

AWARDING EXAMINATION GRADES: CURRENT PROCESSES AND THEIR EVOLUTION

Colin Robinson

Abstract

The aim of this chapter is to describe the process by which each awarding body tries to ensure that the grading of candidates is comparable, no matter when, by whom, or on what aspects of the subject the candidate is assessed. These processes have developed over the years, mainly as a result of the experiences of the awarding bodies themselves, but also because of the impact of regulation, which has grown over time. The chapter describes this key activity of the awarding bodies, looking at the personnel involved, the information that is available to them and the decisions they have to make in order to ensure that the results maintain the multifaceted comparability requirements laid upon the awarding bodies. Changes that have been made over the years and the reasons for them are also outlined. The April 2007 edition of the code of practice lays out the comparability requirements quite clearly:

The awarding body's governing council is responsible for setting in place appropriate procedures to ensure that standards are maintained in each subject examined from year to year (including ensuring that standards between GCE and GCE in applied subjects, as well as between GCSE and GCSE in vocational subjects, are aligned), across different [syllabuses] within a qualification and with other awarding bodies.

QCA (2007)

1 Awarding grades

Necessary conditions for comparability involve choosing appropriate subject matter to be embodied in the syllabus, selecting personnel with appropriate background knowledge and experience, developing question papers that give candidates the opportunity to demonstrate their knowledge, preparing marking schemes that give fair rewards and appointing markers who are reliable. But accurate marking is only part of the process. No matter how well the markers follow the mark scheme, the marks this session will have been obtained in different ways from last session and there remains the issue of how to relate the marks to the grades.

The purpose of grading is to group candidates' results in ways that allow the user to compare performances. It is not essential to have a two-stage process: grades can be awarded directly without the intervention of marks. In higher education, for example, it is common for marks to be awarded that can be directly related to grades,

on the basis of judgements of quality. For example, a mark of 70 is a first-class mark. However, when examinations are marked and the marks are then converted into a limited number of grades, it is necessary to consider the relationship between the marks and the grades. Public examinations have very large numbers of students, so marking has to be carried out by large teams of markers. Research has shown that it is easier to train people to mark reliably to a mark scheme than to grade holistically (Black, 1962). In higher education, there are fewer students and often only one or two markers for a class, so standardisation is not as much of a problem.

Marks in GCSEs and A levels are dependent upon the specific questions asked and the marking scheme attached to the question paper. Because the question papers are different on every occasion, the marks cannot retain a common standard. The two-stage process allows real differences between question papers to be taken into account so that the grades are comparable.

The function of the marking process, including standardisation, reviewing and re-marking, is to judge the candidate's work against the marking scheme. Having done that, the assumption is that all candidates on the same mark in a component are equivalent in terms of their attainment (where a component may be one of a series of question papers or coursework that comprise the entire examination assessment). Throughout the marking process, the focus is on ensuring that the rank order of the candidates is as accurate as possible. Given this, the task of 'awarders' (members of the awarding committee) is to determine 'grade boundary' marks (the cut-off points) that indicate the same standards of achievement this year as they did last year. Grade boundary marks are identified separately for each component of an examination. The award of subject grades, for the examination as a whole, requires the identification of subject grade boundaries; although these are aggregated differently if the syllabus is modular, as explained below. Students' subject grades derive from their total subject mark, the aggregation of their component marks, rather than being dependent upon a specific profile of component grades.

2 The awarding committee

With the exception of the reviser and the scrutineer, the people involved in the meeting are those who were responsible for designing the question papers and supervising the marking: the Chair of Examiners is responsible for all syllabuses within a subject area and chairs all the awarding meetings, the Chief Examiner has overall responsibility for the particular syllabus and the Principal Examiners lead on individual components. They are supported by one or more awarding body officers who will be responsible for making sure the procedures are followed and for advising on the statistics. If a single awarding committee cannot deal with all syllabuses in the subject, there will be overlapping membership with other awarding committees, 'to consider the consistency of the recommendations in the light of standards applied in other [syllabuses]' (QCA, 2007, p. 35). Some of the committees, dealing with a large number of syllabuses, can be very big, with as many as twenty awarders, not counting awarding body officers and other participants. The greater the number of specifications in a subject, the bigger the awarding meeting, but the

average number of awarders across all subjects at all levels in the major awarding bodies is 8–10 (Baird & Dhillon, 2005).

Awarding meetings occupy most of the time between the end of the marking process and the publication of results, and consume a vast amount of resources. Extrapolating from figures in one awarding body (the Assessment and Qualifications Alliance), Baird & Dhillon (2005) estimated that across the country senior examiners spend a total of more than 3,400 person days each year in awarding meetings at a cost of more than £1.5 million in examiner time alone. In addition to the awarders themselves, each meeting is attended by one or more members of staff from the awarding body and will sometimes include observers from QCA, or from Ofsted, representatives from subject or teaching associations and even members of the press. The judgements are all made by the awarders who will be senior examiners with many years' experience. They are provided with information that helps them to ensure that the awards remain in line with those in other syllabuses and with other awarding bodies. Their job is to weigh all the evidence about the performance of candidates at each judgemental grade and come to a consensus as to the mark on the year's paper that represents the lowest level of achievement worthy of the award of that grade. The outcomes are then recommended to the Accountable Officer of the awarding body, who makes the final decision.

The evidence presented to the awarders is substantial as the following list of minimum requirements from the code of practice (QCA, 2007) indicates:

Qualitative

- i. copies of question papers/tasks and final mark schemes
- ii. reports from the principal examiner(s)/principal moderator(s) on how the question paper functioned
- iii. archive scripts and examples of internally assessed work (including, in appropriate subject areas, photographic or videotaped evidence) at the relevant grade boundaries, together with relevant question papers and mark schemes
- iv. samples of current candidates' work (marked scripts and/or internally assessed material) distributed evenly across key boundary ranges for each component, with enough representing each mark to provide a sound basis for judgement so far as the size of entry and nature of work permit. The material should be selected from a sufficient range of centres where work has been marked/moderated by examiners/moderators whose work is known to be reliable
- v. any published performance descriptions, grade descriptions and exemplar material, where available
- vi. any other supporting material (such as marking guides for components where the evidence is of an ephemeral nature)

Quantitative

- vii. technical information – including mark distributions relating to the question papers/tasks and individual questions for the current and previous series, where available
- viii. information on candidates' performance in at least two previous equivalent series, where available
- ix. details of significant changes in entry patterns and choices of options
- x. information on centres' estimated grades for all candidates including:
 - qualification-level estimates for linear (including linear unitised) [syllabuses]
 - unit-level estimates for externally assessed units in all other unitised [syllabuses]
- xi. information about the relationship between component/unit-level data and whole-subject performance, where available

Regulatory authority reports

- xii. relevant evidence from the regulatory authorities' monitoring and comparability reports.

How do they deal with all this information? In the following section, references to the above extract from the code of practice are shown in italics.

Of course, all examiners will have received *copies of question papers/tasks and final mark schemes* as part of their normal marking duties but the main purpose of them at the award is to help to decide if the paper is more or less difficult compared with previous ones. This can only be done by considering both the questions and the credit given. The awarders will need to take into account *reports from the principal examiners... on how the question paper functioned*. The Chief Examiner and Principal Examiners will have marked candidates' work directly themselves and second-marked samples of marking from those examiners they supervise. This is an essential part of the exercise, giving the senior examiners first-hand experience of the question paper in action.

