Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

Qualifications and Curriculum Authority

Archived Content

Inter-subject comparability studies

Study 1b: GCSE, AS and A level sciences

May 2008

QCA/08/3652

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Archived Content

Contents

Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

1 Personnel

The team comprised 12 reviewers and was made up of a mixture of existing senior QCA consultants and those recruited by advertisement. The team was structured so that four consultants had specialist expertise in each of the three main sciences. Each was also asked to consider a second science. It was not possible to make this fit perfectly with their areas of expertise, but all had sufficient experience of teaching all the sciences. This meant that each separate science was considered by eight reviewers, four with that subject as their main specialism and two with a specialism in each of the other two sciences. All reviewers considered GCSE science double award. Three of the existing QCA consultants, each with a background in a different science, were asked to act as lead consultants. The names of participants are provided in Appendix A.

Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Archived Content

2 Materials

The syllabuses reviewed were selected on the basis of the size of candidate entry.

Table 1 The syllabuses used for the study

	Biology	Chemistry	Physics	Double award
GCSE	AQA (3411)	AQA (3421)	AQA (3451)	AQA (3462)
A level	OCR (3881/7881)	AQA (5421/6421)	Edexcel (8540/9540)	N/A

The inclusion of the double award generated particular challenges for reviewers. While much of the examination material was shared with the individual science subjects (the content for the separate sciences comprises that in the double award with a quantity of extension material, and the examination papers have a good deal of overlap), they had to judge whether the total demand in the double award was twice as great as that in each separate subject.

Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

3 Methodology

3.1 Form A

Reviewers used Form A to provide a factual analysis of the syllabuses, question papers and mark schemes. A generic Form A, as used across QCA review work, was considered by the lead consultant(s) and slight alterations were made to the questions to focus reviewers' attention more clearly on issues relevant to the particular nature of the task.

3.2 Form B

Reviewers used Form B to identify differences in demand between the subjects they were reviewing at individual qualification level. One Form B was completed by each reviewer for each pair of review subjects/levels. Reviewers were asked to complete Form B in the light both of their comments on Form A and their completed CRAS forms (see below).

Reviewers used a five-point numerical scale to make judgements about demand, ranging from 1 (very undemanding), to 3 (about right) to 5 (very demanding) to assess the qualification for each subject reviewed, as illustrated in Table 2. After making each numerical judgement, each reviewer was asked to give a brief summary of the reasons for that judgement. Reviewers then used these numerical judgements and their explanatory comments to make comparative summaries of the demand in the two qualifications. Each reviewer came to a conclusion about overall demand.

Table 2 Judgement scale

Key	**Very undemanding**		**About right**		**Very demanding**
Score per team member	1	2	3	4	5

3.3 Form C

Reviewers completed Form C as a summary of all their judgements. Form C enabled reviewers to gain an easy overview of their pattern of judgements across levels as well as between subjects. Reviewers transferred the numerical judgements made on Form B for each individual qualification to Form C and then, where necessary, added summative comments. The sections in Form C were identical to the sections in Form B.

© 2008 Qualifications and Curriculum Authority

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Archived Content

3.4 CRAS analysis

The CRAS analysis was used to enable the reviewers to reach judgements about the cognitive demand of the question papers, based on the nature of the questions, rather than the subject content. Reviewers were asked to assess the extent to which the question papers made demands in terms of:

- the *complexity* of the processes required to answer a question
- the extent to which the *resources* needed to answer the question were provided on the paper
- the level of *abstractness* of questions
- the extent to which candidates were required to generate a *strategy* in their answers.[1]

To do this, they used a numerical scale and recorded their judgements on forms designed for the purpose.

Reviewers were given a detailed explanation at the initial briefing about each aspect of the CRAS analysis and there was a general discussion about the ways in which the demands of a particular question could be manipulated by making adjustments to the question in terms of *complexity*, *resources*, *abstractness* or *strategy*.

