

Five year review of standards

A level biology

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 four sample years: the year of the study and its predecessors from five years, 10 years and 20 years earlier.

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 and its predecessors for the other three years in the study. 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.

In general, syllabuses and question papers were available from all awarding bodies for all years in a study. Unfortunately, prior to 1988, few mark schemes and few documented details about awarding procedures had been retained.

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 for each syllabus studied in stage one. They were asked to submit the complete examination work of candidates, including all examination papers, coursework and any oral examinations.

On the whole, the samples provided for the most recent year of each study were complete. However, the coursework was sometimes missing and work from modular syllabuses presented a problem, in that it was seldom possible to provide the entire work of individual candidates. Usually, several modules from one candidate were provided, supplemented by modules from other candidates to produce the appropriate overall result.

Samples of work from earlier years were much less complete. The awarding bodies could rarely provide work from enough candidates or did not have the complete work of candidates – coursework and orals were usually missing and the work consisted of individual components. No work from the earliest year of the reviews was available.

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 5.

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

In some subject areas, syllabuses and examination papers changed radically over the period of the review. For example, in assessing modern foreign languages the relative importance of the skills of reading, writing, speaking and listening has changed considerably. 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. The archiving practices of the awarding bodies vary, each keeping different amounts of evidence for any year. This applies particularly to examination scripts. What tended to be available from earlier years is work for separate components of the examination rather than the whole work of candidates. Coursework and any oral examinations were usually missing.

A national archive of essential evidence on examination standards has been established by the regulatory authorities. This should ensure that difficulties in this area are reduced in future studies.

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 biology: review of standards 1977–97

Introduction

Changes in A level biology examinations between 1977 and 1997 were influenced several factors. Among the most significant were:

- the introduction in the 1980s of a common core;
- the introduction in 1988 of GCSE examinations in biology with a defined set of national criteria.

Examination demands

Materials available

The reviewers considered a range of syllabuses, question papers and mark schemes from 1977, 1987, 1992 and 1997. Full details of the materials used in the review are given in Annex A.

Over 56,000 candidates took A level biology in 1997. About 55 per cent of these entered for the syllabuses used for that year in the study.

Syllabus content

Over the period of the review, there was a general trend towards amplification of details and greater transparency of content. This trend was reflected in changes in the statements referring to aims and assessment objectives. While these changes had little influence on the examination demands, the introduction of a greater range of support materials (such as the dedicated textbooks which appeared from 1992 onwards) reduced the burden on students by defining their reference base. Moreover the practice of some awarding bodies of drawing on these materials to provide stimulus material in examination papers changed the emphasis of questions. Where they had originally assessed the higher-order skills of interpretation and application, they were now largely assessing recall.

All syllabuses in the review contained essentially the same basic core of biological knowledge. However, this became defined in greater detail over the period. Although overall breadth changed little, there was a substantial change in emphasis. The 1977 syllabuses contained a significant proportion of descriptive biology based to a considerable degree on structure and taxonomy. Those in 1997 reflected the subsequent development of the subject, with the inclusion of material that placed greater emphasis on molecular biology in the fields of genetics and physiology. This change of emphasis substantially increased the examination demands.

Changes in syllabus content have been accompanied by greater specificity of syllabus definition. The 1977 syllabuses were defined more in terms of themes and concepts than in the precise detail that characterises all the 1997 syllabuses. A trend

towards describing syllabus content in terms of assessment objectives is characteristic of the 1997 syllabuses from OCR and, to a lesser extent, AQA/N. This contributed to a reduction in demand by defining and, to some extent, limiting areas where application and interpretation are assessed.

Assessment structure and coverage

Given the lack of syllabus specificity and absence of mark schemes, it was difficult to determine the depth of knowledge and understanding required by the 1977 question papers. All later papers, however, were considered appropriate in this respect. Although the use of objective tests was reduced between 1982 and 1997, greater use was made of short structured questions so that overall syllabus coverage remained acceptable.

Skills or processes were fairly represented in syllabuses and question papers, apart from those of 1977, which made little reference to either. The only significant exception was the extent to which synoptic skills were assessed in 1997 modular courses, where syllabus statements were not always reflected in the content of the question papers.

The general trend over the period of the study was towards reducing the proportion of marks awarded for recall and increasing the proportion awarded for showing more demanding skills. The effect of this trend on demand was balanced by increased structuring and greater direction in questions. Decreased emphasis on recall was also matched by an increase in the proportion of questions requiring interpretation of data and the application of knowledge to unfamiliar material. This shift was particularly marked in the case of AQA/A and OCR, where it was judged to represent an increase in demand. The 1997 AQA/N papers were considered to test these higher-order skills at a less demanding level than other awarding bodies.

