GRADING THE SPECIALISED DIPLOMA

Martin Taylor, Jeremy Pritchard and Elizabeth Gray

Executive summary

This paper describes a number of aggregation methods for generating an overall result for the Diploma, some producing just a Pass/Fail result and some a graded result. It illustrates certain difficulties which arise with aggregation, including regression to the mean (whereby the overall results contain more middle grades and fewer top and bottom grades than the results for individual units), the unreliability associated with aggregating results which are expressed on a coarse scale, and the effect of imposing hurdles (such as a requirement to pass all units). It gives examples of the overall results, under certain aggregation methods, for candidates with various profiles across the units, showing that:

(i) the imposition of a hurdle is liable to cause candidates with very similar profiles to obtain completely different overall results;

(ii) the effect of including ungraded (ie Pass/Fail) units in the aggregation can vary enormously, depending on the value assigned to a Pass result in such units;

(iii) the rank order of candidates, in terms of their Diploma grades, can vary depending on the aggregation method used.

Most of the methods which generate a graded Diploma result require at least some of the units to be graded. As well as the problems of dealing with ungraded units in these methods, the relative values of the grades across the graded units need to be determined. Such an exercise has been carried out by UCAS for the Level 3 qualifications introduced into its tariff structure alongside GCE and also by QCA for the purpose of school and college performance indicators, but there would undoubtedly be serious risks to resolve in deciding upon tariff values for the grades in all different types of unit included in the Diplomas. The opportunity for candidates to take units at a higher or lower level than the Diploma itself would add a further complication. Many other diploma frameworks (in other European countries, for example) do not have overall grades. In those which do, the elements which are aggregated are normally similar in nature or, at least, are all graded in the same way.

The aggregation method for the Diploma must be fair in the sense that it reflects a candidate's profile across the units and must be transparent in the sense that the relationship between the profile and the overall result can be readily understood, at least in broad terms. The method must not act in opposition to the educational aims of the Diploma by encouraging centres and candidates to choose units solely for the purpose of maximising grades. At a given level, Diplomas across all lines of learning, as well as all routes within a particular line of learning, must be seen as comparable.

Any method which generates a graded result for the Specialised Diploma (as opposed to a simple Pass/Fail result) will be mechanistic and somewhat arbitrary, with the grades being of little value

Copyright © 2014 AQA Education and its licensors. All rights reserved.
AQA Education (AQA) is a registered charity (registered charity number 1073334) and a company limited by guarantee registered in England and Wales (company number 3644723). Registered address: AQA, Devas Street, Manchester M15 6EX.

at best and misleading at worst. Diplomas in the various lines of learning might have quite different overall results largely because of the vagaries of the aggregation method. Such differences will be difficult to justify and pose a serious risk to the credibility of the Diploma. The problems with the first Curriculum 2000 A level awards in 2002 might seem trivial in comparison. It is strongly recommended that the success of the Diploma will be more assured if there is no grading of the overall result, with the quality of a candidate's achievement being provided by his/her grade profile across the units and by the evidence in his/her transcript.

GRADING THE SPECIALISED DIPLOMA

1 INTRODUCTION

Part of the Government's response to the report of the Working Group on 14-19 Reform (DfES, 2004) was to announce the introduction of Specialised Diplomas in fourteen lines of learning, each Diploma including both academic and vocational material and covering a specific occupational sector. The intention, as stated in the White Paper *14-19 Education and Skills* (DfES, 2005a), was to 'rationalise the existing very wide array of 3500 vocational qualifications available to young people into much more easily recognisable and understandable Diplomas, containing both specialised material and GCSEs and A levels' (Paragraph 6.11).

The later *14-19 Education and Skills Implementation Plan* (DfES, 2005b) envisages that, of about 450 000 14-15 year-olds (Years 10 and 11) studying for vocational qualifications[1] in 2014 (up from about 320 000 in 2004), about 350 000 will be studying for Specialised Diplomas. To put these figures in context, the size of the cohort (14 and 15 year-olds combined) in 2014 will be about 1.2 million (down from about 1.4 million in 2004) and in 2004 the GCSE subject with the largest entry, Mathematics, had about 740 000 UK candidates.

Before the first five Diplomas are introduced for first teaching in September 2008 (and first certification in Summer 2009 or Summer 2010), there are several important issues which need to be resolved. One of these concerns grading. Results in individual units will be graded where appropriate, but there are a number of difficulties in grading the Diploma overall. While successful completion clearly needs to be recognised, the benefits of certificating achievement beyond the basic pass threshold are more debatable. Essentially, the question is whether there should be just a two-point scale (Pass/Fail) or a multi-point scale (eg Distinction/Merit/Pass/Fail). This is the issue addressed in this paper.

2 BACKGROUND

Each Specialised Diploma is intended to be available at Level 1 (broadly equivalent to GCSE Grade D-G), Level 2 (broadly equivalent to GCSE Grade A*-C) and Level 3 (broadly equivalent to GCE Grade A-E). In terms of amount of learning, the Level 1 Diploma is intended to be comparable to a programme of four to five GCSEs, the Level 2 Diploma to five to six GCSEs and the Level 3 Diploma to three GCE A levels. There will eventually be Diplomas for each of the fourteen lines of learning as defined in the White Paper *14-19 Education and Skills* (op cit). The first five Diplomas will be available for first

[1] 'Vocational qualifications' include, for example, GCSEs in vocational subjects and NVQs. It is envisaged that, following the introduction of Specialised Diplomas, no 14-15 year-olds will take NVQs as stand-alone qualifications.

teaching from September 2008, the next five from September 2009 and the remaining four from September 2010. Each Diploma is being developed by the relevant Diploma Development Partnership (DDP), a consortium consisting of one or more Sector Skills Councils (SSCs) together with representatives from awarding bodies, higher education and other groups.

Each Diploma is likely to consist of some units which contribute to existing qualifications and some units which will be new. While the DDPs will scrutinise and approve units for their suitability in terms of factors such as vocational relevance, there will also need to be generic rules which define, for example, the permitted sizes of units and the proportion of the units which must be at the same level as the Diploma. It could be unhelpful to be over-prescriptive about the permitted sizes, as to do so would risk disqualifying units for largely bureaucratic reasons. It could also be unhelpful to prescribe a certain number or proportion of newly-designed units which should be incorporated – providers and users should not be forced away from courses and qualifications which they already value and which are fit-for-purpose as elements of the Diploma.

