



Neutral Citation Number: [2013] EWHC 211 (Admin)

Case Nos: CO/11409/2012 AND CO/11413/2012

**IN THE HIGH COURT OF JUSTICE**  
**QUEEN'S BENCH DIVISION**  
**ADMINISTRATIVE COURT**

Royal Courts of Justice  
Strand, London, WC2A 2LL

Date: 13/02/2013

**Before :**

**LORD JUSTICE ELIAS**

**- and -**

**MRS JUSTICE SHARP**

-----  
**Between :**

**THE QUEEN**  
**(on the application of**  
**LONDON BOROUGH OF LEWISHAM & ORS)**

**Claimants**

**- and -**

**(1) ASSESSMENT AND QUALIFICATIONS**  
**ALLIANCE (“AQA”)**

**(2) PEARSON EDUCATION LIMITED**  
**(“EDEXCEL”)**

**(3) OFFICE OF QUALIFICATIONS AND**  
**EXAMINATIONS REGULATION**  
**(“OFQUAL”)**

**Defendants**

**- and -**

**(4) OXFORD AND CAMBRIDGE and**  
**RSA EXAMINATIONS t/a OCR (“OCR”)**  
**(5) WJEC**

**Interested**  
**Parties**

-----

**Mr Clive Sheldon QC, Ms Joanne Clement and Mr Joseph Barrett**

(instructed by **LB of Lewisham Legal Services**) for the **Claimants**

**Mr Clive Lewis QC and Ms Jane Oldham**

(instructed by **Eversheds LLP**) for the **Defendants AQA**

**Mr Nigel Giffin QC and Mr Christopher Knight**

(instructed by **Herbert Smith Freehills LLP**) for the **Defendants EDEXCEL**

**Ms Helen Mountfield QC, Ms Sarah Hannett and Mr Raj Desai**

(instructed by **Wragge & Co LLP**) for the **Defendants OFQUAL**

Hearing dates: 11-13 December 2012

-----

**Approved Judgment**

**Lord Justice Elias :**

*Setting the scene.*

1. This case involves two claims for judicial review brought in relation to the award of GCSE English qualifications in August 2012. In England and Wales, those qualifications are awarded by four different awarding organisations (“AOs”) under the supervision of the statutory regulator, the Office of Qualifications and Examinations Regulation (“Ofqual”). Each of the judicial review claims is brought against just one AO. The first is against Assessments and Qualifications Alliance (“AQA”) and the second against Pearson Education Ltd (operating under its trading name “Edexcel”). Ofqual is a second defendant to each claim. The other two AOs were named as interested parties; OCR served an Acknowledgement of Service but WJEC took no part in the proceedings. This is a rolled-up hearing; technically the court must first consider whether to grant the claimants permission to pursue the claim and if it does, it must then determine the substantive merits. We have not, however, dealt with the permission question as a preliminary issue but only after having considered fully the merits of the legal challenges.
2. New GCSE examinations were introduced in all subjects in 2009 and 2010. The new set of GCSE English qualifications was first taught from September 2010, and qualifications were awarded for the first time in the summer of 2012. In place of courses in English and English literature, three courses were provided in English, English language and English Literature. The English GCSE included both language and literature elements.
3. There are three relevant features to note about the new qualifications when compared with their predecessors. First, changes were made to the weightings of external and internal assessment. Internal assessment increased to 60% from 40%, and traditional coursework was replaced with “controlled assessments”. This was partly at least because of concerns that the course work may not always have been the student’s own unaided work, or may have been plagiarised. Examination boards produce a range of controlled assessment tasks. Teachers select the assessment which is then carried out by all the students and is conducted in the classroom under supervised conditions. The controlled assessments are marked by teachers internally although they are subject to moderation by the relevant AO. The 40% subject to written examination is marked externally by examiners appointed by the AO.
4. Second, the new courses were modular. This meant that students were able to take examinations or submit controlled assessments at various points during the course, or at the end of it. For the course which began in September 2010, assessment dates were in January 2011, June 2011, January 2012 and June 2012. Each school could choose the order in which students studied and completed units and when to take the examination or be subject to an assessment. However, this freedom was subject to what is termed the “terminal rule” which required candidates to complete at least 40% of the course at the terminal date in June 2012. This could be by way of written examination or controlled assessment. The written examinations would necessarily change for each assessment date, but the assessment topics were identical for all candidates at a particular school. Candidates could re-sit a unit once and take the better result (unless it was taken to satisfy the terminal rule).

