

Five year review of standards

GCSE chemistry

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

Chemistry 16+: review of standards

Introduction

In 1996, SCAA, QCA's predecessor body, together with Ofsted, conducted an enquiry into standards of public examinations between 1975 and 1995. This work included 16+ examinations in chemistry. The results of the work, published as *Standards in Public Examinations, 1975–1995* (SCAA, 1996), included a series of recommendations concerning future examinations in each subject reviewed.

Also in 1996, new GCSE chemistry syllabuses, based on revised subject criteria, were introduced, leading to first examination in 1998. In terms of the demands they set and the grading standards they applied, these syllabuses offered the first real opportunity to evaluate the extent to which the revisions to the criteria and syllabuses based upon them had met the recommendations in the first report.

The key changes to GCSE chemistry examinations between 1995 and 1998 were:

- changes to the content of the national curriculum orders for science;
- changes to the assessment weighting of core and extension material;
- the change from a three-tier examination to a two-tier examination.

Syllabus and assessment demand

Materials available

Reviewers used syllabuses, question papers and mark schemes from chemistry syllabuses in 1998 and the materials used in the 1996 enquiry.

About 47,000 candidates took GCSE chemistry in 1998. About 94 per cent of those entered for the syllabuses used for that year in this study.

Background

In 1995, syllabus content was largely based on the awarding bodies' interpretations of the national curriculum programmes of study and statements of attainment. Most of the statements were general in nature but closely tied to national curriculum levels. In 1998, most syllabuses were based on learning outcomes which were set out in precise terms and often illustrated by examples and contexts. In general, the 1998 syllabuses indicated more clearly the expected level of demand. All awarding bodies incorporated changes made to the national curriculum orders for science into the core content of their 1998 syllabuses. This was not judged to have affected the overall level of demand.

Content and assessment objectives

In 1995, differences between the awarding bodies in core content were limited. CCEA had no geology or radioactivity but had more group chemistry and environmental issues. Edexcel had no gas laws or radioactivity. The latter was also missing from AQA/N and WJEC. OCR and AQA/A were considered the most demanding syllabuses because they contained the most material in terms of core content. Variations in the extension material were, however, more considerable. In this respect too, OCR was considered the most demanding with electrochemical cells, shapes of molecules and ions, and extraction of tin, zinc and lead all included.

In 1998, with the exception of CCEA, all awarding bodies had common assessment objectives and core content. (CCEA omitted earth science.) The strong emphasis on application of knowledge and social, economic and technological (SOCET) aspects in the 1995 WJEC syllabus was not present in 1998. Some level 8 content from the 1995 AQA/N and WJEC syllabuses appeared in the 1998 versions as Foundation Tier material.

There were variations in the extension material and in the way core and extension material was presented in the 1998 syllabuses. The CCEA syllabus did not distinguish between core and extension material. In the AQA/N syllabus, core and extension material were presented together under three major themes. Edexcel, OCR and AQA/N had the most extension material, although the changes to the core and the redistribution of the extension material by Edexcel between 1995 and 1998 made the extension material less demanding at the Higher Tier. AQA/N's extension material was considered more demanding in 1998 than in 1995 though AQA/A's was the most demanding. WJEC had a high proportion of less demanding material. For example, the section on acids, bases and salts was key stage 3 material and chemical calculations were confined to extension material.

Scheme of assessment

The most significant changes from 1995 to 1998 were in the scheme of assessment. The change in the assessment weighting of the extension material in the 1998 syllabuses from 35 per cent to 25 per cent, and the equivalent increase in the core from 65 per cent to 75 per cent, was seen as lowering the demand in the more difficult aspects of chemistry such as quantitative work and balanced equations.

There was some variation in the way the awarding bodies presented core and extension material in their question papers in both 1995 and 1998. In 1995, AQA/N, WJEC and OCR used a single paper to examine the full range of content. The other awarding bodies examined core and extension material in separate papers, allotting more time overall. It is difficult to assess which of the two models was more demanding. Although the one-paper examination placed a high demand on candidates, the longer examining time of the two-paper examination meant there was more syllabus coverage, which was also demanding.

In 1998, AQA/N and WJEC continued to examine core and extension material through a single examination paper. CCEA used two examination papers with core and extension material included in both. As in 1995, the other awarding bodies

assessed core and extension material in separate papers, with the core papers common to the double award science examination.

Spelling, punctuation and grammar (SPAG) were assessed in all components of the 1995 examination. This requirement was changed in 1998 syllabuses, where this was assessed in coursework only. This was not considered to have affected demand.

Examination time

In 1995, CCEA, which included a multiple choice paper, had the longest examination by one hour and Edexcel allocated 30 minutes more than the other awarding bodies. In 1998, awarding bodies made adjustments to accommodate the new two-tier system. AQA/N and WJEC increased the examining time for Foundation Tier candidates while reducing time for the Higher Tier papers. AQA/A and Edexcel used an equal amount of time for each tier; the others gave more to the Higher Tier. Although CCEA had reduced its total examining time by 30 minutes, Higher Tier candidates were still examined for one hour longer than those of all other awarding bodies. With the exception of CCEA's Higher Tier examination, there was no significant difference in the demand of papers owing to examining time.