3 PRINCIPLES FOR AGGREGATION AND GRADING

The Diploma will have to satisfy several different and possibly competing requirements (for example, those of employers, higher education, centres and, most crucially, of candidates themselves), and it is important to define some principles against which any aggregation scheme[2] should be judged.

3.1 The aggregation scheme must be *fair* in the sense that it must ensure that a candidate's overall result reflects his/her profile across the various elements of the Diploma. The overall result of one candidate in relation to another must be justifiable from their profiles.

3.2 The aggregation scheme must be *transparent*. It must be possible for centres, candidates and users broadly to understand the relationship between a candidate's profile and his/her overall result. (Unfortunately, in qualifications with a complex structure *fairness* and *transparency* are not always compatible.)

3.3 The Diploma must be *achievable*. The eagerness to make it a highly respected and sought-after qualification must not cause it to be so hard that few candidates are able to obtain it.

3.4 If the purpose of the Diploma is to ensure balanced learning and achievement, any aggregation and grading scheme must ensure that all elements have been completed to the required standards. However, many current grading schemes use a *compensation model*, whereby candidates can make up for poor performance on some units by performing very well on others. If the Specialised

[2] The *aggregation scheme* for the Diploma is the method of combining the results of the various elements to generate an overall result for the qualification as a whole, whether this is just Pass/Fail or a graded result.

Diploma uses a different model, candidates may be at a disadvantage compared to their counterparts who take other qualifications.

3.5 Every Diploma at a given level must be seen as *comparable* in the sense that one line of learning should not be seen as easier than another. This principle suggests that the structure across the various lines of learning should be similar, if not identical. Different grading models for different lines of learning (which might be needed if the structures varied) cannot be acceptable. In particular, the proportion of the Diploma taken at each level must be consistent. For example, for a Level 3 Diploma candidates might be required to pass at least 75% of the contents at Level 3 – this figure should not vary across lines of learning.

3.6 The permitted contribution from units taken at a higher level or at a lower level than the Diploma must be defined in a consistent way across all lines of learning.

3.7 Many existing vocational qualifications and units are not graded but are just Pass/Fail. The aggregation system must not automatically advantage or disadvantage candidates taking such units as part of a Diploma.

3.8 The aggregation scheme must ensure that appropriate recognition is paid to all aspects of the Diploma, otherwise centres and candidates are liable to concentrate on the parts which are likely to maximise success.

4 DIPLOMA VOLUME

Before grading can be considered, a method must be agreed for determining the demands of the various elements of the Diploma simply in terms of *volume*, ie the amount of learning required. The Diploma would lose credibility if there were obvious variations in this respect between the different lines of learning or between the different routes within any one line of learning.

The *Final Report of the Working Group on 14-19 Reform* (DfES, 2004) states the following (Paragraph 107).

> 'Any system which combines units or components into whole programmes needs a means of measuring the amount or volume of learning successfully undertaken by the learner. We believe that a credit system provides the most appropriate means of expressing and measuring the amount of learning undertaken by learners within the reformed 14-19 framework, and of providing a mechanism for consistent aggregation of components into whole diplomas.'

It goes on to recommend the following (Recommendation 16).

> 'Each available diploma component should be assigned a credit value according to the volume of learning it contains, and each

diploma should require successful achievement of a minimum number of credits.'

Many units and qualifications have specified numbers of *guided learning hours* (glh). Of course, it is not expected that all institutions will adhere rigidly to these numbers in terms of the time allocated for supervised learning, but the guided leaning hours do provide a broad indication of volume. For example, both a GCE A level and a BTEC National Award have 360 guided learning hours, and should therefore be regarded as the same in terms of volume (the amount of learning required). However, for some qualifications (notably NVQs) it is much more difficult to determine volume, as they are not assigned guided learning hours. When the Framework for Achievement has been fully developed, every unit and qualification will have an assigned credit value, but this work will not have been completed before the first Diplomas are introduced in September 2008. Where units or qualifications do not have guided learning hours assigned, the only existing designations of volume are in the tariffs developed by QCA for use in the School and College Achievement and Attainment Tables, in the LSC funding tariffs, in the UCAS tariffs (Level 3 only) and in frameworks developed in Wales, Northern Ireland and Scotland. However, depending on the range of existing units and qualifications used in the Diplomas, it is unlikely that any of these will be adequate. Also, because they have been developed for varying purposes, they are not entirely consistent with one another. A considerable amount of work may be needed to determine credit values both for existing units and for new units within the Diploma.

A credit system like the one being developed for the Framework for Achievement defines only *level* and *volume* – it makes no attempt to define how *well* a candidate has performed, ie the *quality* of a candidate's achievement. This limitation is a feature shared by other credit frameworks which exist or are being developed such as the Credit & Qualifications Framework for Wales, the Scottish Credit & Qualifications Framework (SCQF), the European Credit Transfer system and the New Zealand National Qualifications Framework. For example, the SCQF states that 'judgements on how well a learner has performed, and the grading and classification systems used to report that performance, are the responsibility of awarding bodies' (SCQF, 2003), ie they are not part of the framework. Most methods of determining an overall grade for the Diploma (see section 5 below) would require some means of recording the *quality* of a candidate's achievement for at least some of the units and on a common scale – credit values on their own would not be sufficient.

5 AGGREGATION MODELS

It is suggested that there are essentially two models for aggregating the results of individual units[3] to produce an overall outcome, whether a grade or simply Pass/Fail. One (Model 1) involves using the level(s) of the units taken and the volume while the

[3] It is assumed, for simplicity, that *unit* results are aggregated to produce an overall Diploma result. However, incongruities may occur where a qualification forms part of a Diploma. For example, in GCE and BTEC candidates who pass the qualification do not necessarily pass every unit. In such a situation it might seem unduly severe if there were a requirement to pass every unit to gain the Diploma when there is no such requirement to pass one of its constituent qualifications.

other (Model 2) also takes account of how well a candidate has performed (ie his/her grade) within some or all of the units. There are numerous versions of each model (for example, whether or not a Pass is required in each unit and whether or not a graded overall result is based on all units or just some). Some of these are described later in this section and a few are exemplified in Section 7.

It is broadly agreed that a Diploma may contain units which are at a higher or lower level than the Diploma, although there has so far been no consensus on the proportions which might be allowed at different levels (QCA, 2006a).

It is assumed that each unit has a credit value (see Section 4 above). In Model 1 a candidate's performance on each unit is described as simply Pass/Fail. In Model 2 performance on each unit needs to be described by a grade.[4] A candidate's *tariff points* for a unit can then be derived from the credit value and the grade.

