

Five year review of standards

GCSE history

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

GCSE History: review of standards 1977–97

Introduction

Changes in 16+ history examinations between 1977 and 1997 were influenced by:

- the introduction in 1985 of national criteria for assessing history at GCSE;
- the introduction of GCSE examinations in 1988.

These linked developments led to major syllabus revisions. The national criteria specified assessment objectives and techniques of assessment, including coursework. There were no later revisions to the criteria, so the syllabuses in 1997 were virtually unchanged from 1992.

Examination demands

Materials available

The reviewers used syllabuses, question papers and mark schemes from history syllabuses in 1977, 1987, 1992 and 1997, although not all these materials were available for the three earlier years. In some cases, examiner's reports allowed reviewers further insight into the expectations of the examination. Annex A shows in detail the materials used.

Between them, the syllabuses included in this study attracted about 57 per cent of the 230,000 candidates for GCSE history in 1997.

Presentation of syllabuses and question papers

One major impact of the introduction of GCSE was that the examinations had to meet the full range of abilities in a single tier of assessment. Partly as a result, syllabuses became more explicit over time in terms of content specificity, assessment objectives, type and nature of questions and grade descriptions. Examination papers became accessible to the wider ability range in terms of layout, rubrics, levels of language used in source material (in most cases) and the wording of questions. Reviewers did not consider that such improvements in the accessibility of question papers had altered the demands that they made

Content

The amount of content candidates had to cover was not reduced between 1977 and 1997. Although the 1977 and the majority of 1987 syllabuses covered a wide range of content, there was no equivalent expectation within the examination papers that candidates would study the whole range or period. The limited evidence from mark schemes in 1987 suggested the requirement was for generalised knowledge. Candidates were rewarded for what they knew, and what they did not know was simply never revealed. Little else was being tested other than the ability to select and

organise relevant knowledge. The only exception to this was the Edexcel multiple-choice paper, which had the potential to guarantee coverage that would be broad, albeit superficial.

In 1992, reviewers judged the breadth of content laid down in the syllabus as unchanged. However, there was greater use of structured questions, increasing demand on specific knowledge required of particular topics. There was also less question choice than there had been in 1977 or 1987. There were, therefore, increased demands made on knowledge in depth of topics and issues, while demand on breadth had at least been maintained.

Assessment

The most significant changes that occurred between 1977 and 1992 were in terms of the skills tested and the amount of time spent on assessment activities.

The syllabuses, question papers and mark schemes for 1992 and 1997 tested a much wider range of skills than the recall of knowledge, organisation and understanding that were required of candidates in 1977 and 1987. The additional skills tested were:

- interpretation of a wide range of sources and evaluation of their use as evidence;
- understanding of historical terminology and concepts (cause and consequence, change and continuity, similarity and difference);
- and empathy.

These skills were specifically tested across questions and papers and coursework, and candidates had to demonstrate attainment against four assessment objectives to achieve a good grade. This was reinforced by the use of 'levels of response' mark schemes which assess understanding/skills and knowledge. The demands placed on candidates were judged to have increased.

There were two other significant increases in the demands made on candidates between 1977 and 1992/97. First, there was a significant rise – approximately 30 per cent – in the external examination time between 1977 and 1997. Second, in addition to the written papers, the examination required candidates to complete a substantial element of coursework.

Demand across awarding bodies in 1997

Some differences were noted in the breadth and depth of content required by the different awarding bodies in 1997. In particular, CCEA and WJEC required coursework content coverage outside the core content, and AQA/N required coursework content to focus specifically on 'Colonialism and Human Rights'. However, this was balanced by the fact that AQA/N's question papers exclude areas of the syllabus content. Overall, the demands were considered broadly similar. However, the WJEC and CCEA syllabuses were of the 'Schools History Project'-type (SHP) so comparisons were not like for like. These SHP syllabuses were considered slightly more demanding in terms of content.

Presentation of question papers

There was a high degree of comparability across awarding bodies in terms of the presentation of the examination papers. Layout was spacious, font size reasonable, and key words and instructions were highlighted. Alternative questions were generally comparable and were achieved by most awarding bodies without resorting to the 'template' approach of OCR Paper 2. Good practice was noted in the case of AQA/A, which provided sources that were detachable from the main booklet, allowing them to be studied alongside the questions.