Before the meeting can consider the grade boundaries for the current series of examinations, the awarders must familiarise themselves with the quality of work that characterises the grade they are looking at. This they will do by checking the standards upon which the grades were awarded previously, by looking at *archive scripts... at the relevant grade boundaries, together with relevant question papers and mark schemes*, typically from the previous year's examination. This evidence will be considered alongside the statistical information such as *mark distributions relating to the question papers and individual questions for the current and previous series*. From these, the awarders will be able to see how the component performed: what range of marks was obtained, how the average mark and spread of marks compared with last year, whether the mark distribution was skewed (with lots of candidates scoring either high or low marks and fewer candidates scoring marks in the middle of the range). The aim is to refresh their understanding of the quality of work that typified the

performance of candidates at the lowest level deemed to be worthy of the grade, bearing in mind the questions asked and what the examiners were crediting in the mark scheme. For A level syllabuses, they will also have copies of performance descriptions and exemplification of the standard of work at each of the boundaries scrutinised by awarders, prepared by QCA in the light of the Tomlinson Report (Tomlinson, 2002). For GCSE, they will have the 'mid-range' grade descriptions contained in the syllabus and/or the relevant GCSE subject criteria.

Between examination sessions, a number of investigations take place to identify possible issues. Examination teams within an awarding body may ask for research to be carried out to check on apparent problems of comparability or on the likely effects of any planned changes in the nature of the assessments. Awarding bodies conduct their own internal enquiries on examinations that appear to be out of line with performances in other awarding bodies; the regulatory bodies report on scrutiny exercises that have been carried out during the year, and the awarding bodies jointly may have undertaken comparability studies. Some analyses are conducted routinely, such as OCR's 'unit pairs' analyses, in which the performances of the same candidates are compared across different question papers. The results of all these exercises will be collated and fed into the awarding meeting's deliberations by the awarding body officer. In some cases there may be a specific requirement to ease or tighten standards in a syllabus, in order to bring the results into line with national standards.

Significant changes in entry patterns and choices of options can have a considerable effect on the proportions of candidates getting each grade without reflecting any real change in the standard of the examination. If able candidates (that is, those with higher prior achievements) move to another syllabus in large numbers, leaving the syllabus dominated by lower-attaining candidates, it is to be expected that smaller numbers of candidates will achieve higher grades. There is now a considerable amount of data which can be drawn upon – from earlier national assessments and also from other subjects in the same examination series. The awarding bodies use this information to provide the awarding committee with comparisons of *candidates' performance in at least two previous examination series*.

At A level, one way of doing this is by grouping candidates according to their overall performance across GCSE examinations and then looking at how those groups perform in a particular A level subject across the country as a whole. There is usually a strong correlation between these performances and it is possible, therefore, to use the relationship to predict the expected results of candidates for whom the awarding body has similar prior achievement data. This information can be useful both before and after the examinations. As Eason writes,

Not only can these methods be used to predict examination results, they can be used *post hoc* to identify whether there are inconsistencies between a particular awarding body's outcomes and the national pattern for a given subject.

Eason (2004)

In addition the awarding bodies collect information about *centres' estimated grades for all candidates*. These are a useful flag that can alert the awarding body to possible discrepancies in the awards, leading to a review of grade boundaries.

A large proportion of the time at the meeting will be spent looking at the *samples of current candidates' work (marked scripts or internally assessed material)*. On the basis of their experience of previous award decisions and their marking of the current year's scripts, the Principal Examiner and the awarding body officer will have agreed the range of marks within which they expect each judgemental boundary to lie – usually a spread of about six or seven marks – and will have arranged for sufficient scripts (booklets of candidates' work in the examination) to be available in the meeting room so that each awarder will be able to scrutinise scripts at each of the marks.

3 The conduct of the awarding meeting

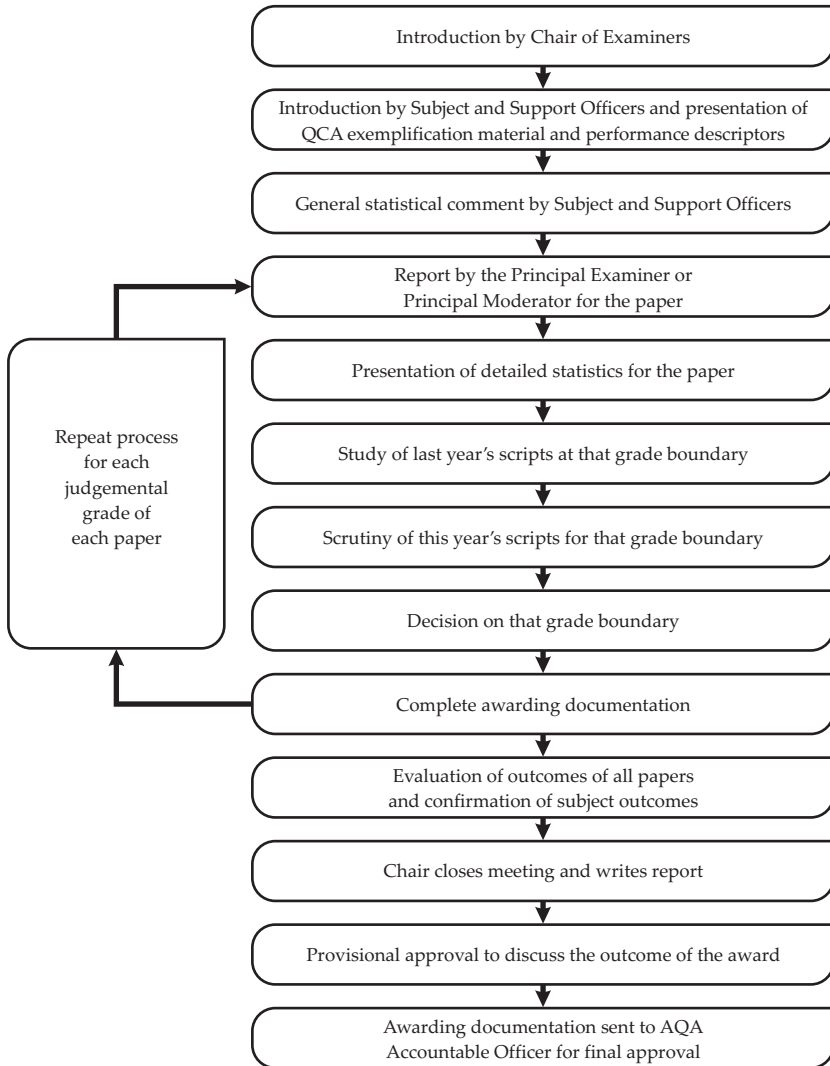
The format of the meeting follows broadly the same pattern across the awarding bodies (where minor variations exist, this chapter discusses the procedures used by AQA). Figure 1 presents an overview of the process. First, the Chair of Examiners opens the meeting by outlining the task ahead for the committee. The awarding body officer will then present the statistics and point out any implications that the committee needs to consider. Reports from the Principal Examiners for each component are given next, indicating how the paper performed, what questions were found difficult or easy by the candidates, and any issues arising from the marking. Unless they have done so in advance of the meeting, members of the committee will refresh their understanding of the standards by reference to the grade descriptions, the archive scripts and exemplar material.

Armed with the information provided by the awarding body officers and the Principal Examiners, the awarders will then consider the quality of the work in this year's examination scripts. The time available is limited and it is certainly not possible (even if it were desirable) for the awarders to read through every response. The exercise is one of determining whether the script as a whole merits the award of the grade under consideration. Marks and any explanatory comments are shown on the scripts.