Some versions of Models 1 and 2 are described below. Section 7 gives examples of the grades which candidates with various profiles across the Diploma units would receive under these models.

Model 1.1

- There is no overall grade for the Diploma.
- In order to obtain the qualification, candidates must pass all units.
- No account is taken of the levels of the units provided that the proportions of the total credit value taken at higher or lower levels fall within the agreed limits.

Model 1.2

- There is no overall grade for the Diploma.
- In order to obtain the qualification, candidates must pass all units of Generic Learning.
- Candidates must also gain at least a given proportion of the total credit value of Principal Learning and at least a (possibly different) given proportion of the total credit value of Additional/Specialist Learning.
- No account is taken of the levels of the units provided that the proportions of the total credit value taken at higher or lower levels fall within the agreed limits.

Other versions of Model 1 could include:

[4] Even where all units are graded on a similar scale (eg a three-point scale Distinction/Merit/Pass), there will be comparability issues, for example whether Distinction in one unit represents a similar degree of challenge to a Distinction in another unit. Where grades for different units are not on similar scales (eg some on a three-point scale and some on a five-point scale), they will need to be translated to a common scale, again with comparability issues. For ungraded (ie Pass/Fail) units, a decision will need to be made about the grade to be assigned to Pass - for example, if a three-point grading scale is used, should the top grade, the middle grade or the bottom grade be assigned to Pass? There are even more serious comparability issues here.

(i) a requirement to pass a given proportion of the total credit value at the level of the Diploma and another (higher) given proportion at no more than one level below;

(ii) an overall grade (eg Distinction, Merit, Pass or Fail) awarded on the basis of the proportions of the total credit value passed at various levels (for example, in order to gain Pass candidates might need to pass 60% of the total credit value for Principal Learning at the level of the Diploma while to gain Distinction they might need to pass *all* Principal Learning at the level of the Diploma, with corresponding criteria for Generic Learning and Additional/specialist Learning);

(iii) an overall grade awarded on the basis of additional volume (for example, in order to obtain Merit or Distinction, candidates might need to take extra units, beyond the minimum requirement, and they might also need to pass some units at a higher level than the Diploma).

The second and third of these options would introduce an overall grade without the need for the contentious issues of assigning a grade to Pass in ungraded units (required in all versions of Model 2, unless this type of unit is artificially excluded from the units which contribute to the overall grade) and aligning grades across different types of graded units. Certainly, the *Final Report of the Working Group on 14-19 Reform* (DfES, 2004) envisaged that higher grades might be awarded for broader achievement (Paragraph 180). However, the requirements would need to be carefully framed in order to avoid adding to the assessment burden by encouraging candidates to accumulate units for no educational benefit. Also, an overall Diploma grade could lack credibility if it failed to take any account of high achievement in graded units.

Model 2.1

- Candidates must pass all units (the 'hurdle').
- For ungraded units, Pass is assigned a specified grade on the grading scale, according to an agreed rule.[5]
- A candidate's overall grade is determined by adding up the tariff points (but the candidate fails if the hurdle has not been met).

Model 2.2

- Candidates must pass Functional Skills at the appropriate level[6] as well as PLTS[7] (the 'hurdle').
- For ungraded units, Pass is assigned a specified grade on the grading scale, according to an agreed rule.
- A candidate's overall grade is determined by adding up the tariff points (but the candidate fails if the hurdle has not been met).

[5] For transparency, the same rule should apply to all units in all Diplomas. For example, it would be confusing if the top grade were to be assigned to Pass in some ungraded units and the bottom grade to Pass in others.
[6] ie at the level of the Diploma at Levels 1 and 2 and no more than one level below the Diploma at Level 3.
[7] Personal, Learning and Thinking Skills

Model 2.3

- Candidates must pass Functional Skills at the appropriate level as well as PLTS (the 'hurdle').
- A candidate's overall grade is determined by adding up the tariff points for the Project and Principal Learning (but the candidate fails if the hurdle has not been met). All units of Principal Learning will be graded – the adoption of Model 2.3 involves the need to prohibit Pass/Fail units in this Diploma component.

Other versions of Model 2 could include:

(i) a requirement to pass Functional Skills and PLTS in order to be eligible for overall Pass and an additional requirement to pass a certain proportion of the other units (which could be 100%) in order to be eligible for a higher overall grade;

(ii) reduced tariff points for units taken at a lower level than the Diploma and increased tariff points for units taken at a higher level.

These variations would arguably make the grading fairer but at the same time they would make it more complex and less transparent.

6 THE EFFECTS OF AGGREGATION

Section 7 below describes how the models in Section 5 would work for specific examples of candidate profiles. First, in this section, some more general issues regarding aggregation are described and are illustrated in terms of existing qualifications. These issues are relevant to the examples in Section 7.

Experience with existing qualifications shows that aggregation of the results from individual elements can affect overall outcomes in unfortunate ways. In particular:

(i) aggregation causes a phenomenon known as *regression to the mean*, whereby the overall results contain more middle grades and fewer top and bottom grades than the results for individual elements;

(ii) aggregating results expressed on a coarse scale is liable to produce an overall grade distribution which is significantly different from the distribution obtained when aggregating the same results expressed on a finer scale;

(iii) the introduction of hurdles can make enormous (and unpredictable) differences to overall grade distributions.

Point (i) is illustrated from an existing qualification by aggregating the unit results from a six-unit A level (actually Edexcel English Literature, with 18190 candidates) in stages: Unit 1 alone, then Units 1 and 2, then Units 1, 2 and 3, and so on. The effects of the aggregation are shown in Figures 6.1(a) - 6.1(f) below. (The grade distributions for Units 2-6 individually are not shown, as each one is almost identical to the distribution for Unit 1, illustrated in Figure 6.1(a).)