The assessment of practical skills also changed. There was a gradual reduction in the emphasis placed on the ability to follow instructions. In 1977, this requirement, although only implicit, was a significant element in the practical tests. In 1997 syllabuses, it was neither specified nor a major factor in the assessment. The change followed the move to centre-based assessment and was accompanied by a greater emphasis on planning investigations, on designing experiments and on evaluation. Assessment in 1997 was generally through lengthy investigation rather than through the shorter, more discrete practical exercises that characterised the earlier practical tests. These features were considered to increase the demands made by 1997 syllabuses. However, centre-based assessment allowed more than one opportunity for assessment and provided a less stressful environment for assessment.

The written papers that formed the basis of the 1977 examinations typically offered a choice from a wide range of questions, many of which were relatively unstructured. The format of the 1997 papers was substantially different. There was little choice, a greater range of question style, and a greater emphasis on application of knowledge and interpretation of data at the expense of recall. These changes had been accompanied by general improvements in clarity and layout. Papers have become more accessible, with a greater consistency between awarding bodies in the use of

terminology and a simplification of the language used. Overall, these changes were considered to have had little effect on demand.

Options

In all years, there was a substantial difference between awarding bodies in the provision of options. Several offered little or no choice; others, notably OCR and AQA/N, offered extensive choices. Choice was further expanded in the 1997 modular syllabuses by allowing for combinations with modules drawn from subject areas other than biology. Extensive choice, and the resulting multiplicity of routes, raised considerable problems associated with parity of demand. Although wide-ranging syllabuses effectively reduce the breadth of study for individual students, this is offset to a considerable extent by the greater depth of knowledge that is required of particular topics. The inclusion of a 33 per cent mandatory core in these 1997 syllabuses resulted in a somewhat superficial treatment of certain key areas of physiology and ecology, which were only studied in appropriate depth by candidates taking the relevant optional modules. For those whose courses did not include these particular options, this represented a significant reduction in demand.

Summary

There were many changes between 1977 and 1997. These included greater attention to the presentation of examination papers, the switch from practical tests to the assessment of coursework, and the introduction of modular syllabuses. In addition, there was improved feedback from awarding bodies to teachers by way of INSET programmes and more comprehensive reports, as well as the provision of textbooks that targeted specific syllabuses. Not all the changes had an effect on demands. Moreover, many of the effects tended to cancel each other out, with increases in some areas being offset by decreases in others. Overall, the changes identified were judged to have had no appreciable effect on examination demands.

Standards of performance at grade A and grade E

Materials available

Scripts were available from all awarding bodies for 1997. Small numbers of scripts were available for 1992 (or adjacent years) for AQA/N, AQA/A, and CCEA at grade A and for AQA/N, AQA/A, CCEA and WJEC at grade E. None was available from 1977 to 1987. The small numbers and use of a different specification to identify the earlier work make any findings about performance over time very tentative.

Full details of the materials are given in table A2 of Annex A.

The descriptions of expected performance used in this exercise were developed from published grade descriptions, adjusted to take into account the fact that the work was from borderline candidates.

Standards expected at grade A

Candidates were expected to show a sound and consistent knowledge of facts, principles and concepts across a broad range of subject material. They were expected to be able to describe the trends and patterns shown by data presented in tabular or graphical form, where appropriate indicating anomalies and inconsistencies. They should be able to select relevant facts and principles to produce coherent, logical and comprehensive explanations and to solve problems involving both familiar and unfamiliar situations. Ideas should be communicated effectively, involving coherent explanations. Answers to longer questions should be logical and detailed. Appropriate biological language should be used both in referring to specific technical terms and expressing more general concepts. Numerical problems involving a range of arithmetic and algebraic skills should be solved accurately and supported, where appropriate, by logical explanation.

Performance at grade A

In 1997 at grade A, there was considerable consistency between awarding bodies. Although all grade A scripts showed evidence of some of the required skills, it was not always easy to determine the overall standard of performance in the modular schemes. This was partly because a number of the modular tests targeted a rather limited range of assessment objectives, while assessment in the linear syllabuses was more integrated. There was also some inconsistency between individual scripts from modular schemes, which may have been attributable to the form these scripts took. Overall, in scripts from AQA/N and WJEC, the standard attained was judged to be below that expected for grade A. The scripts of the other four awarding bodies were considered to meet the expected standard.