Their individual judgements on each script – worthy of the higher grade, not worthy, or unsure – are then tallied for each mark in the range, starting at the highest mark and working down until they come to the point where there is reasonable consensus that all scripts at or above that mark are worthy of the higher grade. This mark is referred to as the Upper Limiting Mark. The committee then works up from the lowest mark until they reach consensus on a mark that is the highest that is not worthy of the grade. The mark above this is referred to as the Lower Limiting Mark.

In the example shown in Figure 2, the awarders agreed that all the scripts they had considered at a mark of 64 were not worthy of the grade and although there were a couple of scripts at 65 that were of an acceptable standard, following discussion it was agreed that by and large scripts with this mark were also below the grade.

Figure 1 Overview of the process followed in an awarding meeting (from AQA, 2005a, p.7)



(AQA material is reproduced by permission of the Assessment and Qualifications Alliance)

Working from the top, only two awarders had managed to review scripts with a mark of 70, but they were convinced of their standing. There were a couple of scripts at 69 that were thought to be below standard and only one at 68. The awarders therefore came to the conclusion that the Upper Limiting Mark was 68 and the Lower Limiting Mark was 66.

Figure 2 Example of a tick chart (from AQA, 2005a, p.6)

Statistically Recommended Boundary = 67
Principal Examiner's Recommendation = 67–70

Mark	Awarder 1	Awarder 2	Awarder 3	Awarder 4
70	✓			✓✓
69	✓	×✓✓	✓✓✓	×✓
68	✓✓	✓✓	✓✓✓×	✓✓
67	✓✓×✓	×✓✓	✓✓	✓×?
66	✓✓	✓?	✓✓×	××
65	✓×	××	×	✓×
64	××	×	×	×

Upper limiting mark: marks of 68 and above are definitely worth the grade
 Lower limiting mark: marks of 65 and below are definitely not worth the grade

✓ script 'worthy' of the grade in question
 × script 'not worthy' of the grade in question
 ? awarder unsure of whether the script is worthy of the grade in question

(AQA material is reproduced by permission of the Assessment and Qualifications Alliance)

Having established these limits, the awarders seek to agree a single mark that represents the boundary for the grade, using their collective professional judgement. They also consider statistics – the proportion of candidates that would get the grade if it was awarded on 66 or 67 marks – and they may also consider further scripts. If the marks on individual questions were recorded, the item-level statistics can provide additional pointers.

Only certain grade boundaries are determined through this judgemental process. Typically, these are the A/B, C/D and F/G boundaries (for GCSE), and the A/B and E/U boundaries (at A level).

4 Setting subject boundaries for linear syllabuses

Having set all of the component boundaries, the awarding committee will need to consider the appropriate grade boundaries for the subject as a whole. For linear syllabuses – those, as for most GCSEs, whose examinations are all taken at the end of a course of study – two indicators are calculated: one based upon marks (the addition method), the other on percentages of candidates (the percentile method). As the code of practice explains, the lower indicator is chosen.

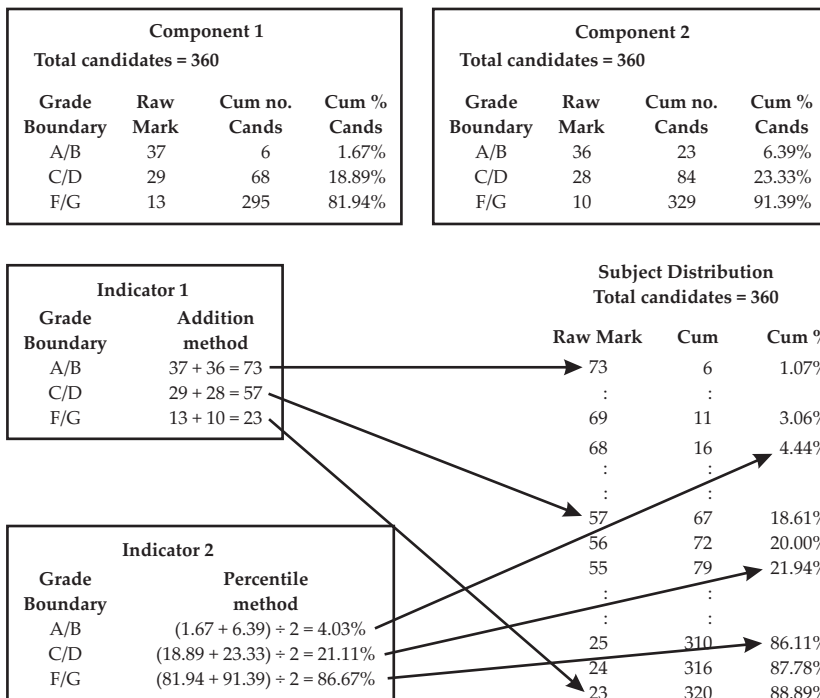
The obvious way of doing this is to add together the grade boundary marks for each component: the addition method. So, for example, if the boundaries for the grade are at 37 on one component and 36 on another, the total would be 73 (if they are equally weighted). Yet this can lead to apparent anomalies. As is shown in Figure 3, the addition method – Indicator 1 – would lead to only 1.07% of the candidates being awarded grade A overall, which might seem a bit harsh (1.67% of students were awarded grade A on component 1, 6.39% were awarded grade A on component 2). At the other end of the range, it would lead to 88.89% of candidates being awarded grade F or better, which might, if anything, seem a bit generous (although 91.39% of students were awarded grade F or better on component 2, only 81.94% of students

were awarded grade F or better on component 1). Clearly, different proportions of candidates might be thought to deserve the grades for different question papers because the awarders may consider performances on paper 2 to be better than those on paper 1, for example.

Having said that 1.67% deserve a grade A on paper 1 and 6.39% deserve a grade A on paper 2, why is the proportion getting the subject grade less than either of those, at 1.07%, using the addition method? The reason is that it is not *the same* candidates who were thought worthy of a grade A on each question paper. Candidates who scored over 37 on component 1 were likely to score less than 36 on component 2, so scored less than 73 overall. In the early 1990s, the method based on percentages was introduced to combat this effect because of the perceived unfairness to candidates.

Figure 3 Effects of the addition and percentile methods of boundary score aggregation (from AQA, 2005b)

Example: Applying the indicators to two equally weighted components



(AQA material is reproduced by permission of the Assessment and Qualifications Alliance)

In the percentile method (Indicator 2), the weighted average of percentages of candidates considered to be worthy of the grades is calculated. Going back to the previous example (see Figure 3), the average of 1.67% and 6.39% is 4.03%. There is no subject mark with exactly 4.03% of candidates attaining that mark, so the nearest percentage to 4.03% has to be found. That is 4.44%, on a mark of 68.

The code of practice (QCA, 2007) explains how the indicators are calculated and how to choose which one represents the subject boundary mark.

Indicator 1

- i. The boundary mark for each component is scaled as necessary to reflect the mark allocation for that component as detailed in the [syllabus].
- ii. The resulting scaled component boundary marks are added up and the result is rounded to the nearest whole number (0.5 rounded up).

Indicator 2

- iii. The percentage of candidates at and above the boundary mark on each component is determined.
- iv. A weighted mean of these percentages is calculated, using the weightings of the components as detailed in the [syllabus].
- v. The mark is identified on the distribution of total examination marks at which the cumulative percentage of candidates most closely corresponds to the weighted mean.

Conclusion

- vi. Whenever the two indicators do not coincide, the grade boundary should normally be set at the lower of the two indicator marks, unless, in the awarders' judgement, there is good reason, as a result of a review of the statistical and technical evidence, to choose a higher mark within the range spanned by the indicators.

Only judgementally determined boundaries lead to the calculation of Indicator 1 and Indicator 2. The remaining subject grade boundaries are determined arithmetically; by creating bands of equivalent width (as equivalent as possible) according to procedures specified in the code of practice (QCA, 2007).