Assessment pattern

The 1997 syllabuses had been standardised across awarding bodies in terms of the four assessment objectives required by the national criteria. However, differences between the awarding bodies emerged in terms of the relative weightings given to each assessment objective; moreover, the weightings were not clearly shown for AQA/N or OCR. These variations were not considered to have a significant impact on comparability of demand across the awarding bodies, although the equal weighting given to the four assessment objectives and the testing of assessment objective 3 by examination made the WJEC syllabus marginally more demanding.

Question papers

The examination papers demonstrated a reasonable degree of comparability across the awarding bodies. There were some exceptions. WJEC paper 1 included a number of source questions where a high percentage of marks were awarded for low-level skills. AQA/N paper 2 contained a large number of questions that assessed source interpretation, although the syllabus stated that this was the focus of paper 1. Therefore, paper 2 made less demand on recall, conceptual understanding and application of knowledge than the papers of most of the other awarding bodies. OCR paper 1 questions were conceptually demanding and required considerable recall, which was confirmed by the mark schemes. In addition, OCR paper 2 used more demanding source material in terms of language than that of the other awarding bodies, and it required candidates to reach conclusions based on all the sources. Therefore, the OCR question papers were considered slightly more demanding than the other syllabuses. The AQA/N papers were considered the least demanding.

All the awarding bodies used 'levels of response' mark schemes, but differences were noted. Edexcel, CCEA, OCR paper 2 and AQA/A paper 2 provided little guidance in terms of indicative content or examples in order to make explicit what 'knowledge' was demanded. On the other hand, WJEC used content as the major discriminator in terms of measuring a candidate's performance.

Summary

The changes in history syllabuses over the period were judged not only to have significantly altered the nature of the demand but also to have increased that demand. What candidates were required to do had not been reduced in terms of breadth, depth or recall; but they were being required to do more in 1992/97 than in 1977. What candidates were required to do was also tested more effectively.

Despite slight differences between the awarding bodies in 1997, the level of demand was reasonably comparable.

Standards of performance at grade A, C and F

Materials available

Reviewers considered candidates' work from all the awarding bodies in 1997, and in far more limited quantities from 1992 and 1987. The details of what was used are provided in 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

The demands of the question will be understood and the answer will be focused and well organised. Historical knowledge will be used accurately, effectively and with consistency to substantiate arguments and reach historical judgements. Answers will be largely analytical, will demonstrate a good understanding of key concepts and show an appreciation that many of these are complex. Sources will be used critically and in relation to their historical context; they will be evaluated and used to reach substantiated and relevant conclusions. Events and issues will be explained with a clear understanding of the perspectives of people in the past, and the importance of looking for motives will be understood.

Performance at grade A

In 1997, the reviewers identified marginal differences in the performance of candidates from the various awarding bodies. For OCR, performance was felt to be above that expected. Candidates revealed extensive and accurate knowledge, critical evaluation of sources which were not stereotypical responses, and extended writing was analytical and displayed a developed understanding of key concepts.

Candidates from AQA/A were judged to have met expectations. Scripts from AQA/A revealed effective and focused use of knowledge, a good understanding of the key concepts with an ability to make judgements. There was some evidence of prepared or mechanical responses to source-evaluation questions. The coursework showed some excellent use of sources in context, although there was a wide variety of performance across centres as a result of their freedom to design tasks. Performance by candidates from Edexcel also met expectations on the externally assessed components, but there was some evidence of inappropriate targeting of assessment objectives within the coursework tasks.

For AQA/N, CCEA and WJEC, performance in 1997 was judged just below the expected standard. In each case, this was partly attributable to the nature of the question papers. With WJEC, structured questions with narrow targets impeded candidates' ability to meet the grade A descriptor. Source-evaluation questions did not target utility and reliability and were largely used as stimulus. Answers were also

descriptive rather than analytical, and there were fewer opportunities to develop conceptual understanding. There was also evidence within the mark schemes of rewarding knowledge simply on the basis of the quantity displayed. In scripts from CCEA, answers were predominantly narrative with little evidence of developed analysis or judgements. The broad nature of the question types with stems beginning 'Explain the events...' did not guide the candidates sufficiently towards analysis. Source evaluation was inhibited by the structure of the questions. The nature of the structured questions in the AQA/N papers did not encourage the higher levels of source evaluation or the ability to develop complex and critical answers, especially on the relationship of source to context. Where extended writing was required, it was felt that candidates lacked the critical edge demanded at this level; answers were predominantly narrative with some analysis. Candidates' understanding of key concepts did not meet the expected standard.

In 1992, performance in both AQA/A and WJEC was considered below the expected standard. For WJEC, the situation was very similar to that in 1997, with the papers not tending to produce the skills needed for the grade. For AQA/A, it was partly a result of the marking scheme for one of the papers, which tended to reward low levels of thinking too generously.