For this study reviewers used a ten-point scale, with the assumption that foundation level questions would be likely to fall within the range 1–4, intermediate questions 3–6, AS questions 5–8 and A2 questions 7–10. This provided a four-point scale for each level, with what seemed a reasonable degree of overlap. It was made clear, however, that reviewers did not need to restrict themselves to the range for the level. In the event most ratings fell within the target ranges, but there were some occasions when reviewers rated particular factors outside that range.

[1] These factors had been identified in a study into question structure by University of Cambridge Local Examinations Syndicate (UCLES) commissioned by QCA. Each factor has the capacity to make examination questions more or less difficult, irrespective of the subject content. The exact interpretation of the four factors is often, to a degree, subject dependent. Explaining any subject-specific aspects was one of the tasks carried out by the lead reviewers.

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

4 Findings of the review of examination materials

4.1 Assessment objectives.

Reviewers considered that the assessment objectives resembled aims of a course rather than assessable objectives, particularly at foundation tier. Often, the assessment objectives were irrelevant to the actual demand not only because of this but also because of a variable approach to allocating certain items/questions to particular assessment objectives at the paper construction stage. It was pointed out that having common objectives across the sciences often limited the development of discipline-specific objectives, particularly at a higher tier. There was also a poor match of experimental/investigative objectives to relevant work in biology. The impact of common science objectives remained at AS and A2.

Although reviewers recognised the rationale for using common assessment objectives, they concluded that the approach did not reflect the individuality of each subject and did not actually lead to comparability. For example, chemistry at AS could be seen as more demanding in relation to assessment objective 3, given that its coursework component was less important than in biology and the higher order assessment objective 2 was comparatively more so. Equally, in the physics syllabus used, assessment objective 3 was assessed through a practical exam.

Reviewers considered that the assessment objectives were demanding for foundation tier candidates in biology, physics and double award at GCSE. They also judged that the assessment objectives were slightly more demanding in higher tier chemistry and double award than in biology and physics. At AS and A2 they found that chemistry was slightly more demanding in terms of assessment objectives than biology and physics.

Reviewers noted that progression from foundation tier GCSE through to higher then to AS and A2 did not reflect reality. This led to the important question of the relevance of the current science content and assessment for foundation tier candidates.

4.2 Schemes of assessment

All reviewers commented that the length of the papers in the GCSE separate sciences unduly increased demand.[2] This was particularly acute for foundation tier candidates where high mathematical demands exacerbated this issue. At AS, total examination times [assuming

[2] The particular syllabuses used were a suite in which the external assessment comprised a single long paper. Other syllabuses divide that assessment into two, which presents different demands.

© 2008 Qualifications and Curriculum Authority

coursework] were markedly different, with biology having 2 hr 45 minutes in total while chemistry had 4 hrs 15 minutes.

In addition, there were differences in the way choice was used across the subjects at A level. The physics syllabus offered some options in the AS, whereas biology offered choice only at A2. This physics syllabus also had a compulsory practical rather than coursework.

There were differences between the subjects in the assessment of the ability to synthesise knowledge, understanding and skills (assessment objective 4). This had a major effect on demand. In chemistry[3] a multiple choice/completion test was used to test knowledge recall across the course, but this approach required little analysis or synthesis and reviewers judged that it was less demanding of synoptic skills than the approaches adopted in the other two sciences. Physics and, to a greater extent, biology had more demanding ways of assessing assessment objective 4, often through extensive writing. In addition, the allocation of 40 per cent of A2 marks to a two-hour synoptic paper in physics significantly increased the demand.

In terms of schemes of assessment, reviewers found that the separate sciences were much more demanding than double award at GCSE foundation tier. This was also the case at higher tier, although physics was judged the least demanding of the three separate sciences. Reviewers judged the schemes of assessment at AS to be about right in their overall demand, despite differences in examination time. Reviewers found that A2 chemistry was more demanding than biology or physics, although the approach to synoptic assessment in chemistry was not as demanding as in the other two sciences, as detailed above.

4.3 Syllabus content

Reviewers identified several distinct problems with syllabus demand, noting in particular that there was a heavy content load in all subjects both at GCSE and A level. Chemistry was judged to be very demanding in terms of content at all levels.

Reviewers identified the following issues at GCSE in all four subjects (to a greater or lesser extent).