5 Setting subject boundaries for modular syllabuses

The introduction of modular syllabuses has led to a completely new approach to the aggregation of component boundary marks. The rationale of modular syllabuses is that candidates are assessed after they have completed each module – more commonly referred to as a 'unit'. In the current GCE A level, there are typically six units to be taken over two years. Three of these (the AS units) are designed to be taken in the first year of the course and can lead to the Advanced Subsidiary award. These AS units also count towards the full A level, when combined with three further units (the A2 units), designed to be taken in the second year. Candidates taking the same unit on different occasions will not have been assessed on exactly the same

materials and it would be impossible to determine subject grade boundaries (from component ones) in the same way as for linear syllabuses.

We need to be able to save the details of each candidate's attainment for use when all units have been completed and the overall subject grade needs to be calculated. We could, of course, just 'bank' the grades achieved by the candidate in each unit, and then aggregate unit grades, but that would suffer a number of problems – candidates who gained a very high mark within the grade would be indistinguishable from those who just scraped through and when we aggregated across the units we would end up with a distribution that showed less discrimination between the candidates. What we need is a mechanism that records the candidate's attainment in a form that maintains the fine discrimination of the marks but that also recognises the 'grade-worthiness' of marks in different units.

The mechanism for this is the Uniform Mark Scale (UMS). In the UMS, the scaled marks for each grade are pre-determined and depend only on the weighting given to the unit. Table 1 gives the UMS for units of different weight and shows how they contribute to the overall subject awards. Thus, for an A level which comprises six units, an individual unit might well have a weighting of 16.7%; this would mean that grade boundaries on its raw mark scale would be converted to grade boundaries on the 100-mark UMS.

Table 1 Grade boundaries in terms of uniform marks according to weighting of unit (from AQA, 2006)

Weighting as % of total AS assessment	30%	33.3%	35%	40%	100%	AS subject award	Advanced subject award
Weighting as % of total Advanced assessment	15%	16.7%	17.5%	20%	50%		
Max uniform mark	90	100	105	120	300	300	600
A	72	80	84	96	240	240	480
B	63	70	74	84	210	210	420
C	54	60	63	72	180	180	360
D	45	50	53	60	150	150	300
E	36	40	42	48	120	120	240
(N)	27	30	32	36	90	–	–

The method for linking the unit raw mark grade boundaries to the UMS and the aggregation procedure is outlined in Box 1 (adapted from AQA, 2006).

Despite the reference to marks in its name, the UMS marks are really a form of grade. Just like grades, the UMS marks have a common interpretation irrespective of the

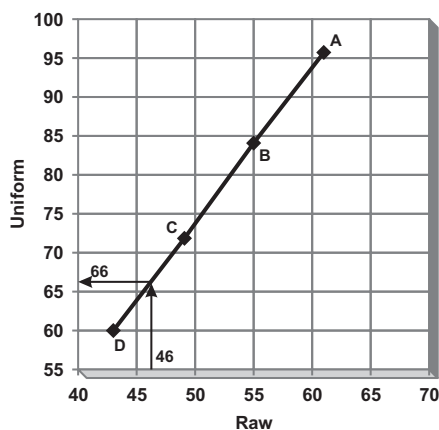
Box 1

This example is based on Curriculum 2000 A levels, which had a total maximum UMS of 600. The table below shows typical grade boundaries for a GCE unit which is marked out of 80 raw marks and has 20% weighting. For this unit, the total maximum UMS is 120 (20% of 600). The second column shows the raw mark boundaries. The grade A boundary is 61 (approximately 76% of the raw mark total). The third column shows the uniform mark boundaries. For a GCE unit with 20% weighting, the maximum uniform mark is 120 and uniform marks in the range 96–120 correspond to grade A. This does not mean that the paper is marked out of 120 or that a candidate has to score 80% of the raw marks (96/120) to obtain grade A on the unit. A candidate who scores 61 (the lowest raw mark for grade A) will receive a uniform mark of 96 (the lowest *uniform* mark for grade A). Similarly, a candidate who scores 43 will receive a uniform mark of 60 and a candidate who scores 49 will receive a uniform mark of 72. A raw mark between 43 and 49 corresponds to a uniform mark between 60 and 72; for example, a raw mark of 46 (exactly half way between 43 and 49) corresponds to a uniform mark of 66. This is illustrated in the figure below.

GCE unit with maximum raw mark of 80 and accounting for 20% of the A level assessment: typical raw mark grade boundaries, together with the uniform mark boundaries

Grade	Lowest raw mark in grade (maximum 80)	Corresponding uniform mark (maximum 120)
A	61	96
B	55	84
C	49	72
D	43	60
E	37	48
(N)	31	36

Conversion to uniform marks (for part of the mark range) for the above data



underlying raw marks of individual units. The advantage is that the UMS marks retain much finer distinctions than the grades. Once candidates have taken all of the units, their UMS marks are aggregated and their grade is evident from their final score. As Table 1 shows, a UMS of 480 means that the candidate has attained a grade A at A level.

6 Making the final decision

The balance between the examiners' qualitative assessment of standards and the statistical information is a delicate one. In large-entry subjects it may be argued that the performance of the candidature as a whole is unlikely to change substantially from one year to the next unless there are changes in the entry pattern (for example, more students with high prior attainment scores) or changes within the examination itself (for example, more marks available in a particular component) that would explain the difference. Such changes may be taken into account in a 'statistically recommended boundary' for a judgemental grade, which is based on the quantitative evidence available for the question paper under discussion and calculated as the mark which maintains standards from last year as closely as possible, statistically speaking (AQA, 2005a, p. 9). The statistically recommended boundary takes into account the subject-level standards, not just those at component level, as it would be unwise to wait until the end of the process to find out what the judgements on each component did to the grading of the examination overall. Analyses of the types of candidate entering the examination in each year are carried out prior to the meeting, and if the entry type has changed between years the statistically recommended boundaries can be modified to reflect that fact (Baird, 2000). Within AQA, if the awarding committee wishes to place a subject grade boundary at a mark that departs substantially from the statistically recommended boundary, it is required to document the reasons for its recommendation.

The recommendations of the awarding committee are referred to the Accountable Officer within the awarding body, upon whom the code of practice places responsibility for setting the final grade boundary, 'to ensure that grades awarded represent continuity and parity of standards across years, over time and across [syllabuses]' (QCA, 2007, p. 38). The Accountable Officer is appointed by the awarding body's governing council and 'is accountable directly to its governing council for ensuring the quality and standards of its qualifications' (QCA, 2007, p. 6). For every syllabus the awarding body offers, the Accountable Officer must approve the grade boundaries.

If the Accountable Officer does not believe the grade boundaries recommended by the awarding committee to be appropriate, it is the responsibility of the Accountable Officer to change them. In such a case, the Accountable Officer would discuss changes to the boundaries with the Chair of Examiners. If the changes were outside the range of marks considered at the meeting, it would be necessary to recall the awarding committee. The Accountable Officer, having received recommendations from all of the committees, may make a change to boundaries to bring consistency across the decisions being taken in, for example, new qualifications.

7 Evolution of the awarding process

7.1 School Certificate

The processes of awarding have evolved over the last century, though the basic principle – that of converting the rank order of the marks into a smaller number of grades whose standards are comparable across years and across syllabuses – has remained the same since the first local examinations in 1858. In the words of Circular 849 issued by the Board of Education in July 1914, the standard of a pass in the School Certificate ‘will be such as may be expected of pupils of reasonable industry and ordinary intelligence in an efficient Secondary School’ (quoted in Board of Education, 1943, p. 27). Subjects were grouped into three sets, (i) English subjects, (ii) Languages and (iii) Science and Mathematics and candidates were ‘expected to show a reasonable amount of attainment in each of these groups...’ (ibid.).