For 1987, only scripts from WJEC were available. These reflected the fact that the focus of assessment was very different from that in later years. They were responses to O level papers which tested recall, selection, organisation and deployment of knowledge in the first part of the essay question, together with conceptual understanding in the second part. The scripts displayed impressive, detailed, accurate and mostly relevant knowledge but, in terms of conceptual understanding, the answers did not, for example, attempt to prioritise causal factors or discuss continuity and change.

Standards expected at grade C

Most of the demands of the question will be understood and the answer will show relevant links to the question. Historical knowledge will be deployed, generally with accuracy and relevance to support the descriptions. Answers will demonstrate limited analysis and some understanding of key concepts. There will be evidence of some ability to evaluate and assess sources and to use them to draw relevant conclusions with limited supporting evidence. Some understanding of the perspectives of people in the past will be evident.

Performance at grade C

In 1997, performance was more consistent than at grade A. Candidates from all the awarding bodies, except WJEC, were judged to have met the expected standard. As at grade A, the problem in WJEC was felt to stem from the question papers and performance, as a result, did not meet expectations.

In 1992, as at grade A, standards of scripts from both AQA/A and WJEC were considered just below the expected standard, and for the same reasons. Scripts from 1987 at this grade reflected the very different nature of the examination. They were

judged to exceed expectations in terms of recall and deployment of knowledge, but to fall below it in terms of conceptual understanding.

Standards expected at grade F

Answers will include some relevant and accurate information that is from the syllabus content. The knowledge will show some basic understanding of key concepts.

Answers will show ability to understand particular pieces of evidence and to make simple comparisons between two sources. These will tend to be understood at face value, with some limited/generalised information extracted. Limited knowledge of the perspectives of people in the past will be evident.

Performance at grade F

The picture at grade F was very similar to that at grade A. Scripts from OCR were judged above the expected standard. They met the demands of the question with some relevant and accurate knowledge and limited source evaluation. Scripts from AQA/A and Edexcel matched expectations, while those from AQA/N, CCEA and WJEC were below the required standard. With both AQA/N and CCEA, in particular, it was evident that the weaker candidates did not understand the demands of the questions and there was also evidence of weak scripts having been rewarded too generously.

Summary

There were small variations in standards of performance across the awarding bodies in 1997 as a result of the different demands placed on candidates in terms of question types and expectations within mark schemes. Across all awarding bodies, the differences are marginal in that performance was never considered well above or below the expected standard. Overall, OCR was marginally more demanding, while AQA/N, CCEA and WJEC were marginally less demanding. The main area of concern lay in the nature of the questions, which impeded candidates' performance across all levels of response but, in particular, at grade A.

Comparisons over time were limited to two awarding bodies in 1992 and only one – at two grades – in 1987. There had also been significant changes in the nature of what was being assessed. Thus, conclusions in this element of the study are tentative. The performance of candidates in 1987 was particularly strong in the use and deployment of accurate knowledge, but performance was weaker than in 1997 in conceptual understanding, displaying analytical thinking, prioritising causal factors and discussing continuity and change. The performance of candidates in 1997 was not weak in the use of accurate knowledge; it was just not as strong as in 1987. This has to be balanced by the fact that candidates in 1997 were being assessed against more objectives – source evaluation and sensitivity to perspectives of people in the past. The demands made on candidates in 1997, against a wider portfolio of skills, were greater and more explicit. So, while performance in 1997 in the one skill that was identical to that of 1987 was not as strong, it has to be remembered that candidates in 1987 were assessed solely on that one skill. Overall, the reviewers considered that the performance of candidates had improved over time.

Annex A: materials used in the review.

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

Awarding body	AQA/N	AQA/A	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 A2 shows the materials available for the script review.

Awarding body	Grade	AQA/N	AQA/A	CCEA	EDEXCEL	OCR	WJEC
1997	Grade A	15	15	14	15	15	15
GCSE	Grade C	15	15	14	15	15	15
	Grade F	15	11	11	15	15	15
1992	Grade A		5				5
GCSE	Grade C		5				6
	Grade F		5				3
1987	Grade A						15
O level	Grade C						15
1977	Grade A						
O level	Grade C						

Note:

The coursework was also available for every candidate only for the batches of scripts in bold, although there were limited quantities available at most grades in 1997.

Key to the awarding bodies

During the period of the reviews, the number of awarding bodies operating fell². 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/01/768