Figure 6.1a

One unit award (Unit 1)

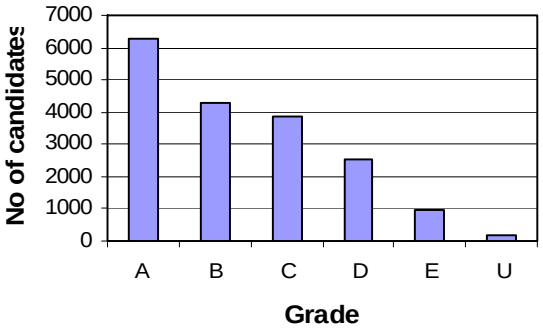


Figure 6.1b

Two unit award (Units 1 and 2)

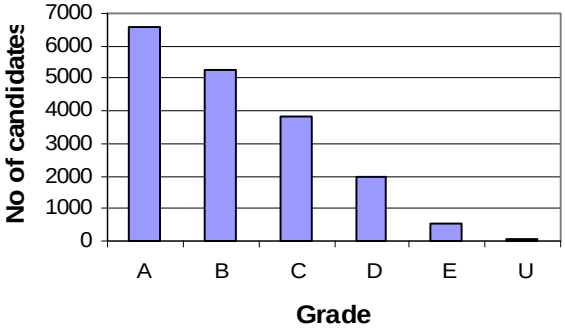


Figure 6.1c

Three unit award (Units 1-3)

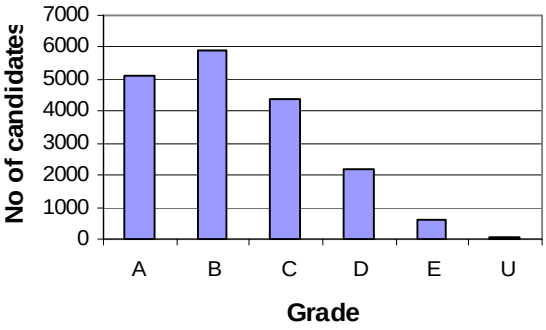


Figure 6.1d

Four unit award (Units 1-4)

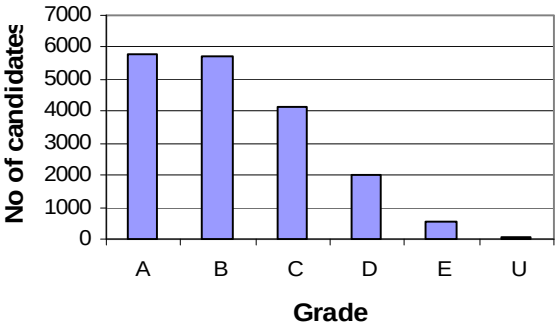


Figure 6.1e

Five unit award (Units 1-5)

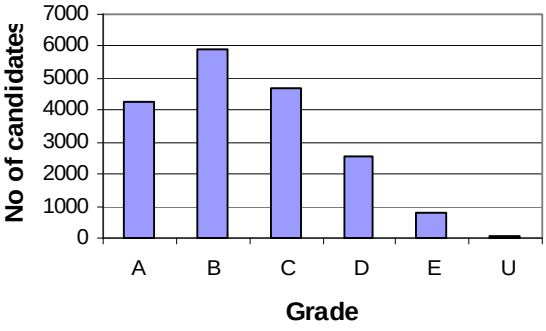
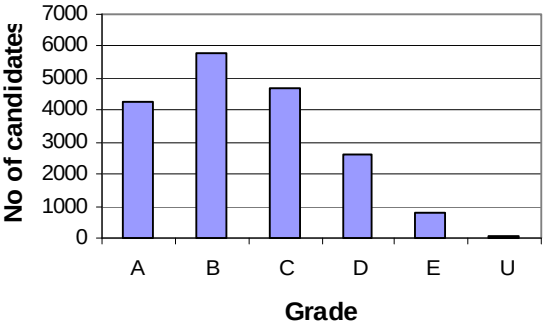


Figure 6.1f

Six unit award (Units 1-6)

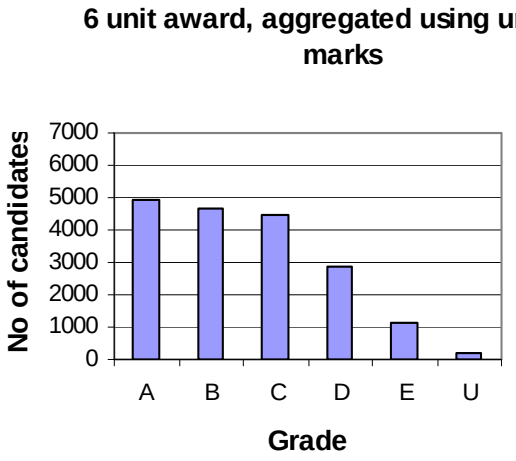


The diagrams show that aggregating across six units greatly reduces the number of candidates obtaining Grade A and greatly increases the number obtaining Grade B. The

number obtaining Grade C increases slightly, Grade D is almost unchanged, and Grades E and U decrease slightly. The diagrams show the progressive effect of regression to the mean as more units are brought into the aggregation.

The aggregation in the above diagrams was carried out using a points system, with A=5, B=4, C=3, D=2 and E=1 in each unit. For live GCE examinations the unit results are aggregated using uniform marks, which means that unit performance *within* a grade is taken into account in the aggregation. In this case the use of uniform marks gives the grade distribution shown in Figure 6.2. Comparison of Figure 6.1f and 6.2 provides an illustration of point (ii) above, that the overall outcomes are significantly different depending on whether a coarse scale (Figure 6.1f) or a fine scale (Figure 6.2) is used.

Figure 6.2



To illustrate point (iii) regarding the imposition of a hurdle, candidates were awarded a putative twelve unit qualification using their unit results from two (six unit) A levels. The units were graded using A=5, B=4, etc (as before) and overall grades were awarded as follows:

Distinction	44-60
Merit	28-43
Pass	12-27.

The results are shown in Table 6.3. The third column of the table shows how the cumulative percentages are reduced when there is a requirement (ie a hurdle) that candidates must pass all twelve units.

Table 6.3 Change in cumulative percentages of candidates obtaining the grades in a putative twelve-unit award when a hurdle is introduced
(number of candidates = 1034)

	Cumulative percentage of candidates	
	Compensation allowed	**All units passed**
Distinction	61.1	58.0
Merit	90.0	66.8
Pass	98.9	66.8

Thus, 90.0% of candidates achieve Merit or above if the criterion is simply that they must score a total of 28 points or more, but this figure falls to 66.8% if the hurdle that they must pass all units is imposed. The fact that, when the hurdle is imposed, the percentage of candidates achieving Pass or above is still 66.8% indicates that there were no candidates who obtained the 12-27 points needed for a Pass *and* passed all twelve units (ie all those who passed all twelve units gained sufficient points for a Merit).