The introduction of the School Certificate (SC) and Higher School Certificate (HSC) as the first truly national qualifications brought pressure on the examining boards to ensure similar standards, even though the Ministry had no power to enforce agreement and there was much dispute. The HSC was the basis of both the award of State Bursaries and the matriculation requirements for many universities. Any divergence of standards would therefore result in injustice to some candidates. Accordingly the Secretaries of the examining boards met regularly and exchanged views on issues of comparability, agreeing amongst themselves what adjustments needed to be made in order to bring the different boards into line. Nevertheless, the Norwood Committee was concerned that the standards were not necessarily the same: ‘...even under present conditions two apparently similar certificates mean very different things, and illusory uniformity can be bought too dearly’ (ibid.).

7.2 General Certificate of Education

When the General Certificate of Education (GCE) was introduced in 1951, the examining boards had to devise new procedures to ensure that the standards of the old SC and HSC were carried forward into the new Ordinary (O) and Advanced (A) level respectively. The Ministry of Education had laid down, through the Secondary School Examinations Council (SSEC), the basic form of the certificates: both O level and A level were to be pass or fail with no grading.

7.3 Ordinary level

The O level pass would be continuing the standards of the Credit in the School Certificate. This was to compensate for the fact that candidates were no longer required to pass all their subjects before a certificate could be issued. The practicalities of how this was to be achieved were left to the boards and they devised their own processes. UCLES (1950), for example, placed great emphasis on the value of school estimates.

Although prohibited from indicating grades on the formal certificates, from the start some released information about the marks obtained; others developed informal grading schemes that were released to schools and to universities. The approach

proposed by UCLES was based upon percentages of the raw marks available, as Table 2 shows. This process would now be deemed unsatisfactory, as it would not maintain standards from year to year. Even if the *same candidates* took two years' examination papers, different proportions of them would be awarded a pass under this system because, as described earlier, a raw mark from one year is not equivalent to a raw mark in another year. That is the whole reason for having an awarding process.

Table 2 Rankings of candidates (from UCLES, 1951)

Pass	Fail
1 – 82%	6 – 40%
2 – 75%	7 – 35%
3 – 64%	8 – 30%
4 – 54%	
5 – 45%	

On the other hand, the Joint Matriculation Board (JMB) approach related grades separately to the proportions of candidates above and below the pass mark (Table 3).

Table 3 JMB grading guidance (from Forrest, 1970)

For candidates to be awarded passing grades	
Grade	Cumulative percentages of candidates
1	9
2	18
3	45
4	63
5	85
6	100
For candidates to be awarded failing grades	
7	30
8	70
9	100

7.4 Advanced level

The change from the Higher School Certificate to the Advanced Level was less of a problem, with the standard for the new examination being that of the pass in a principal subject in the HSC. The rationale for the grouping of subjects in the HSC had been almost exactly the opposite of that of the SC. Whereas in the SC the aim was to ensure that candidates covered a wide range of different subject areas, the aim

of the HSC groups was to enforce specialisation. However, it swiftly fell into disuse and awarding bodies found ways around the regulations on grouping.

The relationship of the SSEC to the boards remained advisory throughout its existence, with the result that most changes were instituted by the boards themselves rather than as a response to the SSEC. The only major area over which it held sway was what was published on the actual certificate issued to successful candidates partly because, to add to their credibility, they were counter-signed by an official on behalf of the Ministry of Education. It exercised this oversight with considerable caution, with changes to what was recorded on the certificate happening very slowly.

In the years following its introduction, the A level moved quickly from a pass or fail approach to one in which degrees of achievement were recognised: the distinction grade was re-introduced in 1953, at the same level as the distinction in the HSC. This satisfied the immediate need to distinguish candidates for state scholarships and bursaries, but it was not enough for the selection purposes of the universities. In that same year the JMB proposed that the A level should be reported in terms of grades. It took the best part of a decade before the SSEC put forward guidelines for the grading of the GCE A level (SSEC, 1960), which were adopted by all the boards.

In spring 1960, Professor Oliver, Professor of Education at Manchester and a member of the JMB, gave an address to the Association of Assistant Mistresses in which he argued not only that the GCE was providing insufficient information for university selection but also that 'the harmful effects of a defective examination on the work of the schools and the education of adolescents are positive evils which call for changes in the G.C.E irrespective of its use for selection' (Oliver, 1960, pp. 3–4).

In the same year, the SSEC issued a report that looked at the tensions between the use of the GCE A level as a selection instrument for university entrance and its use as a certification of attainment at the end of the sixth form course (SSEC, 1960). The report concluded that,

The most promising reform would be the introduction of a system of grading of passes at the Advanced level, with supplementary grading for those in the higher ability ranges, effected by means of a special or 'S' paper available with the advanced level papers. Grading should be recognised by the Minister and the results entered on the candidate's certificate as well as being circulated to universities and other users.

SSEC (1960)

The recommendations for grading the 'basic' A level papers were couched in terms of approximate proportions of the candidature: 10% grade A, 15% grade B, 10% grade C, 15% grade D and 20% grade E, with a further 20% being allowed an O level pass. The unequal divisions (and in particular the narrow grade C band), and the fact that the grades were described in terms of the candidature, was to store up problems for the future.

The report emphasised that these proportions were guidelines and the awarding bodies did not stick rigidly to them; for example, there were, and still are, large differences in

grades awarded between subjects. Nevertheless, the overall proportion of A level candidates passing remained around the 70% mark from its inception until well into the 1980s. At the same time, the report called for a greater degree of uniformity:

the Examining Boards should all agree both on a uniform set of symbols for presenting candidates' results, on both basic and S papers, and, so far as is possible, on similar methods of assessment in arriving at those results, so that universities and other users can feel some confidence that they are comparing like with like.

SSEC (1960)

8 A variety of grading systems

The advent of the Certificate of Secondary Education (CSE) in 1965 was a clear challenge to the status quo. It was from the outset a graded examination: four passing grades with two grades defined by reference to 'the calibre of candidates',

The definition of standards will therefore be as follows:

In each subject the examination should be so constructed that:

- (i) a 16-year-old pupil of average ability who has applied himself to a course of study regarded by teachers of the subject as appropriate to his age, ability and aptitude, may reasonably expect to secure grade 4; and
- (ii) a 16-year-old pupil whose ability is such that he may reasonably have secured a pass in the 'O' level of the GCE examination, had he applied himself to a course of study for that examination, may reasonably expect to secure grade 1, having followed a course of study defined in (i) above.

SSEC (1963a)

In addition to the four passing grades a fifth grade was agreed for those who had been correctly entered but whose performance was 'not quite up to the standard of grade 4'. Candidates whose performances were below grade 5 were ungraded – a term that was to be used in preference to 'fail' in the examinations that followed.

The SSEC recognised the problems of comparability between boards, different modes of examination and between the two examination systems, but it stood firm against any heavy involvement.

The difficulty of this range of problems is increased by the fact that no solution can be regarded as acceptable if it involves detailed intervention by the council in the conduct of the boards' examinations, formalising pressures by the boards on the work of the schools, or alignment of syllabus content and methods of examining as between CSE and GCE examination systems. In short, the job must be done without interfering in any way with the natural development of the work of schools, or the independent development of the two examination systems.