- They expressed concern about the distribution of content between the tiers.
    - In some cases, a particular piece of content within a topic was assigned to the higher tier when it could have easily rested in foundation tier and made for more

[3] This was a particular feature of the syllabus used in the study. In other chemistry syllabuses, different approaches are used for assessing synoptic skills.

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Archived Content

coherent teaching and learning. As a result, questions addressing the topic at foundation tier were narrow in their focus.

- In other cases a whole topic area was assigned to foundation tier, with the result that large elements of the topic were demanding for the candidates. Conversely, where a whole topic area was assigned to a higher tier, large elements of topic were often undemanding for the candidates.

- There were serious reservations about the usefulness of foundation tier subject content. Reviewers did not consider it relevant for those not going on to further study.

- Some areas of subject content were conceptually very demanding. They were perhaps too abstract, less relevant to everyday life or low in pedagogical opportunity.

- Overall, reviewers considered the level of demand of the syllabus content in chemistry to be too high for both foundation and higher tier.

In biology, for example:

- how the body exhales air was the only higher tier content in a very simple breathing module
- accommodation was the only element of eye structure/function which was on the higher tier
- the detail of structure and function of bone was confined to higher tier in the locomotion unit
- the reasons for energy loss in food chains was the only higher tier content within energy and nutrient transfer.

This would make coherent teaching and learning difficult, as well as resulting in a constraint for principal examiners when setting question papers.

In chemistry the problem was slightly different. For example, the ionic nature of neutralisation was the only part assigned to higher tier in a demanding and large unit on chemical reactions. Similarly, representing reactions, quantitative chemistry and reactions involving enzymes contained some very challenging content for foundation tier candidates. Conversely, useful products from oil, atomic structure, rates of reaction and the rock cycle were left as common units and were therefore simplistic and lacking in demand at higher tier.

In physics the characteristics of waves, the solar system and electromagnetic forces were common to both tiers and therefore simplistic and undemanding at higher tier. Frictional forces

© 2008 Qualifications and Curriculum Authority

and non-uniform motion were included in a common unit with accessible work on stopping distances and less accessible content on acceleration due to gravity. In addition the treatment of this topic on question papers might be constrained because *f* = *ma* was in a unit on forces and acceleration at higher tier.

In science double award, the differences between biology and chemistry, and, to a lesser extent, physics remained. Here, however, the distribution of content between foundation and higher tiers seemed generally more logical and coherent.

Overall, reviewers considered the level of the demand of the syllabus content of chemistry was too high at both foundation and higher tiers. For foundation tier, this was mainly owing to the inclusion of three topics that extended the national curriculum: industrial processes, quantitative chemistry and aqueous chemistry. Each of these topics required a great deal of rote learning and included concepts that were challenging for foundation tier candidates. The quantitative chemistry topic was very difficult indeed for most foundation tier candidates. In addition many of the units were very large. Aqueous chemistry, for example, encompassed key stage 3 water cycle, hard/soft water, treatment of drinking water, solubility and interpretation of solubility curves, acids, alkalis and the reactions of metals, salts with acid and the development of ideas.

It was also noted that there were different approaches to the examination of key stage 3 material which constrained question construction, with the lack of direct assessment of key stage 3 material leading to gaps in the assessment of some key ideas.

In physics and, to some extent chemistry, at higher tier, demand was significantly increased by the inclusion of demanding mathematics skills.

Reviewers judged that progression in content terms should only really apply from higher tier to AS and then to A2. Although it is possible to obtain a grade C on foundation tier, in terms of the breadth and depth content covered, reviewers judged that it was not an appropriate springboard to AS in any of the separate sciences or from the double award. The topics at foundation were from a traditional science programme and reviewers considered that more use could be made of the more applied and interesting aspects of science.

Reviewers considered the value of the content for those not progressing to the next level. They considered that opportunities for introducing research, presentation, discussion and real investigation were missed at GCSE. Applied aspects of science were more of an added-on section than a driver and were missing completely from the double award. This meant that the

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content
Archived Content

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Archived Content

content was largely irrelevant to everyday life and future, unless a candidate was going to study science at the next level.