The 2004 BTEC National Diploma results are also used to illustrate the effects of imposing a hurdle. The BTEC National Diploma is an eighteen unit qualification in which each unit is graded Pass (2 points), Merit (4 points) or Distinction (6 points).[8] Candidates are awarded an overall grade of DDD, DDM, DMM, MMM, MMP, MPP or PPP based on their total points for all eighteen units (for example, DDD requires a points total of 86 or more, while PPP requires 36-42). Table 6.4 shows the 2004 results for the 31 148 candidates who obtained a BTEC National Diploma at this time, together with the results which would be generated if hurdles were imposed. For example, 76.7% of candidates obtained MPP or higher, but this figure falls to 53.6% if the hurdle that all units must be passed is imposed.

Table 6.4 Actual results for BTEC National Diploma (2004 certification) and results when hurdles are imposed (cumulative percentages of candidates)

	Grade						
	DDD	**DDM**	**DMM**	**MMM**	**MMP**	**MPP**	**PPP**
Actual results	17.5	26.7	33.2	47.4	59.5	76.7	100.0[9]
All units passed	13.0	19.9	24.9	34.9	42.9	53.6	66.7
At most one unit failed	17.4	26.4	32.7	46.3	57.2	72.2	91.0

It can be seen that the results fall dramatically at the lower grades when there is a requirement that all units must be passed. A less stringent requirement, that candidates must pass all but one unit, makes much less difference to the actual results (for example, 72.2% of candidates would obtain MPP or above compared with the actual figure of 76.7%). However, there is a hurdle in the live qualification that candidates must pass all but two units, and a requirement to pass all but one unit is only slightly stricter. The

[8] Some externally-set units have double weighting, ie Pass = 4 points, Merit = 8 points, Distinction = 12 points.
[9] There is no Fail, because candidates do not claim the award unless they have a graded result (ie PPP or better).

modest decrease in the actual results in the last row of Table 6.4 must be seen in that context.

Another finding from the investigation of BTEC National Diploma results is the enormous variation across sectors in the effect of the imposition of a hurdle. As noted in relation to Table 6.4, candidates do not claim this qualification unless they have at least an overall Pass, so in that sense the pass rate can be said to be 100%. If a hurdle that all units must be passed is imposed, the pass rate varies from under 10% in some sectors to over 90% in others. There is much less variation if the hurdle allows candidates to fail one unit – in this case the pass rate is over 80% in all of the sectors investigated. However, this more limited variation must be seen in the context of the hurdle in the live qualification, as noted in the previous paragraph.

The conclusion from these examples must be that, if a hurdle is imposed, it should not be an unduly strict one (such as a requirement to pass all units).

7 EXEMPLIFICATION OF THE MODELS IN SECTION 5

In this section the overall grades obtained by candidates with various unit profiles are calculated using the aggregation models described in Section 5. The following Diploma structure has been used:

- Principal Learning consists of eight units and Additional/specialist Learning consists of two units, each with 120 notional learning hours (nlh);
- each of the Functional Skills (English, Mathematics, ICT) has 80 nlh;
- PLTS (Personal, Learning and Thinking Skills) has 60 nlh;
- the Project has 120 nlh.

In terms of the learning time allocated to each part of the Diploma, this is almost identical to the structure of a Level 2 Diploma.[10]

Each unit is assigned a credit value, using the rule that 1 credit is equivalent to 10 nlh. It is assumed that each unit is *either* graded as Pass/Merit/Distinction *or* is simply Pass/Fail.

The structure used in the exemplification is summarised in Table 7.1.

Examples of candidate profiles and the results for the Diploma overall are shown in Table 7.2, with the following rules and conventions.

(i) The results for graded units are shown on the scale 0-3 (3 = Distinction, 2 = Merit, 1 = Pass, 0 = Fail) while the results for ungraded units are shown as P (Pass) or F (Fail).

[10] See, for example, *The Specialised Diploma: qualification structure* (QCA, 2006b). The number of notional learning hours (which include unsupervised learning) is taken to be twice the number of guided learning hours.

Table 7.1 Diploma structure used in the exemplification (Level 2)

		nlh	credit value	nlh	credit value	nlh	credit value
Generic Learning	**Functional English**	80	8	420	42	1620	162
	Functional Maths	80	8				
	Functional ICT	80	8				
	PLTS	60	6				
	Project	120	12				
Principal Learning	**Unit 1**	120	12	960	96		
	Unit 2	120	12				
	Unit 3	120	12				
	Unit 4	120	12				
	Unit 5	120	12				
	Unit 6	120	12				
	Unit 7	120	12				
	Unit 8	120	12				
Additional/ specialist Learning	**Unit 1**	120	12	240	24		
	Unit 2	120	12				

(ii) A similar convention is used for overall Diploma results. These are shown on the scale 0-3 in Model 2 but as P or F in Model 1.

(iii) For the purpose of aggregation in Model 2, a candidate's result in each unit is converted to *tariff points* by multiplying the credit value by the candidate's grade on the unit.

(iv) A candidate's overall grade (Distinction, Merit, Pass or Fail) in Model 2 is determined by using appropriate thresholds to convert his/her total points score to a grade. The Pass threshold is calculated by adding the tariff points of a candidate who has a Pass for each unit, and the range between this threshold and the maximum possible total points score is divided into three equal parts to determine the Merit and Distinction thresholds.

(v) Two versions of Models 2.1 and 2.2 are exemplified. In the first version (shown as Models 2.1.1 and 2.2.1 in Table 7.2), Pass in an ungraded unit is assigned the same value as Pass in a graded unit. In the second version (shown as Models 2.1.2 and 2.2.2 in Table 7.2), Pass in an ungraded unit is assigned the same value as Distinction in a graded unit.

(vi) In Model 1.2 the proportions of Principal Learning and Additional/specialist Learning which candidates must pass are taken as 6 (out of 8) units and 1 (out of 2) units respectively.

Table 7.2 illustrates that the overall grades obtained by candidates are highly dependent on the aggregation model used. Not only do candidates obtain different grades, depending on the model, but, more alarmingly, the rank order of candidates is sometimes reversed if the model is changed.

Table 7.2 Examples of candidate profiles and the overall grades under various grading models

Where results are graded (ie for the Project, for the Principal Learning units, for the Additional/specialist Learning units and for the overall Diploma in Models 2.1.1 - 2.3), they are shown as 3 (Distinction), 2 (Merit), 1 (Pass) or 0 (Fail). Where results are Pass/Fail (ie for Functional Skills, for PLTS, for certain other units in 7.2.3 and for the overall Diploma in Models 1.1 and 1.2, they are shown as P (Pass) or F (Fail).