Tiering

The changes to the tiering system were, however, highly significant. The use of three tiers in 1995 allowed demand to match candidates' abilities more closely than the two tiers in 1998. This affected candidates of differing abilities in different ways. The change meant that the Higher Tier papers had more lower demand questions and fewer at the highest level: so demand was reduced. In contrast, the Foundation Tier papers contained fewer questions at the lowest level and some at higher levels thereby increasing demand. The effect of the change for candidates previously entered for the Intermediate Tier would clearly depend on which tier they were entered for in 1998.

Structure of examination papers

There were some changes between 1995 and 1998 in the types of questions set by all awarding bodies. In 1995, the comparative demand of question papers across awarding bodies varied between tiers. At Foundation Tier, most questions were structured requiring a simple, short answer. These ranged from 70 per cent of questions on AQA/A papers to 53 per cent for Edexcel. Questions involving simple choices, such as those in which the candidate had to match a number of given terms (such as polymerisation and neutralisation) with a range of definitions or cloze exercises, were most frequent with CCEA, WJEC and OCR. Questions requiring an answer written in prose were least frequent on CCEA and WJEC papers. At the Higher Tier, simple choice questions were far fewer with the exception of CCEA. Questions requiring answers in prose were more common in the Higher Tier papers and similarly demanding across awarding bodies.

In 1998, at least 50 per cent of the Foundation Tier papers for all awarding bodies consisted of simple structured questions requiring short answers. AQA/N, Edexcel and OCR Foundation papers were made up principally of this question type, while CCEA was notable for the increase in the number of questions to be answered in

prose. On the Higher Tier, AQA/A had the largest number of short answer questions and CCEA had a higher proportion of questions requiring an answer in prose than other awarding bodies.

Calculations and chemical equations

In 1995, questions requiring candidates to use chemical equations were fewest on WJEC and OCR Foundation Tier papers. On the Higher Tier papers, calculations featured most with WJEC and Edexcel and least with AQA/A and OCR. Chemical equations and the use of formulae appeared most on AQA/A and Edexcel papers.

The CCEA Foundation Tier papers showed an increase in demand between 1995 and 1998 as candidates were required to do more calculations and chemical equations. The Higher Tier papers also featured more chemical equations. OCR similarly increased the weighting of chemical calculations, formulae and equations in the Higher Tier papers. The Foundation Tier questions remained mainly short answer but did have more questions about equations, showing an overall increase in demand. OCR reduced the number of questions dealing with social, economic and technological aspects and environmental issues. AQA/A Foundation Tier papers included more chemical calculations and the number of marks available for calculations and chemical equations was slightly increased on the AQA/N Foundation Tier papers. At Foundation level, AQA/N papers were, however, low on word equation testing and Edexcel papers had relatively few calculations.

In 1995, CCEA and AQA/N Foundation Tier were considered the least and Edexcel the most demanding. CCEA papers had little extended prose and gave high weighting to simple choice questions. At Higher Tier, OCR was judged the least demanding and Edexcel again the most.

In 1998, CCEA and AQA/A were judged to be the most demanding overall at Foundation Tier and Edexcel the least. At Higher Tier, CCEA was considered the most demanding and AQA/N the least. Higher Tier papers from all boards had a similar proportion of calculations but the demand for understanding chemical equations was greatest with CCEA and AQA/A and least with AQA/N.

Coursework

For coursework, changes in the specification of Sc1 from 1995 to 1998 were most significant at the highest levels of attainment. In 1995, it was considered that levels 9 and 10 were too demanding and changes in the national curriculum addressed this by deliberately lowering demand for exceptional performance. These changes were reflected in all syllabuses. It was also considered that the move away from all coursework being based on whole investigations had decreased demand slightly. In contrast, in 1998, evaluation was required of all candidates in their coursework. This increased demand particularly for lower ability candidates.

Summary

There was more uniform demand across awarding bodies in 1998 than in 1995. This arose from common assessment objectives, more common examining time allocation, a common coursework demand and a common core content which had become a greater percentage of the whole syllabus. The extension material still varied between awarding bodies and, although demand in terms of question difficulty

was generally similar, some awarding bodies had a greater quantity of syllabus content to cover. The depth of treatment of the same content appeared from the mark schemes to be similar across awarding bodies.

The reduction in the extension content weighting meant a potential lowering of demand since calculations, use and knowledge of chemical formulae and chemical equations tended to be focused in the extension paper or in questions associated with the extension material. Changes in the tiering structure had a differential effect on demand, with that on the Higher Tier reduced and that on the Foundation Tier increased. There had been a conscious move to reduce demand for the most able in coursework, while the requirement for evaluation had increased demand for lower attaining candidates.