There were also some concerns specific to AS and A2:

- the level of content demand in all three subjects was unnecessarily high
- chemistry was more demanding in content terms when compared with physics and biology
- optional routes led to differences in demand within a subject
- the interpretation of socioeconomic and technological aspects was narrow and there was no compulsory coverage of post-classical physics or modern chemistry
- the core concepts module in biology was very content heavy.

Reviewers noted the interrelationship of the sciences and their relationship to mathematics. This was particularly the case for progression from AS to A2. It would be difficult, for example, for a candidate to succeed in biology at A level without some confidence and competence in chemistry, as well as some mathematical knowledge. Similarly, A level physics had less overlap with the other sciences but required fluency in mathematical and symbolic thinking, in addition to a knowledge of specific mathematical techniques. Reviewers felt that these overlaps could be highlighted more by awarding bodies. However, it was noted that in biology it was difficult to find any evidence within the content that students were expected to think quantitatively, let alone make calculations with understanding.

The large range of topics was felt to be a major influence on the increased demand within the separate sciences at GCSE and in A level biology. It was also noted that there was often no clear rationale for the range of topics, particularly in biology. Not only did this add to the content burden but it also increased demand through the assessment due to the effect of content sampling within question papers. This was particularly acute in synoptic papers.

The consideration of the range of topics brought out an interesting difference of view between reviewers. Some reviewers considered that a broad range generated an increase in demand by providing more diverse content to be mastered. Others thought that a narrow range increased demand. On this argument, the narrower range in chemistry at GCSE made this subject more demanding than the other two. The reasoning was that if candidates found changing materials difficult, then half of the syllabus would be inaccessible to them. On the other hand, the broader range of topics in physics and double award science meant that if candidates found one topic difficult, there was still a large proportion of the syllabus which was accessible to them.

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

In addition to the sheer volume of content that candidates had to learn, especially at GCSE, reviewers judged that the key points were often standalone facts, all of which the candidates needed to learn. Reviewers took a similar view of chemistry at A2, which had a very high number of topics that were not especially interrelated.

An overall judgement concerning the demands of syllabus content raised the following observations:

- reviewers judged chemistry to be very demanding at all levels
- at GCSE, reviewers found that biology was about right in terms of demand whereas physics and double award were slightly too demanding
- in biology, there was a large jump in terms of content from GCSE higher tier to AS
- at AS, physics was seen as the least demanding in terms of content, but the step from AS to A2 was large
- the step from GCSE double award to A level in the separate sciences was large
- there was variation in the level of demand between options, for example, the genetic option in biology was very demanding, whereas the environmental science option was not.

4.4 Nature of assessment materials

Reviewers looked at question papers and mark schemes to arrive at judgements about the coverage of assessment objectives, time available per question paper, the nature and accessibility of the assessment tasks.

Although the coverage of assessment objectives was broadly about right, reviewers noted that at GCSE, the assessment of application of knowledge and understanding, analysis and evaluation (AO2) was at its upper limit (or beyond) in physics. Elsewhere, there was little assessment involving AO2, particularly for the double award, as this contained few applications and was mainly concerned with academic content. In addition it was noted that what might have seemed to be application at the paper construction stage was in fact testing recall, with limited opportunity for stretching higher attaining candidates. Evaluation was covered only to the extent of interpreting data from tables and graphs.

In physics there were also fewer questions, but with higher tariffs, so syllabus coverage was narrower than in biology. Consequently, topics in physics were tested to greater depth. However, there was a heavy content load in biology that represented a major demand for candidates.

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

Often, particularly in chemistry, difficult content was tested in an inaccessible way that increased demand. The opposite was seen in some biology papers where easier, more concrete content was tested in an engaging relevant way.

Reviewers varied in their judgements about the demand of double award science, particularly at foundation tier. The examination used shorter question papers based on the [less demanding] core material, coupled with a narrower range and number of topics covered. Several reviewers therefore considered it less demanding. However, others judged that the physics and chemistry content were inaccessible, coming together to form an inappropriate level of demand for this candidature.