7.2.1 Candidates whose overall grades are a fair reflection of their unit profiles

	Generic Learning					Principal Learning								Add/Spec		Result under each model						
	Eng	Ma	ICT	PLTS	Proj	1	2	3	4	5	6	7	8	1	2	Pass/Fail		Graded				
nlh	80	80	80	60	120	120	120	120	120	120	120	120	120	120	120	1.1	1.2	2.1.1	2.1.2	2.2.1	2.2.2	2.3
credits	8	8	8	6	12	12	12	12	12	12	12	12	12	12	12							
Cand																						
1	P	P	P	P	3	3	3	3	3	3	3	3	3	3	3	P	P	3	3	3	3	3
2	P	P	P	P	2	2	2	2	2	2	2	2	2	2	2	P	P	2	2	2	2	2
3	P	P	P	P	1	1	1	1	1	1	1	1	1	1	1	P	P	1	1	1	1	1
4	P	P	P	P	0	0	0	0	0	0	0	0	0	0	0	F	F	0	0	0	0	0
5	P	P	P`	F	3	3	3	3	3	3	3	3	3	3	3	F	F	0	0	0	0	0

For four of these candidates the overall grade is a fair reflection of their profile across the units. However, candidate 5 is defeated by having failed PLTS, which is a hurdle in all models.

7.2.2 High-achieving candidates

	Generic Learning					Principal Learning								Add/Spec		Result under each model						
	Eng	Ma	ICT	PLTS	Proj	1	2	3	4	5	6	7	8	1	2	Pass/Fail		Graded				
nlh	80	80	80	60	120	120	120	120	120	120	120	120	120	120	120	1.1	1.2	2.1.1	2.1.2	2.2.1	2.2.2	2.3
credits	8	8	8	6	12	12	12	12	12	12	12	12	12	12	12							
Cand																						
6	P	P	P	P	3	3	3	3	3	3	3	3	3	3	3	P	P	3	3	3	3	3
7	P	P	P	P	3	3	3	3	2	2	2	2	2	2	2	P	P	3	3	3	3	3
8	P	P	P	P	2	2	2	2	2	2	2	3	3	3	3	P	P	3	3	3	3	2
9	P	P	P	P	0	3	3	3	3	3	3	3	3	3	3	F	F	0	0	3	3	0

All candidates appear to be equally qualified under Models 2.2.1 and 2.2.2 (in which only the Functional Skills and PLTS are hurdles) but candidate 9 fails under the other models. Candidate 8 has only overall Merit under Model 2.3, although arguably his/her profile (four Distinctions and six Merits) is no worse than that of candidate 7.

7.2.3 The effect of ungraded units

	Generic Learning					Principal Learning								Add/Spec		Result under each model						
	Eng	Ma	ICT	PLTS	Proj	1	2	3	4	5	6	7	8	1	2	Pass/Fail		Graded				
nlh	80	80	80	60	120	120	120	120	120	120	120	120	120	120	120	1.1	1.2	2.1.1	2.1.2	2.2.1	2.2.2	2.3
credits	8	8	8	6	12	12	12	12	12	12	12	12	12	12	12							
Cand																						
10	P	P	P	P	1	1	1	1	1	1	1	1	1	1	1	P	P	1	1	1	1	1
11	P	P	P	P	1	P	P	P	P	P	P	P	P	P	P	P	P	1	3	1	3	1

For candidate 11, who has done all ungraded units, the overall grade varies considerably depending on the value assigned to a Pass in an ungraded unit. Thus, in terms of overall grade, candidate 11 can appear much better than or the same as candidate 10, depending on the grading model used.

7.2.4 Rank order – Distinction/Merit

	Generic Learning					Principal Learning								Add/Spec		Result under each model						
	Eng	Ma	ICT	PLTS	Proj	1	2	3	4	5	6	7	8	1	2	Pass/Fail		Graded				
nlh	80	80	80	60	120	120	120	120	120	120	120	120	120	120	120	1.1	1.2	2.1.1	2.1.2	2.2.1	2.2.2	2.3
credits	8	8	8	6	12	12	12	12	12	12	12	12	12	12	12							
Cand																						
12	P	P	P	P	2	2	2	2	2	2	2	3	3	3	3	P	P	3	3	3	3	2
13	P	P	P	P	3	3	3	2	2	2	2	2	2	2	2	P	P	2	2	2	2	3
14	P	P	P	P	3	3	3	3	3	3	3	1	1	1	1	P	P	2	2	2	2	3
15	P	P	P	P	1	1	1	2	3	3	3	3	3	3	3	P	P	3	3	3	3	2

The rank order of these candidates varies depending on the grading model used. Under Models 2.1.1, 2.1.2, 2.2.1 and 2.2.2 candidates 12 and 15 appear better than candidates 13 and 14. However, under Model 2.3 the opposite is true.

7.2.5 Rank order – Merit/Pass

	Generic Learning					Principal Learning								Add/Spec		Result under each model						
	Eng	Ma	ICT	PLTS	Proj	1	2	3	4	5	6	7	8	1	2	Pass/Fail		Graded				
nlh	80	80	80	60	120	120	120	120	120	120	120	120	120	120	120	1.1	1.2	2.1.1	2.1.2	2.2.1	2.2.2	2.3
credits	8	8	8	6	12	12	12	12	12	12	12	12	12	12	12							
Cand																						
16	P	P	P	P	1	1	1	1	1	1	1	3	3	3	3	P	P	2	2	2	2	1
17	P	P	P	P	2	2	2	2	2	2	2	1	1	1	1	P	P	1	1	1	1	2
18	P	P	P	P	1	1	1	1	1	1	1	3	3	3	3	P	P	2	2	2	2	1
19	P	P	P	P	1	3	3	3	2	1	1	1	1	1	1	P	P	1	1	1	1	2

These four candidates also illustrate that the rank order depends on the grading model used.

7.2.6 Pass/Fail

	Generic Learning					Principal Learning								Add/Spec		Result under each model						
	Eng	Ma	ICT	PLTS	Proj	1	2	3	4	5	6	7	8	1	2	Pass/Fail		Graded				
nlh	80	80	80	60	120	120	120	120	120	120	120	120	120	120	120	**1.1**	**1.2**	**2.1.1**	**2.1.2**	**2.2.1**	**2.2.2**	**2.3**
credits	8	8	8	6	12	12	12	12	12	12	12	12	12	12	12							
Cand																						
20	P	P	P	P	1	1	1	1	1	1	1	1	1	1	1	P	P	1	1	1	1	1
21	P	P	P	P	1	1	1	1	1	1	1	1	1	1	0	F	P	0	0	0	0	1

In this table candidate 21 has failed just one unit but fails the overall Diploma under all but two of the grading models. The outcomes under Model 2.2.1 and 2.2.2 could be varied by changing the threshold (for Pass) in terms of the total tariff points. For these examples the system has been set up so that the Pass threshold is obtained by adding the tariff points of a candidate who has Pass for each unit (see (iv) earlier in Section 7.2).

