performance descriptors. In particular, work at grade E tends to be characterised by the lack of pattern of performance by individual candidates in different aspects of mathematics.

Standards expected at grade A

Candidates usually recognise what is being demanded of them and are generally able to choose appropriate techniques. They are able to reason in a logical way with only occasional errors.

Candidates are able to manipulate mathematical expressions with a high degree of accuracy and make very few algebraic blunders. They are able to produce and interpret diagrams and graphs accurately. They use mathematical language correctly. They use their calculators appropriately and present results to an appropriate degree of accuracy.

Candidates can formulate problems mathematically and select standard models introduced in the syllabus. They can generally move into them from realistic

situations, and usually interpret results fully and accurately. They make relevant comments on the appropriateness of models.

Candidates make substantial progress through problems requiring the development of a multi-stage solution.

(Relatively good performance in one area may compensate for relatively poor performance in another.)

Standards expected at grade E

Candidates often recognise what is being demanded of them and are sometimes able to choose appropriate techniques.³ They are able to reason in a logical way in straightforward and very standard cases but nevertheless often make elementary errors.

Candidates are able to manipulate simple mathematical expressions but often make elementary errors. They are often able to produce and interpret diagrams and graphs reasonably accurately. They sometimes use their calculators appropriately and often present results to an appropriate degree of accuracy.

Candidates have some knowledge of standard models introduced in the syllabus. They can generally move into indicated models from simple situations, and can interpret results partially. They make some relevant comments on the appropriateness of models.

Candidates usually make little progress through problems requiring the development of a multi-stage solution.

(Good performance in one area may compensate for poor performance in another.)

It was noted that the separate paragraphs into which the descriptions were divided should not be expected to carry equal 'weight', and that, perhaps particularly in examinations from 1995, some syllabuses would not necessarily provide opportunities for some of the descriptions to be adequately met. It also proved easier to apply the descriptions to some of the linear scripts than to some of the modular sets.

Findings

For most cases, there was a consensus that the scripts under review matched the performance descriptions reasonably closely. The main exception concerns the 1998 OCR examination, at both grades; in this case, the majority rated the scripts as below the expected standard. For the OCR 1995 examination, at grade A the performance was judged to match expectation or even, just, exceed it. It must be remembered, however, that this is the awarding body which reviewers felt they could not fairly judge in 1998. It was also possible to identify a slight overall difference in the pattern of reviewers' responses to the 1998 and 1995 Edexcel scripts at grade A, although any such difference was much less marked than was the case for OCR. WJEC scripts were judged to be of the same standards in both years at both grades.

^pAs a result of the exercise, it was agreed that this sentence would have been more accurate had 'often' been further qualified with 'fairly'.

In terms of performance at grade A in 1998, Edexcel scripts were judged to be of a higher standard than those from OCR; the picture was less clear for CCEA and WJEC. At grade E, Edexcel and WJEC were judged comparable and of a higher standard than OCR.

Reviewers also noted various ways in which the descriptions had not accurately reflected the nature of the examinations. First, the reference, in the descriptions for both A/B and E/N candidates, to 'problems requiring the development of a multi-stage solution' had not been useful, since in practice there were virtually no such questions in any of the papers. Also, there was very little that related to the third paragraph of the descriptions (mathematical models; interpretation; comment) in the 1995 papers – and very little more in many of the 1998 papers. The OCR syllabus for 1998 was considered to focus on these areas more than most, and this emphasis might have been even more apparent had the missing coursework been available. Much of the reviewers' judgement depended therefore on the first two paragraphs of the descriptions.

The largely content-free nature of the performance descriptions presented a further problem. The descriptions themselves included no explicit reference to the difficulty of the mathematical tasks to which they are supposed to refer, so reviewers had to decide how (or whether) to allow for the examination, or part of it, being at an inappropriate standard. A particular instance appeared in one of the syllabuses under review: the M2 module in the Edexcel 1998 examination was clearly too hard, as was acknowledged in the Examiners' Report. The resulting low marks and poor standard of work made for great difficulty in judging the candidates' standard against the descriptions, and in comparing the standard of this examination against others where candidates were given a fairer opportunity of showing what they could do.

Several further factors made the reviewers' task difficult: linear/modular examination structures; choice/no choice of questions; coursework/no coursework syllabuses; and differing combinations of components. It was also found difficult to relate the 'complete work of a candidate' (or of a pseudo-candidate), as supplied for 1998 examinations, to a sample of scripts all at or near each component borderline, as supplied for 1995.

Summary

Reviewers said that, for several reasons, their judgements were not made with any degree of confidence. There was, however, some cause for concern about standards at both grades in OCR in 1998, which were judged below expectation and below that required in 1995.

Key to the awarding bodies

During the period of the reviews, the number of awarding bodies operating fell. There are currently five: AQA, CCEA, Edexcel, OCR and WJEC. However, the three English awarding bodies came together through a number of mergers and a government requirement for unitary awarding bodies which could offer the range of GCSE, A level and GNVQ/VCE qualifications. This means that the qualifications used in the reviews came from a number of earlier examination boards and examining groups.

For the purposes of the reports the following abbreviations will be used:

AQA/A, AQA/N, CCEA, Edexcel, OCR and WJEC.

AQA/A covers AQA legacy A level syllabuses offered by AEB; legacy GCSE syllabuses offered by SEG; and O level syllabuses offered by AEB.

AQA/N covers AQA legacy A level syllabuses offered by NEAB, NEA and JMB; legacy GCSE syllabuses offered by NEAB and NEA; and O level syllabuses offered by JMB.

CCEA covers A level and GCSE syllabuses offered by CCEA, NISEAC and NISEC; and O level syllabuses offered by NISEC and NIGCEEB.

Edexcel covers A level and GCSE syllabuses offered by Edexcel, ULEAC and ULSEB; GCSE syllabuses offered by Edexcel, ULEAC and LEAG; and O level syllabuses offered by ULSEB.

OCR covers A level syllabuses offered by OCEAC, OCSEB, UCLES and UODLE; GCSE syllabuses offered by MEG; and O level syllabuses offered by OCSEB, UCLES and UODLE.

WJEC has retained the same name throughout the period.

© **Qualifications and Curriculum Authority (QCA) 2001**
83 Piccadilly
London W1J 8QA

www.qca.org.uk/

Order ref: QCA/01/764