At GCSE the range of tasks within the chemistry papers was particularly narrow. There was little except short or structured questions in any of the examinations. Biology and physics showed much more ramping in difficulty across the levels by increasing the amount of unstructured answers. In addition there was very little graphic or pictorial stimulus on the chemistry papers. This made them particularly inaccessible at foundation tier. This problem was compounded by crowded question papers, the use of unusual contexts and the use of much technical language. Reviewers were divided in their opinions about contextual information in relation to demand. Many felt that increasing the contextual information within a question increased accessibility, particularly at foundation tier, whereas others argued the danger of increasing superfluous information and distracters.

At higher tier there was again a major element of recall and straightforward application, particularly in biology. There was also a difference in the way higher tier papers set more challenging questions. Biology required several-sentence descriptions or explanations; in physics there was a dependence on calculations to show higher level demand. However, many calculations at this level were relatively routine and there was a need for other types of task, for example, explanations, to ensure demand was adequate.

At foundation tier the following were identified as issues:

- there was a shortage of readily accessible tasks targeted at grades F and G
- there was limited assessment of relevant mathematical skills and ideas of scale and measurement
- the assessment of ideas and evidence in science and the quality of written communication was not always adequate.

At higher tier the following were identified as issues:

© 2008 Qualifications and Curriculum Authority

- there was insufficient challenge for the top grades, particularly in biology and the biology assessment in double award
- the use of common questions across the separate sciences and the double award, as well as between foundation and higher tiers, helped to ensure that standards were consistent, however, there was an unfortunate side-effect that assessment in double award lacked most of the more interesting and relevant applied science.

At AS level, the question style of the subjects remained very different and in biology it changed considerably between units. However, the bulk of the marks were allocated either to very directed questions that required a word or short phrase answer, or to questions requiring extended writing, worth from four marks up to 10 or 11. However, these extended writing questions were not as demanding as might be expected, as many involved little more than recall. There was no requirement for candidates to provide their own expression of complex ideas and relationships. Several questions in biology at AS were similar those at GCSE higher tier, with the difference in demand reflected only in the mark scheme.

In physics the two most common question styles were calculations and short explanations. The shortage of extended writing meant that some skills could not be shown, but the tasks set were at an appropriate level. However, at AS the calculations were significantly more demanding than at higher tier GCSE and required a range of higher skills.

All examination papers set a large number of questions, involving a variety of question types. Reviewers judged AS chemistry to be the most demanding because of a high requirement for recall and extended writing. The difference between the subjects continued at A2 with the extra demand coming, to a greater or lesser extent, from the synthesis of abstract concepts.

Reviewers found the time allocation for some question papers was insufficient. In particular, they judged that the one hour available for the AS chemistry units was not sufficient. They acknowledged that it was almost certainly possible for most candidates to answer the structured questions in 45 minutes (as recommended) but they felt that to answer question 5, which in each paper required considerable time and thought to recall and organise ideas, required much more than the 15 minutes allowed. It would be very difficult for candidates to produce their best work in response to this type of question in such a short time. Similar points applied to chemistry at A2. The length of the physics AS topics test, at 30 minutes, was also considered demanding as it hardly gave a candidate time to overcome exam nerves.

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Archived Content

Reviewers judged the physics A2 synoptic unit challenging for candidates, as it was relatively long at two hours and of a style with which they would be unfamiliar, requiring extended writing. Furthermore, it was worth 40 per cent of their A2.

Reviewers identified an issue related to the level of the questions matching the demands of the syllabus. For example, there was very little extended writing in physics, which meant that the quality of written communication and some higher skills could not be easily assessed. In addition, there was little detail with regard to the assessment of the quality of written communication across all the sciences, so that the standards applied were very uncertain. At the same time, extended writing could pose its own problems in terms of time management. At A2, for example, the biology paper included a comprehension passage to read and absorb and there was more demand in terms of extended prose compared to the other subjects. There was a major difference in the mark schemes for extended writing. In chemistry, for example, there were 15 possible points for 15 marks, whereas in biology a candidate could score 15 marks for any 15 out of 20 possible marking points. This reflected the writing tasks in biology which were far more open-ended so that a candidate could deal with a question on a particular topic in breadth or depth.