As noted in point (ii) of Section 6, and illustrated by the A level English Literature example (see Figures 6.1f and 6.2), aggregating unit results expressed on a coarse scale is liable to produce overall grades significantly different from those obtained when aggregating the same results expressed on a finer scale. In GCE examinations (and in unitised GCSE examinations) unit results are expressed on a fine scale (known as the *uniform mark scale*[11]), while in some other qualifications, such as BTEC National, unit results are expressed on a coarse scale (eg Distinction/Merit/Pass). Current thinking on aggregation within the Specialised Diploma appears to be that a coarse scale will be used. For some units this system will mean that, although more precise data are available, they will be discarded. Table 7.3 below shows how overall grades might vary if a finer scale (such as a UMS) were to be used. The examples in the table show that a candidate's grade may vary under UMS depending on how well the candidate has performed *within* a grade and may be different from the grade obtained under the tariff points system of Model 2.

Table 7.3 Examples of candidate profiles and the possible grades if uniform marks are used for the aggregation

For consistency with Table 7.2, grades are again shown as 3 (Distinction), 2 (Merit), 1 (Pass) or 0 (Fail). See next page for an explanation of the uniform mark ranges.

	Generic Learning					Principal Learning								Add/Spec		Result under each model					
	Eng	Ma	ICT	PLTS	Proj	1	2	3	4	5	6	7	8	1	2	Graded using uniform marks	Graded using original models				
nlh	80	80	80	60	120	120	120	120	120	120	120	120	120	120	120		2.1.1	2.1.2	2.2.1	2.2.2	2.3
credits	8	8	8	6	12	12	12	12	12	12	12	12	12	12	12						
max um	11	11	11	8	41	41	41	41	41	41	41	41	41	41	41	492					
Cand																					
22	P	P	P	P	2	2	2	2	2	2	2	3	3	3	3	2 or 3	3	3	3	3	2
23	P	P	P	P	1	1	1	1	1	1	1	3	3	3	3	1 or 2	2	2	2	2	1
24	P	P	P	P	1	1	1	1	2	2	2	2	2	2	2	1 or 2	1	1	1	1	1
25	P	P	P	P	3	3	3	3	3	3	3	3	3	3	3	2 or 3	3	3	3	3	3
26	P	P	P	P	1	1	1	1	1	2	2	2	2	2	2	1 or 2	1	1	1	1	1

[11] Normally abbreviated to UMS.

In Model 2, each grade is associated with a specific number of tariff points. For example, in a unit with credit value 12, Pass has 12 tariff points, Merit 24 and Distinction 36. With a uniform mark scale (UMS), each grade is associated with a range of uniform marks. For example, in the same unit, 12 is now taken to be the mid-point of Pass, thus Pass is associated with the range 6-17 uniform marks. Similarly, Merit is associated with the range 18-29 uniform marks and Distinction with the range 30-41 uniform marks.

When aggregating using uniform marks, a candidate's overall grade depends not just on the grade obtained for each unit but on how well the candidate has performed *within* that grade. For example, for the unit in the previous paragraph a candidate may obtain a top Distinction, represented by 41 uniform marks, or a bottom Distinction, represented by 30 uniform marks. Therefore, as can be seen from Table 7.3, the same grade profile does not lead to a well-defined overall grade. For example, candidate 22 obtains 2 (Merit) or 3 (Distinction) under UMS, depending on how well he/she has performed within each unit grade.

To the extent that the various versions of the Model 2 aggregation method also produce either Merit or Distinction for candidate 22 and either Pass or Merit for candidate 23, the outcomes for these candidates under UMS are unremarkable. However, for candidates 24-26 the outcomes under UMS are less consistent with the outcomes under Model 2, as explained below.

Candidate 24
Pass under all versions of Model 2. However, unless the candidate receives the minimum uniform mark for each unit grade, he/she obtains Merit under UMS.

Candidate 25
Distinction under all versions of Model 2. However, depending on performance within the unit grades, the candidate may obtain Merit under UMS.

Candidate 26
Pass under all versions of Model 2. However, even with only middling performance within each unit grade, the candidate obtains Merit under UMS.

8 OTHER GRADING ISSUES

The previous sections have illustrated some of the difficulties and potential inconsistencies associated with aggregating the unit results in the Specialised Diploma. Further issues are discussed in this section. Many of these have been documented in previous reports and responses, for example in awarding body responses to the interim and final reports of the *Working Group on 14-19 Reform* and to the *Specialised Diploma Phase 2 consultation*.

8.1 Aggregation of disparate elements

Difficulties may arise from the aggregation of a wide variety of disparate elements. Where existing units/qualifications are used, these will have their own approaches to grading which cannot necessarily be validly used in combination.

8.2 Value of different units

The difficulty of assigning credit values to units was mentioned in Section 4 above, and attempting to grade units on a common scale for the purpose of aggregation would add a further complication. There is a risk that centres and candidates would concentrate on the elements which contribute to an overall grade, or on those which carry greater weighting. There would be no equitable way of defining appropriate relative values between grades in graded units and Pass in Pass/Fail units. The apparent arbitrariness of, and potential for controversy about, the tariff values for different units would threaten the credibility of an overall grade.

8.3 Value to users

If there were to be an overall grade, the aggregation model might not meet the requirements of users who were looking for different emphases. These users might determine their own criteria, based on unit outcomes, which could appear to be at odds with the official overall grades, leading to controversy and loss of credibility.

8.4 Regression to the mean

As discussed in Section 6 above, regression to the mean is likely to reduce the differences between candidates on the aggregated scale. The consequence will be either to place the majority of candidates in a single grade or to require large grade separation between candidates with very similar overall scores (eg there might be little difference, in overall performance, between the top of Pass and the bottom of Distinction). The first outcome would make the grading almost pointless, while the second would lead to accusations of a lack of reliability in the grades.