Of the three awarding bodies providing scripts from 1992, the work of AQA/A and CCEA candidates was considered to meet the expected standard. For AQA/N, however, performance in 1991 was considered to be above that shown by the other awarding bodies, in marked contrast to the work from 1997, which was below expectation.

Standards expected at grade E

Candidates were expected to show a knowledge of biological facts, principles and concepts. This may be displayed consistently across the questions or vary between good and poor on different areas. They should be able to describe broad trends shown by data presented in graphical or tabular form, and to select appropriate facts and principles to produce limited but relevant explanations and to solve problems concerning familiar material. When tasks deal with unfamiliar material, material should be relevant to the subject area even if there are difficulties in applying the facts and principles involved. Candidates were expected to communicate, with some degree of logic and coherence, although not all information used may be relevant. The use of technical language may be limited, but some specific terminology should

be used correctly. Simple numerical problems involving the rules of number and the calculation of percentages should be solved accurately and it should be possible to follow the steps in the calculation.

Performance at grade E

The 1997 scripts revealed that candidates at grade E were often characterised by the inconsistency of their performances. There were many different ways of gaining the total mark necessary to be awarded this grade, but these candidates generally tended to gain credit for recall rather than for their ability to interpret information or to apply their knowledge to the solving of problems. The structure of many of the papers, particularly of those forming part of modular schemes of assessment, was such that few grade E candidates were able to show consistency of performance across the full spectrum of assessment objectives.

The pattern displayed with the 1997 grade A scripts was repeated at grade E, with the quality of the work of candidates from four of the awarding bodies meeting the expected standard, while those from AQA/N and WJEC were judged below the expected standard. In many of these scripts it proved difficult to find evidence of skills other than those associated with recall of knowledge. This situation was compounded by ineffective mark schemes. The result was that candidates frequently gained sufficient marks from randomly scored points that were rarely assembled to provide a comprehensive answer. In addition, it was clear that the different options offered by some awarding bodies made different demands on candidates. In some cases, the relative weighting of assessment objectives did not appear to be closely matched between options, leading to doubts about whether standards at grade E were comparable within an awarding body.

As at grade A, performance was considered to meet expectations in CCEA and WJEC scripts for 1992 and for AQA/N in 1993. For AQA/N scripts for 1991, however, performance was again considered to be above expectations, in marked contrast to 1997. Reviewers were of the opinion that candidates gaining a grade E with this awarding body in 1991 might have been awarded a substantially higher grade in 1997.

Summary

Most awarding bodies applied appropriate standards at both grades and over time. However, neither AQA/N nor WJEC candidates were considered to have met the expected standards at either grade in 1997. Strikingly, AQA/N candidates were considered to have exceeded expectations in 1991.

Annex A: Materials used in the review

Table A1 shows the materials available for the review of examination demand.

Awarding body	AQA/A	AQA/N	CCEA	EDEXCEL	OCR	WJEC
1997						
Syllabus	✓	✓	✓	✓	✓	✓
Question papers	✓	✓	✓	✓	✓	✓
Mark scheme	✓	✓	✓	✓	✓	✓
1992						
Syllabus	✓	✓	✓	✓	✓	✓
Question papers	✓	✓	✓	✓	✓	✓
Mark scheme	✓				✓	✓
1987						
Syllabus	✓	✓		✓	✓	✓
Question papers	✓	✓		✓	✓	✓
Mark scheme				✓		
1977						
Syllabus	✓	✓		✓	✓	✓
Question papers	✓	✓		✓		✓
Mark scheme				✓		

Table A1: materials available for the syllabus review

Table A2 shows the materials available for the script review.

Year	Grade	AQA/N	AQA/A	CCEA	EDEXCEL	OCR	WJEC
1997	Grade A	15	15	14	15	13	15
	Grade E	15	15	14	15	13	15
1992	Grade A	4	2	3			
	Grade E	4	3	3			8
1987	Grade A						
	Grade E						
1977	Grade A						
	Grade E						

Table A2: Numbers of sets of candidates' work available for the script review

Notes:

Coursework was not available for AQA/N or WJEC in 1997 nor from any awarding body in earlier years

The AQA/A scripts were from 1993 rather than 1992; those from AQA/N were from 1991.

The OCR scripts from 1992 covered a range of options.

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/01/760