Overall, reviewers identified the following issues when considering assessment materials:

- there were differences in demand between optional routes
- in terms of its assessment style, chemistry was the most demanding at every level
- there was inconsistency in the use of extended writing between and within subjects
- demand relating to mathematical work was variable – quantitative thinking was hardly evident in biology, whereas in physics complex mathematical processes were repeated unnecessarily and used in place of writing and synthesis to increase demand
- different approaches to synoptic assessment affected demand.

4.5 Outcomes of CRAS analysis

Reviewers also carried out a CRAS analysis of the question papers. The results are summarised in Table 3.

Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Table 3 The average CRAS overall ratings for each subject

Subject	Biology	Chemistry	Physics	Double award
Foundation	2.5	2.6	2.9	2.5
Higher	4.1	4.5	4.5	4.3
AS	6.7	6.7	6.8	
A2	8.2	8.4	8.2	

Table 3 shows that all subjects were similar in demand at each level, with chemistry and physics perhaps slightly more demanding than biology and, at GCSE, double award. That they were very similar is not a surprise since all three subjects adopt a very similar approach to assessment, with the majority of questions being short-answer, often within a series of structured questions.

These overall ratings conceal the extent to which there were variations by individual CRAS analysis factor. In particular, the analysis of chemistry at all levels generally gave high ratings for *abstractness*, often with *complexity* and *strategy* lower. Where *strategy* was high, the questions required considerable organisation and the selection of information, for example, from a data sheet. Candidates generally had to generate their own strategies when doing calculation questions.

Resource requirements for chemistry at AS and A2 were high compared to the other two sciences. This is reflected in the comments above about the high amount of recall required in chemistry.

4.6 Coursework

At GCSE the demands of coursework were broadly equivalent across all the subjects, though some reviewers found that double award science was more demanding than the separate sciences because there were two marks per skill area. It was also suggested that the coursework requirements were poorly matched to biological investigations (with so many uncontrollable variables, the difficulty of finding quantitative opportunities, the time over which data must be collected and the need for controls).

At AS and A2, the coursework criteria were natural developments of those at GCSE. Reviewers judged their demand to be about right in biology and chemistry. In physics,[4] however, the practical test offered a very narrow interpretation of the assessment objective. In

[4] This was a particular feature of the syllabus used in the study. In other syllabuses, practical skills can be assessed through coursework, rather than through a compulsory practical examination.

© 2008 Qualifications and Curriculum Authority

Archived Content

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Archived Content

addition very little was required from candidates in terms of devising strategies in the examination; rather, instructions were given which candidates had to follow step-by-step. The complexity of the examination was also much less than the complexity of completing coursework tasks. Overall, the level of demand for physics in this area was lower than it should be.

4.7 Progression

Overall, reviewers found that progression from GCSE to AS and from AS to A2 was not equivalent across the sciences and that this may have been linked to the variable impact of the assessment objectives across the subjects. In biology there was a major leap in conceptual demand between GCSE and AS, and less into A2, whereas in physics, the effect of the optional routes at AS meant that there was less of a step from GCSE to AS, but a much steeper incline of difficulty from AS to A2.

Reviewers judged that the step from GCSE double award science to all three separate sciences at A level was particularly large.

4.7 Overall

- At each level, chemistry was seen as the most demanding in terms of content.
- In all subjects at GCSE there were oddities in the way content had been divided up between the tiers.
- Reviewers expressed reservations about the usefulness of some foundation tier subject content. Reviewers did not consider some content to be relevant to those not going on to further study.
- Reviewers considered that successful progression from AS to A2 depended on knowledge of other sciences and, to a varying degree, of mathematics.
- Options within subjects affected the demand. Some options were conceptually harder and/or tested in a more demanding fashion. The differences between subjects in terms of whether they offered options at AS or A2 affected both comparability between subjects and progression within a subject.

Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

- At A level the approach to the assessment of the ability to synthesise knowledge, understanding and skills (assessment objective 4) varied between the three sciences, with chemistry judged to be less demanding than physics and biology in this respect.