There had been some increase in the testing of chemical equations and calculations, although for several awarding bodies the proportions remained low. Question types were similar between 1998 and 1995. While there were opportunities to demonstrate the ability to organise knowledge, there was slightly less extended prose in 1998. There was generally an increase in the weighting of chemical calculations. The balancing of equations had increased, but there were significant variations between awarding bodies. In general, there was a slight increase in demand in these higher level aspects of chemistry.

Standards of performance at grades A and C

Materials available

Reviewers considered candidates' work from all the awarding bodies² in 1998 and the materials from the 1996 enquiry for comparison.

Standards expected at grade A and grade C

Reviewers used the grade descriptions contained in the 1995 National Criteria for Science. These were discussed in a plenary session led by the co-ordinator. The aim was to ensure that their particular application to GCSE Chemistry was agreed and how they needed to be applied for borderline rather than typical candidates was understood.

Performance at grade A

At grade A, candidates' performance was judged to be either well matched to expectation or just below it. Edexcel and CCEA scripts were found to be consistently well matched to the grade A description, though it must be noted that for CCEA there was no coursework available. For OCR and WJEC, candidates' performance was equally divided between well matched and just below; nearly all AQA/N scripts were judged to be just below.

² There was no work from AQA/A: candidates from the AQA/N syllabus were taken to represent the standard of that awarding body.

Performance at grade C

Performance at grade C at Higher Tier consistently matched the grade description for AQA/N and WJEC candidates. In the case of OCR and CCEA, judgements were divided between well matched and just below. At Foundation Tier, AQA/N, CCEA, OCR and WJEC candidates matched the grade description, although this must be seen in the light of the extent to which the papers allowed the candidates to match expectations. This is discussed in more detail below. The performance of Edexcel candidates at this grade for both tiers was mainly just below and occasionally well below the expected standard.

Targeting of the papers

Reviewers found that the exercise of reviewing the scripts also provided clearer insights into the extent to which the examinations provided opportunities for candidates to show the knowledge, understanding and skills in the grade descriptions and the effectiveness of the questions common to both tiers.

Foundation Tier

The Foundation papers of CCEA, Edexcel and AQA/N had relatively little overlap with the Higher papers, whereas OCR and WJEC papers shared a good range of common questions. The match of performance between Higher and Foundation Tier candidates at the grade C boundary was generally good on the common questions where awarding bodies had substantial overlap.

In general, however, the targeting of the Foundation Tier questions was not always sufficiently precise to allow candidates to demonstrate competence as defined in the grade C description. In particular, candidates at Foundation Tier had fewer calculations and little or no balancing of equations. The Edexcel paper contained no balanced equations and only one calculation. In addition, candidates were not required to explain patterns. The WJEC paper had only one question testing balanced equations with no questions allowing candidates to demonstrate their ability to solve quantitative problems or handle data. AQA/N, CCEA and OCR similarly had few questions on balanced equations or requiring calculations. CCEA did not require Foundation candidates to explain patterns or handle data. The OCR Foundation paper did, however, cover formulae and structures and tested a good spread of skills. AQA/N succeeded in testing recall of information but there was insufficient testing of skills to match the description for grade C.

Higher Tier

On the Higher Tier, CCEA candidates were required to perform at a very high level, so much so that it was difficult for grade C candidates to do well. The paper was based on a limited amount of the syllabus content and calculations were wide-ranging and difficult. Candidates had to tackle a range of balanced equations though they were not asked to apply their knowledge. There were several unstructured questions to be answered in prose but these focused mainly on recall of information. Edexcel's examination was also of high demand, providing a good blend of structured and unstructured questions. There was a range of chemical equations and straightforward calculations though candidates were not required to construct or use balanced equations. The OCR papers had good syllabus coverage with questions structured in a simple and straightforward manner. The demand for answers written

in prose was therefore low. Quantitative questions allowed candidates to show their ability, and testing of application of knowledge was effective but there was limited testing of equations and bonding. AQA/N set most questions in an applied context but calculations were mainly of a low demand. There was only one question on balancing equations and opportunities for writing in prose were few. In contrast, there was a good demand for prose answers in the WJEC papers though much of the descriptive writing involved recall of information. Generally, questions concerned with bonding, structure, equations and calculations were not very demanding. There were no ionic equations and testing of understanding and application of knowledge was limited. Overall, it was thought to be a little harder to obtain a grade C on the Higher than on the Foundation Tier. This was particularly true of CCEA.

Coursework

There was a significant variation in the quality of coursework of candidates at a common grade: much had been marked generously. This was particularly the case with AQA/N and to a lesser extent WJEC and OCR. Some AQA/N work had been marked to the previous assessment scheme. Edexcel coursework was consistently well marked.

In some cases the activity did not offer the demand associated with the marks awarded. Often there was little correlation of predictions or observations with theory and understanding. Generally, candidates' predictions, conclusions and evaluations were weak.

Summary

One awarding body at grade A and one (different) one at grade C (both routes) were judged consistently not to have met the expected standard. Foundation Tier papers were not felt to give candidates sufficient opportunities to show their higher order skills, so that it was harder to obtain a grade C on the Higher Tier. There were significant inconsistencies in the marking of coursework for most of the awarding bodies.

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/01/771