SSEC (1963b)

In terms of ensuring comparability between GCE O level and CSE examinations, the boards faced three major problems. As indicated above, they were precluded from changing their syllabuses to make it possible for candidates to follow a common

course leading to both examinations. The new CSE boards were regional and there was no overlap, so a candidate in one area could not take the examinations of a different regional board. Finally, the boards were required to provide three different modes of examining. The balance of these three modes of examining had an important impact on the mechanisms used for awarding grades. The boards were dominated by teachers and a wide variety of different models were adopted. Most boards required that teachers nominated to their subject panels were actively involved in teaching the subject, so they had a vested interest in having an assessment model that was sympathetic to their preferred model of the curriculum. Whalley (1969) outlined the different approaches:

Some Boards treat the modes as being, in effect, distinct examining systems. In such regions, in Mode 1, the school uses a syllabus constructed by the Subject Panel which also externally examines the candidates. There is an examiner appointed for each subject and he works with the Panel. In Mode 2, the school's syllabus is externally examined, also by the panel and the examiner. Schools using Mode 3 construct their own syllabuses and examine their own candidates. The Board appoints moderators who discuss with the teachers in the schools the Mode 3 syllabuses and methods of examination and, when the examination has been carried out, look at the work of the candidates and their marks so that they can check on standards. There is therefore no link between the modes other than that the same Panel is concerned, in different ways, with all three; the methods of examining might be very different in each case.

Whalley (1969)

Some boards, however, merged the three modes as closely as they could while maintaining their distinctive nature. In one example, teachers marked the examinations of all three modes, using marking schemes supplied by the board for Modes 1 and 2 and one devised by the school (approved by the board) for Mode 3. The teacher then graded the candidates and sent the written work of the lowest candidate in each grade to a member of the Moderation Committee. Schools were grouped into sets of no more than 24 and teachers from those schools attended a Group Moderation meeting to consider the standards represented by these candidates who had been judged to be at the lowest acceptable standard for each grade. After any necessary adjustments had been made to bring centres' assessments into line, the Moderation Committee met to make any final adjustments and award the final grades.

However, it was felt that with a completely new examination it was necessary to have a basis of comparison. The mechanism used was a reference test – a multiple choice test of general scholastic ability, or aptitude. Twenty representative schools from each of the CSE boards and nine from each of the GCE boards were selected and the reference test given to all the examination candidates in early 1965, as Whalley described.

Also early in the year, the candidates' teachers were asked to forecast their CSE or GCE grades in the six subjects concerned. On the basis of these forecasts and the results of the reference test it was possible to suggest to each of the CSE Boards, before the examination was held, the proportion of candidates who might appropriately obtain Grade 1 in each of the six subjects, and a mean grade for the subject. Tolerance limits were supplied. The

Boards were not obliged to conform to these standards but were expected to justify any significant departure from them.

Whalley (1969)

When the 1965 results were analysed, no major discrepancies were found in the final results across the boards but there was a general tendency to award grade 4 to a higher proportion of candidates than had been predicted by the test scores. In the spirit of leaving the final judgement to the boards, this was accepted by the newly formed Schools Council for Curriculum and Examinations (commonly referred to as the Schools Council), which had taken over the role of the old SSEC. An interesting outcome of the first monitoring experiment was the comparison of standards at grade 1 with those of the GCE O level pass. If anything, it showed that the CSE boards might have been slightly more severe than the GCE boards at this level.

In 1975, the Department of Education and Science finally succumbed to pressure to allow grades on the GCE O level certificate. But notions of the old 'Pass' were not lost, despite an attempt to remove it. The grades ranged from A to E with those failing to reach the standard required for an E being Unclassified. The standard of the pass was set at grade C and it was so heavily entrenched that even in the years immediately prior to the introduction of the GCSE, the certificates still distinguished between the upper grades which were printed as A(a), B(b) and C(c) so that the lower grades D and E could not be altered to look like higher grades. In the awarding committees, C/D remained the main judgemental boundary, though judgement was also used at grade A/B and E/U.

9 The attempt at criterion referencing

In 1983, the government set up a new body, the Secondary Examinations Council (SEC), to replace the old Schools Council, which it considered cumbersome and an obstruction to reform. Whereas the Schools Council had been representative of stakeholders, with its membership drawn from the teaching profession and local education authorities, the SEC was entirely appointed by the Secretary of State. One of its first functions was to advise on the introduction of a new examination at age 16.

Pressure for the end to the dual system of assessment at 16+ (CSE and O level) led to proposals for a single system of examining at 16+. These were originally put forward in 1976 but the government of the time was unwilling to take the step of abolishing the O level. However, in 1984 Sir Keith Joseph, then Secretary of State, having agreed that the GCSE should be introduced with effect from 1988, indicated one of his hopes for the new examination

We need to define more precisely, for each subject, the skills, competences, understanding and areas of knowledge which a candidate must have covered and the minimum level of attainment he must demonstrate in each of them, if he is to be awarded a particular grade.

Joseph (1984a)

At the time, criterion referenced assessment was regarded as the panacea to all examining ills. James Popham argued strongly that assessment should be against

clearly defined objectives that were known to the candidates as well as to the examiners (Popham, 1978). Work had already begun on developments in Scotland for criterion referenced certification (CRC, 1987). The trouble with Popham's proposals, however, was that they primarily referred to multiple choice questions. These were not common in the English system and relating criterion referencing to British examinations was problematic.

The proposed answer was to be 'grade-related criteria'. This was based upon the assumption that, because examiners could recognise in a script the qualities that would lead them to award it a particular grade, they could articulate what those qualities were. Working parties in the main subject areas were set up, comprising experienced examiners, practising teachers, local authority advisers, representatives of the subject associations, higher and further education, and industry. The terms of reference of the working parties were,

...having viewed the literature dealing with the construction of grade-related criteria and considered the national criteria:

- i. to identify a maximum of six domains of the subject that may be assessed;
- ii. to specify the skills and competencies which candidates must demonstrate in order to achieve the grades at present denoted by CSE grade 4 and O level grades C and A and to suggest appropriate techniques of assessment;
- iii. to determine the method by which the scores on each individual domain may be aggregated to produce a single overall score preserving a level of criterion-referencing compatible with the requirement that current standards are to be carried forward;
- iv. to ensure that the effects of the proposals on the curriculum will be beneficial.

SEC (1984)

In September 1985, the draft grade criteria were published for consultation. Respondents to the consultation were critical of the complexity of the draft grade criteria and the potential damaging impact they would have on the curriculum. In response, the Secondary Examinations Council (SEC) put forward a different approach:

Rather than defining the criteria from the outset, and then arranging for syllabuses to be produced which incorporated these criteria, it decided to take as its starting point a number of existing approved GCSE syllabuses... and to ask the examiners of these syllabuses to articulate their implicit judgements as to the award of grades: in other words, to develop a performance matrix for each of these syllabuses.

DES (1987)

With hindsight this approach was equally doomed to failure. Even if the working groups had been able to identify the qualities they sought, those reading them would have no choice but to make their own interpretations of the requirements. Nevertheless, awarding on the basis of performance matrices was used in at least one awarding body.

It might be expected that, having failed to devise such criteria for the GCSE, the attempt at criterion referencing would be abandoned. However, the development of the General National Vocational Qualifications (GNVQs), based upon the competence statements designed for NVQ work-based assessments, had similar weaknesses. In this case performance statements were devised to identify what the candidates had to do in order to achieve success in a particular aspect. Indeed, in trying to ensure that there was no ambiguity, hurdles were introduced that made it increasingly difficult for candidates to understand the requirements. Box 2 gives an example of some performance statements for a GNVQ in Business (Edexcel, 2007). Problems of interpretation and hurdles are found: who determines whether the candidate's description of the activity is 'clear'? What happens if the candidate has produced a description of what s/he and the teacher regard as four functional areas of the business, but the moderator decides that there are only three? In these circumstances the result was complete failure for the candidate.