- There was variation in the range of task types used across the subjects at both GCSE and at A level, with biology tending to require more extended writing and physics using more questions requiring short answers or calculations. However, the range of task types within the subjects tended to be limited and this was particularly the case for chemistry, which tended to use questions requiring short answers, with a high amount of recall.

- Progression from GCSE higher tier to AS, and from AS to A2 was not comparable across the subjects.

- There was inconsistency in the requirement for extended writing between and within subjects.

It is clear that awarding bodies, working with the regulatory bodies, can address a number of these issues through syllabus review, and guidance on question writing and question paper construction. Syllabus review is likely to be the first step in order to generate new syllabuses that recognise the above issues and attempt to do something about them.

Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

5 Findings of the review of candidate work

5.1 Materials and methodology

Given the exploratory nature of this work, lessons were inevitably learnt from previous QCA inter-subject comparability studies and changes made to methodology used. This meant that the nature of the materials reviewed and the kind of analyses possible were different from the other three inter-subject comparability studies (studies 1a, 2a and 2b) and perhaps slightly more reliable.

For this study, candidates' work was selected to cover a significant range of attainment within each level of each subject. The range was deliberately designed not to centre on any of the key judgemental grade boundaries, so that reviewers would not be distracted by questions of whether they felt candidates had been correctly graded.[5] For each subject and level, the awarding bodies supplied the compete work of candidates in the externally assessed components of the examination. Reviewers did not consider coursework as part of this exercise.

The work was then divided into three overlapping sets: high, medium and low. These were overlapping to allow some level of evaluation as to how reliable it might be to extrapolate or interpolate findings to specific points on the mark range. Reviewers looked at parallel sets of work (two medium sets) from two of the subjects involved in the work. They made a series of judgements about the merits of the work of various individual pairs of candidates within each set. Each reviewer considered work from candidates in the same two single sciences for which they had reviewed examination materials in the first part of the study. All reviewers also made comparisons with double award science candidates, comparing such work with the single sciences that they were also comparing.

As far as possible, therefore, the study was designed to reduce the possibility of systematic bias. In discussions with reviewers, it was clear that they also recognised the possibility of bias. Interestingly, the general view was that if their judgements were biased, it would be in favour of the subject with which they were less familiar.

Nevertheless, it is important to remember when evaluating these results that the task the reviewers were being asked to perform was highly demanding, much more so than that involved in awarding grades in the first place. Although they were not asked to decide what

[5] The key grade boundaries at GCSE higher tier are grades A and C and at foundation tier C and F. At both AS and A2, they are the A and E boundaries.

© 2008 Qualifications and Curriculum Authority

grades candidates deserved, only to decide if they had performed better or worse than other candidates, they had to make these decisions with a much less defined idea of what represented good performance and across a wider range of material than is usually the case. They were also not given any of the background technical and statistical information that is routinely supplied at awarding meetings.

It is also important to remember that examining is not in itself exact. There is unreliability that arises at several stages of the process. First the initial marking may not have been absolutely accurate, meaning that the actual marks the candidates gained did not perfectly represent their attainment on the papers. Second, the awarding process – and particularly that part of it involving judgement of candidates' work – is not absolutely precise. Those involved in the process are not called on to do more than establish a zone of uncertainty within which the boundary will lie, only determining their final recommended mark in the light of a wide range of other evidence. Moreover, even with the whole range of evidence taken into account, it may still be virtually impossible to choose between two possible marks for a boundary. The decision affects potentially large numbers of candidates, but that does not make the choice of mark any easier, still less more reliable.

For the purposes of the study, it was important that the marks were placed on a common scale. At A level, this was relatively straightforward and defensible. Because of the modular structure of those examinations, raw marks on individual units are already converted to a uniform mark scale that bears a constant relationship to the grading scale, regardless of subject and unit. So the marks used in the analysis of the AS and A2 units were uniform marks, converted into percentages.

At GCSE, however, no such system was used in the syllabuses selected for this study. There is a uniform mark scheme in operation for modular GCSE examinations and the marks were converted to marks on this scale, again expressed as a percentage. It is possible that this conversion may have distorted the raw mark relationships between the various candidates.