8.5 Curriculum 2000

There were major concerns expressed about the awards of the first Curriculum 2000 A levels in 2002, with allegations that the grading process was not properly carried out. These concerns were expressed in the context of an established qualification, for which the awarding bodies had an abundance of archive scripts

and statistical data to carry out grading. Grading of Specialised Diplomas would have to take place in the absence of any reference material. The grades would need to be determined from the constituent elements using a mechanistic and somewhat arbitrary procedure, with consequences which would be completely unpredictable. While one line of learning might produce 5% of candidates with the top grade, another might produce 30%. There would be no way of explaining or justifying such discrepancies.

8.6 The UCAS tariff structure

When bringing new qualifications into its tariff structure, UCAS uses a rigorous process in which these qualifications are benchmarked against an existing qualification in the structure by an expert group using a protocol designed specifically for this purpose by the University of Oxford Department of Educational Studies. However, there is a long lead-in period and only a limited range of qualifications is covered, nearly all at Level 3. It would be impracticable to carry out a similar exercise for the wide range of qualifications and units which are likely to be incorporated into the Specialised Diploma.

8.7 Other tariff structures for UK qualifications

The only other attempt to equate grades in disparate qualifications appears to be the work carried out by QCA on school and college performance indicators (QCA, 2002 and 2004a). At each level, qualifications have been assigned a 'size indicator' and a 'challenge indicator', and a points value or tariff has been determined for each grade on the basis of these two measures. However, because of the larger number of qualifications involved, the exercise has not been carried out with the same degree of rigour as the UCAS exercise, and the 2002 report acknowledges that, while the correspondence between grades and tariffs has been determined by consensus, it 'says nothing about the relative challenge between a Distinction grade in a particular qualification and an A grade in another …. by definition, qualifications at a particular NQF level can differ quite significantly within the broad band of challenge the level represents and the grade criteria may differ even more significantly'. Moreover, the relationships between the tariffs are not always the same as in the UCAS structure.

8.8 Other Diploma frameworks

There are few if any other diploma or certification systems which attempt to aggregate the results from potentially diverse sets of qualifications or units to produce an overall grade. Notably, the Welsh Baccalaureate is not graded, and many other European diploma-type qualifications do not have an overall grade. Exceptions include the following.

(a) France

Candidates receive an overall score (as a mark out of 20) for the *baccalauréat*, with a pass mark of 10 and 'grades' awarded for higher levels of performance (*'assez bien'* for 12 or 13, *'bien'* for 14 or 15 and *'très bien'* for 16 or more). The overall score is a weighted average of a candidate's scores

in the various elements of the *baccalauréat*, with a complex system of weightings designed to take greater account of the candidate's specialist areas. Although a diverse set of elements may contribute towards a candidate's *baccalauréat* result, they are all scored in the same way – the constituent elements do not have different scoring systems. Also, this is a long-standing qualification which has evolved and grown over time.

(b) *International Baccalaureate*
Like the French *Baccalauréat* this is a grouped award with an overall grade. Candidates are awarded points on a scale of 1-7 for each of the six subjects taken as part of the baccalaureate and points on a scale of 1-3 for the Theory of Knowledge and extended essay combined, giving an overall maximum of 45 (= 6x7 + 3). In this qualification the elements being aggregated are homogenous in the sense that they are all traditional academic subjects (apart from the Theory of Knowledge + extended essay, which has relatively low weighting), and they are all scored in the same way. Nevertheless, it must be questionable whether a points total of 36, say, in one combination of subjects is comparable to the same points total gained in a different combination of subjects.

9 CONCLUSION

The issues which are often put forward in the debate about whether or not there should be an overall grade for the Specialised Diploma include the following.

(1) During consultation, users and providers in particular were strongly in favour of the provision of an overall grade.

(2) While it is acknowledged that an overall grade may provide an *incomplete* picture, the transcript will be available to provide more detailed information.

With regard to point (1), the views of users and providers may need to be interpreted with caution. There is a danger in asking a simple question without also pointing out the risks and possible drawbacks. An analogy would be to ask householders if they would like their Council Tax bills to be reduced, to which the obvious answer is 'yes'. Of course, in this analogy respondents would be more aware of the potential drawbacks (deterioration of local services, etc), and these considerations will affect their answers. With a question about grading the Specialised Diploma, most respondents will not be aware of the potential drawbacks unless these are made explicit.

With regard to point (2), the modelling in Section 7 of this paper shows that the information provided by an overall grade would be *misleading* rather than merely *incomplete*. A candidate could obtain different overall grades depending on the grading model adopted, and in some cases the overall grades of two candidates do not reflect their relative merits suggested by their unit profiles.

The Welsh Baccalaureate is not graded. Provided that candidates meet the minimum requirements they are awarded an Intermediate (Level 2) or Advanced (Level 3) Diploma. Thus, the Diploma has a similar function to the threshold 'five GCSEs at A*-C or equivalent', except that it is more useful because of the specific core studies which candidates have to cover. Users who wish to find out more about a candidate's attainment can look at the grades for individual elements. The Specialised Diploma could usefully serve a similar function.

It is strongly recommended that the success of the Diploma will be more assured if there is no grading of the overall result, with the quality of a candidate's achievement being provided by his/her grade profile across the units and by the evidence in his/her transcript.

REFERENCES

Department for Education and Skills (2005a) *14-19 Education and Skills*

Department for Education and Skills (2005b) *14-19 Education and Skills Implementation Plan*

Qualifications and Curriculum Authority (2002) *Including all approved qualifications in school and college performance indicators: Phase one report* available at http://www.qca.org.uk/14-19/developments/downloads/Phase_One_Report2.pdf, accessed 4 April 2006

Qualifications and Curriculum Authority (2004a) *Including all approved qualifications in school and college performance indicators: Phase two report* available at http://www.qca.org.uk/14-19/developments/downloads/phase_2_report.pdf, accessed 4 April 2006

Qualifications and Curriculum Authority (2004b) *Principles for a credit framework for England* available at http://www.qca.org.uk/downloads/ffa_principlescreditframework.pdf, accessed 4 April 2006

Qualifications and Curriculum Authority (2006a) *Specialised Diploma: document for Phase 2 of consultation to support the development of grading, assessment and standards*

Qualifications and Curriculum Authority (2006b) *The Specialised Diploma: qualification structure*

Scottish Credit and Qualifications Framework (2003) *An Introduction to the Scottish Credit and Qualifications Framework* available at http://www.scqf.org.uk/downloads/IntrotoSCQF 2ndEdition.pdf, accessed 4 April 2006

Working Group on 14-19 Reform (2004) *14-19 Curriculum and Qualifications Reform: Final Report of the Working Group on 14-19 Reform* Department for Education and Skills