Box 2

Pass To achieve a pass your work must show:

- a clear description of the activity, aims and objectives of your chosen business
- a description of four functional areas of the business, including human resources, explaining fully how each contributes to the business activity and giving examples of job roles associated with each area
- a description of how the equal opportunities of employees are safeguarded by legislation
- an explanation, using examples, of how different functional areas communicate with each other and external contacts
- a comparison of the organisational structure of your chosen business with a contrasting structure and explain the different communication flows
- an appropriate list of the sources of evidence you used in your research
- an oral explanation of how the customer service of your chosen business meets customers' expectations and suggestions of any necessary improvements based on best practice
- you can speak clearly during your presentation, keeping to the subject and using an image to illustrate your main point(s).

Criterion referencing for GCSE and GCE was abandoned but there remain some attempts at representing the sorts of performances that might be expected at different grades. Grade descriptions which originate from QCA subject criteria are presented in all GCSE syllabuses as is illustrated in Box 3, taken from a GCSE in geography (OCR, 2007). These are much more general than anything that could be regarded as 'criteria' but are nevertheless used as a reference point in the awarding meetings.

Box 3**Grade C**

Candidates recall accurately information about places, environments and themes, at a range of scales, as required by the specification, and show a broad knowledge of location and geographical terminology.

Candidates understand geographical ideas from the specification content in a variety of physical and human contexts. They understand a range of physical and human processes and their contribution to the development of geographical patterns, the geographical characteristics of particular places and environments, and their interdependence. They understand interrelationships between people and the environment and appreciate that considerations of sustainable development affect the planning and management of environments and resources. They understand the effects of attitudes and values of those involved in geographical issues and in decision-making about the use and management of environments.

Candidates undertake geographical enquiry, identifying questions or issues, suggesting appropriate sequences of investigation, collecting appropriate evidence from a variety of primary and secondary sources, using a range of relevant skills and techniques, reaching plausible conclusions, communicating outcomes, and appreciating some of the limitations of evidence and conclusions.

10 Greater transparency

Although the exercise failed to produce the desired shift towards criterion referencing, the GCSE was a much more transparent qualification than had ever been designed before. The GCSE syllabuses had to conform to centrally produced criteria. They had to include defined content areas and to follow agreed assessment approaches. For the first time, syllabuses for 16-year-olds had to be approved by a central government agency.

The introduction of the GCSE was accompanied by an unprecedented (and since unrepeated) programme to ensure that teachers as well as examiners were fully aware of the changes that were being introduced and how the new examination was to be assessed. This was partly because the new examination included a coursework element that was to be compulsory in almost every subject, so teachers had to be brought into the assessment community. However, it was also as part of the belief that teachers needed to be more aware of the requirements so that they were better able to prepare their students.

As was the case with the change from School Certificate to GCE, the standards of the new examination were based on the standards of the old ones. Grades A, B and C were to remain at the standards set by the O level. Grades D, E, F and G transferred the standards of the CSE grades 2–5. In this way, those involved in the transfer could draw upon their experiences of the previous system to ensure reasonable

comparability from 1987 to 1988 (the transition year). In addition, national data on the overlapping grades of the GCE O level and the CSE were used to give the new GCSE examining groups indications of the likely patterns of results.

One of the major issues facing the awarding bodies in the change from the dual system to the GCSE was how to cope with the wide range of attainment for which the new examination would cater. The Secretary of State had indicated his desire for more pupils to take the examination than had taken either O level or CSE (combined): 'I conclude that it is a realistic objective to try to bring 80–90% of all pupils *at least* to the level now associated with the CSE grade 4' (Joseph, 1984b).

The solution to this problem varied across the different subjects. Most opted for a system of differentiated papers – one covering the lower grades, and another the higher grades, with an overlap of grades in the middle. One subject, mathematics, held out for a three-tier system with no overlap between the highest and lowest tiers. This was heavily influenced by the report of a committee of inquiry into the teaching of mathematics chaired by Sir Wilfred Cockcroft (1982), who had been appointed Chairman of the Secondary Examinations Council. The report argued that there were aspects of mathematics that were essential in the curriculum of the most able that would be so far beyond the understanding of the least able that they would experience only failure. However the desire for a common pattern across subjects has finally brought mathematics into a two-tier system. A few subjects – mainly those that had traditionally used a more narrative response format – chose a form of differentiation that relied upon identifying different levels of sophistication in the responses of candidates to the same stimulus.

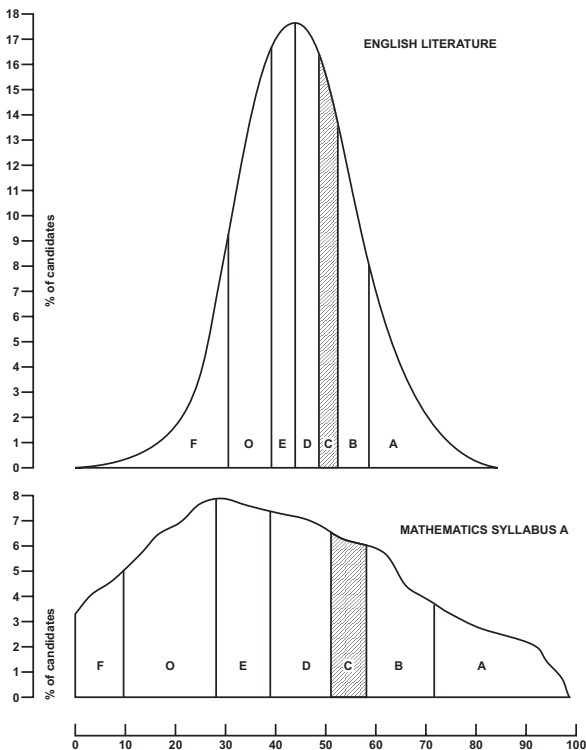
The introduction of tiered papers was the subject of much discussion. Work by Tattersall (1983) and a major research project by Good & Cresswell (1988) both concluded that it was difficult to ensure that standards for the overlapping grades were the same. In the event, the awarding of the first GCSE results was carried out with few problems, despite the changes, though the original plan – to have grades A–C awarded by the O level boards and grades D–G by the CSE boards – was never followed. From the start, awarders from both sides considered all grades. The initial reports by Her Majesty's Inspectorate of Schools (HMI) on the introduction of the GCSE were favourable – particularly in terms of its impact on the curriculum and its motivational value within schools. However, this changed when the focus was turned on the operation of the examination in the boards and HMI started to show concerns about the comparability of systems across the boards.

In 1989, another new examination, the GCE Advanced Supplementary was introduced. Despite pleas by the boards that what was needed was an examination bridging the gap between GCSE and A level, it was decided that the new examination would anticipate a two-year period of study and would be assessed at the same standard as A level but on half the content. Because it was at the same standard and awarded at the same time as the A level, the awarding processes could be integrated. It never achieved the take-up the government had hoped for and was ultimately replaced by the Advanced Subsidiary in 2001.

11 Revised A level grading

The problem with the SSEC (1960) grading recommendations gradually increased with the rapid rise of the numbers of candidates taking A level. In 1969, the JMB put forward proposals for the revision of the A level grading system to alleviate the situation. These resulted in a consultation by the Schools Council in July 1970 on a 20-point grading scale in which the margin of error associated with each point was also specified. However, this was rejected by the Secretary of State in 1972 and the problems continued.