5.2 Outcomes at GCSE foundation tier

At foundation tier, the range of marks extended roughly from grade F to grade D; and in terms of the uniform mark scale from 30 to 50 per cent of the marks. Table 4 shows the relationship between the four subjects at 30, 40 and 50 per cent. The table indicates the uniform mark, centred on the relevant point on the mark range, where the four subjects were equivalent.

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

Table 4 The relationship between the four subjects at 30, 40 and 50 per cent

Subject	Subject mark at 30%	Subject mark at 40%	Subject mark at 50%
Biology	29.35	38.86	48.36
Chemistry	32.69	41.63	50.58
Physics	26.19	38.99	51.79
Double award	30.74	39.93	49.12

From Table 4 it can be seen that, although there were differences between the subjects, these were small. They almost certainly fall within the limits of accuracy in the awarding process and certainly within those of this exercise.

5.3 Outcomes at GCSE higher tier

At higher tier the range of marks was roughly from grade C to grade A, and in terms of the uniform mark scale from 60 to 80 per cent of the marks. Table 5 shows the uniform mark, centred on the relevant point on the mark range, where the four subjects were equivalent.

Table 5 The relationship between the four subjects at 60, 70 and 80 per cent

Subject	Subject mark at 60%	Subject mark at 70%	Subject mark at 80%
Biology	62.85	71.48	80.11
Chemistry	57.81	67.81	77.80
Physics	59.21	69.78	80.35
Double award	62.25	71.55	80.86

Table 5 shows that at higher tier the four subjects were also very well aligned. Notably, the outcomes were marginally more consistent with the findings of the review of examination materials, in that chemistry candidates were consistently judged to be slightly better than those in the other subjects at the same uniform mark, and those from double award science (and to a smaller extent biology) were judged to be slightly weaker.

5.4 Outcomes at AS

At AS the work used ranged roughly from grade D to grade B, and in terms of the uniform mark scale from 50 to 70 per cent of the marks. Table 6 shows the information for same performance at 50, 60 and 70 per cent for the three sciences considered.

© 2008 Qualifications and Curriculum Authority

Table 6 The relationship between the three subjects at 50, 60 and 70 per cent

Subject	Subject mark at 50%	Subject mark at 60%	Subject mark at 70%
Biology	49.18	59.13	69.09
Chemistry	48.83	59.09	69.35
Physics	51.50	61.73	71.96

Table 6 shows that standards at AS were very well aligned. There was a consistent picture, with biology and chemistry almost indistinguishable and physics very slightly generous. The differences were very small and well within the limits of error.

5.5 Outcomes at A2

At A2 the work used ranged roughly from grade D to grade B, and in terms of the uniform mark scale from 50 to 70 per cent of the marks. Table 7 shows the information for same performance at 50, 60 and 70 per cent for the three sciences considered.

Table 7 The relationship between the three subjects at 50, 60 and 70 per cent

Subject	Subject mark at 50%	Subject mark at 60%	Subject mark at 70%
Biology	51.42	62.86	74.30
Chemistry	49.57	57.64	65.70
Physics	50.84	60.12	69.40

From Table 7 it can be seen that the rank order of the three subjects remained consistent throughout the mark range. Moreover, the gap between biology and chemistry increased through the mark range, with chemistry noticeably more severe than biology at 70 per cent. The mark for physics remained in between the two throughout.

Archived Content

Archived Content
This document is for reference only. It may have been discontinued or superseded.
Archived Content

© 2008 Qualifications and Curriculum Authority

Archived Content

This document is for reference only. It may have been discontinued or superseded.

Archived Content

Appendix A: Reviewers

Judy Bate
Peter Chandler
Chris Hickman
Philippa Hulme
Giles Job (lead consultant, physics)
David Kelly
Alan McMurdo (lead consultant, biology)
Geoff Mines (lead consultant, chemistry)
Graham Riley
John Skevington
David Smith
Mike Solomon
Paul Spencer

Archived Content

© 2008 Qualifications and Curriculum Authority