Figure 4 An illustration of the width of A level grade bands for two 1982 JMB examinations



In the late 1970s, increasing concerns were expressed about the inaccuracy of grading at A level. In 1983, the JMB published a pamphlet drawing attention to the narrowness of the mark ranges covering some of the grades. Because the SSEC recommendations for grades were defined in terms of the proportions of candidates, for grades that lay in the middle of the scale, the width could be very narrow indeed; in some cases, the difference between the bottom grade B mark and the top grade D mark – the width of the grade C mark band – was as small as 4% of the mark total (for example, 12 marks out of 300 in an A level geology examination). This is illustrated in Figure 4 (Whittaker & Forrest, 1983, p. 7).

The following year the SEC set up a working party to look into the problems of grading the A level. The working party reported towards the end of 1985, recommending that the

basis for grading should be the partitioning of the mark range, rather than the candidature. It recommended two points that should be determined by examiner judgement: the A/B boundary and the E/N boundary. It also recommended a new grade N to indicate those who narrowly missed grade E. This was an opportunity to bring the boards' different practices into line, but the Secretary of State, concerned about a change to the meaning of the high-profile grade A that might be caused by changing board practices, decided that the boards should continue to use their existing methods for determining this grade (even though these methods differed across the boards just as much as any other grade) and that the upper boundary for the new approach should be the B/C borderline, with the grades between this and the E/N boundary determined arithmetically. This approach was introduced in all awarding bodies in 1987 and it was not until the introduction of the first code of practice that the awarding bodies were required to follow the same procedures for all grade boundaries. The present approach is similar, although the modularisation of A level syllabuses (accompanying the introduction of Curriculum 2000) saw the boundaries to be determined by judgement change from A/B, B/C and E/N to A/B and E/U only in 2002.

12 Conclusion

The awarding process is one of weighing the evidence and coming to a judgement on where to locate grade boundaries. It is an essential part of the comparability process: raw marks are too variable to be relied upon.

Although the awarding process relies heavily on awarders' understanding of their subject, it has been shown that even experienced examiners are relatively poor at making judgements about the 'grade worthiness' of a piece of work on one mark compared with another without something other than subject matter understanding to guide them (Baird & Dhillon, 2005). In the past, examiners had somehow, simultaneously, to account for both the quality of the candidature and the difficulty of the paper. This was essentially a circular task, as the examiners had nothing to judge the quality of the candidature by except the performance of the candidates on the paper whose difficulty they did not know. The statistical information now accessible to awarders means that they have much more available to help them in their task. It is still a matter of judgement, balancing the sometimes conflicting information given by the statistics and their own judgement of scripts. However, awarders are now in a position to test their decisions in a number of ways, retaining their professional input, but complementing it with the power of quantitative analysis, to come closer to the goal of comparability that they are seeking.

References

Assessment and Qualifications Alliance. (2005a). *A basic guide to standard setting*. Guildford: Assessment and Qualifications Alliance.

Assessment and Qualifications Alliance. (2005b). *The use of statistics in examinations processing*. Unpublished, Assessment and Qualifications Alliance.

- Assessment and Qualifications Alliance. (2006). *Uniform marks in GCE, VCE, GNVQ and modular GCSE examinations* (Version 3.1). Guildford: Assessment and Qualifications Alliance.
- Baird, J. (2000). Are examination standards all in the head? Experiments with examiners' judgments of standards in A level examinations. *Research in Education*, 64, 91–100.
- Baird, J., & Dhillon, D. (2005). *Qualitative expert judgements on examination standards: Valid, but inexact*. Internal report RPA 05 JB RP 077. Guildford: Assessment and Qualifications Alliance.
- Black, E.L. (1962). The marking of G.C.E. scripts. *British Journal of Educational Studies*, 11, 61–71.
- Board of Education. (1943). *Curriculum and examinations in secondary schools (Norwood Report)*. London: His Majesty's Stationery Office.
- Cockcroft, W.H. (1982). *Mathematics counts – Report of a committee of enquiry into the teaching of mathematics under the chairmanship of Dr. W.H. Cockcroft*. London: Her Majesty's Stationery Office.
- CRC. (1987). *Criterion-referenced certification project final report*. Glasgow: Jordanhill College of Education.
- Department of Education and Science. (1987). *Improving the basis for awarding GCSE grades*. Paper presented at the 1st Annual Conference of the Joint Council for the GCSE. London: Department of Education and Science.
- Eason, S. (2004). *Statistical screening procedures to investigate inter-awarding body comparability*. London: Joint Council for Qualifications.
- Edexcel. (2007). *Intermediate GNVQ in business (GB206). (Specification)*. London: Edexcel.
- Forrest, G.M. (1970). *Standards in subjects at the Ordinary level of the GCE*. Occasional Paper 33. Manchester: Joint Matriculation Board.
- Good, F.J., & Cresswell, M.J. (1988). *Grading the GCSE*. London: Secondary Examinations Council.
- Joseph, K. (1984a). Speech by Sir Keith Joseph, Secretary of State for Education and Science, at the Assistant Masters and Mistresses Association, Bournemouth, Monday 16 April, 1984. *SEC annual report 1983–84*. London: Secondary Examinations Council.
- Joseph, K. (1984b). Speech by Sir Keith Joseph, Secretary of State for Education and Science, at the North of England Conference, Sheffield, Friday 6 January, 1984. *SEC annual report 1983–1984*. London: Secondary Examinations Council.

Oliver, R.A.C. (1960). *The effectiveness of G.C.E. Advanced level as a criterion for university selection*. Occasional Paper No. 7. Manchester: Joint Matriculation Board.

Oxford Cambridge and RSA Examinations. (2007). *GCSE geography A (1986)*. (Specification). Nottingham: Oxford Cambridge and RSA Examinations.

Popham, W.J. (1978). *Criterion-referenced measurement*. New Jersey: Prentice-Hall Englewood Cliffs.

Qualifications and Curriculum Authority. (2007). *GCSE, GCE, VCE, GNVQ and AEA code of practice*. London: Qualifications and Curriculum Authority.

Secondary Examinations Council. (1984). *The development of grade-related criteria for the General Certificate of Secondary Education – The task of the working parties*. Unpublished, Secondary Examinations Council.

Secondary School Examinations Council. (1960). *The General Certificate of Education and sixth form studies*. Third Report of the Secondary School Examinations Council. London: Her Majesty's Stationery Office.

Secondary School Examinations Council. (1963a). *The Certificate of Secondary Education: Some suggestions for teachers and examiners*. Examinations Bulletin No. 1. London: Her Majesty's Stationery Office.

Secondary School Examinations Council. (1963b). *Scope and standards of the Certificate of Secondary Education*. Seventh Report of the Secondary School Examinations Council. London: Her Majesty's Stationery Office.

Tattersall, K. (1983). *Differentiated examinations: A strategy for assessment at 16+?* Schools Council Examinations Bulletin 42. London: Methuen Educational.

Tomlinson, M. (2002). *Inquiry into A level standards. Final Report*. London: Department for Education and Skills.

University of Cambridge Local Examinations Syndicate. (1950). *Minutes of the awarding committee 12 October 1950 (arrangements for the award of the GCE in 1951)*. Cambridge: University of Cambridge Local Examinations Syndicate Archives.

University of Cambridge Local Examinations Syndicate. (1951). *General Certificate of Education: The award and communication of marks*. Cambridge: University of Cambridge Local Examinations Syndicate Archives.

Whalley, G.E. (1969). *The Certificate of Secondary Education*. Institute of Education Paper No. 9. Leeds: University of Leeds.

Whittaker, R.J., & Forrest, G.M. (1983). *Problems of the GCE Advanced level grading scheme*. Manchester: Joint Matriculation Board.