5. Third, examinations and assessments were marked after each January/June date and the marks and grade boundaries were made public. The raw marks given by examiners for the scripts or assessments do not equate directly to grades. The grade boundaries are set by the AO after the raw marks have been determined.
6. The consequence of publishing marks and grades at each stage of the process is that individual candidates and their teachers know after each module what raw marks they achieved and how that translated into a grade for the particular unit. Many teachers assumed that the boundary mark between grades C and D would be the same, or at least almost the same, from one assessment date to the next. A central issue in this case is whether they were led to believe that this would be so, and whether it would in any event be fair and lawful for an AO to adopt significant differences in the grade boundary for a particular unit from one assessment to the next.
7. There are more than one hundred and fifty claimants represented in the two actions and they include local authorities, schools, teachers and pupils. They share a widespread and deeply held grievance over the way in which the boundary between grade C and grade D was fixed in the English GCSE examinations and controlled assessments assessed in June 2012.
8. This boundary between C and D is a particularly important one for students, schools and teachers alike. For students it may be crucial to their chances of being qualified to go into further education or achieve apprenticeships; and for teachers and schools who are subject to increasing accountability, the proportion of students attaining the C grade in English is one of the more important measures of their success. Furthermore, many teachers quite properly take professional pride in their ability to judge performance and to determine whether a student is of the requisite standard for a C grade or not. If fewer students secure at least a C grade than anticipated, their judgment is in question, and the results may be damaging to the standing of the school and the teacher.
9. The claimants' complaint is that too rigorous a standard was adopted when assessing some of the units in June 2012 with the result that many pupils who confidently and reasonably expected to attain the C grade, on the basis of results which their fellow examinees had obtained in the January 2012 and indeed earlier assessments, inexplicably failed to do so. There was an unheralded and unjustified shift in the grade C boundary. This constituted an elementary unfairness because pupils competing in the same examination were not treated equally. The January cohort of students was graded more leniently than the June cohort, at least in some of the papers assessed by the two AOs. Ofqual, as the regulator, had power to forbid this inconsistent and unfair treatment by issuing statutory directions, and its failure to do so in order to remedy this conspicuous unfairness constituted an error of law.
10. This unfairness was, say the claimants, compounded by two further factors. First, both the AOs and Ofqual had led the pupils and their teachers to understand that the marking standard would be consistent at whatever stage in the two year cycle a unit was completed. The natural inference from this was that in relation to any particular unit, the same, or at least substantially the same, grade boundary would be adopted in June as in the previous January. It is conceded that everyone understood that there might be some minor variation in the mark boundary for written examination papers to reflect the fact that a particular paper may vary in difficulty from one half-yearly

assessment to the next. The marks will then be correspondingly higher or lower depending upon whether the paper is easier or harder and the grade boundary will need to be adjusted accordingly, but no radical change would have been anticipated in such cases. For controlled assessments, where the task remains precisely the same whenever the unit is completed, there is no justification in changing the grade boundary at all. Mr Sheldon QC, counsel for the claimants, submits - and this is not disputed - that many pupils and teachers had acted on that assumption to their detriment. In some cases, for example, there is evidence that once teachers were confident that a student would achieve a C grade on the basis of previous grade boundaries, the student was encouraged to switch focus to other subjects.

11. Second, the claimants allege that the AOs had wrongly given effect to what was in substance, if not in form, a direction from Ofqual requiring them to fix their June grade boundaries by reference to the predicted results for the particular batch of students. Whatever Ofqual's intentions, in practice the AOs acted as if Ofqual was requiring them to set the grade boundary so that the number of students obtaining the C grade did not exceed the predicted number by more than 1% (the "tolerance limit"). The effect was artificially and unfairly to pitch the pass mark for the C grade too high. Insufficient credit was given for the qualitative performance of students in the relevant assessment exercises, and assessments were improperly dominated by quantitative statistical evidence of dubious validity purporting to predict the likely pattern of results for the particular cohort of students. In somewhat colourful language, Mr Sheldon claimed that there had been "an illegitimate grade manipulation as a result of a statistical fix".
12. The defendants reject these criticisms. Each of the AOs claims that the work was assessed in June by adopting precisely the same procedures as had been employed in January. With the benefit of hindsight they concede that it may indeed be the case that the January students were treated more generously than they ought to have been, although this was not apparent at the time. But this was not because of any difference in approach to the assessments. It was a consequence in part of the fact that fuller and more precise information, particularly statistical information predicting likely performance, was available in June than had been available earlier. It would have been wrong for the examiners to have ignored this material. If the examiners had simply applied the January grade boundaries in the June assessments, as the claimants contend that they ought to have done, this would have led to a dramatic increase in the number of students gaining grade C when compared with earlier years. Moreover, it would have been unjust to students in subsequent years unless they too were to be beneficiaries of this striking grade inflation. But that would dilute the value of the qualification. The defendants are adamant that the June examinees received the right grades even if some of the January cohort perhaps received more than their due.
13. More specifically, each of the defendants categorically denies that they ever represented that they would apply the same, or almost the same, grade boundaries in June as had been adopted in January. On the contrary, it was always made plain that the grade boundaries might vary from one assessment to the other. If teachers and students acted on any other assumption, that was unfortunate but it was not the fault of the AOs or Ofqual. Nor did the AOs act on the assumption that there was a direction from Ofqual which slavishly had to be followed. There was guidance from Ofqual which did inform the decisions of the AOs, but they treated it as guidance and

no more than that. There was no unfairness in the process. Indeed, considerable care was taken to ensure that the pupils were treated fairly and consistently.

14. The AOs further contend that since they are non-governmental bodies providing services for reward under private law contracts, they are not amenable to judicial review at all. Ofqual is the statutory regulator and accepts that it is amenable to judicial review. The AOs contend, moreover, that the rationale of the regulatory scheme is that any complaints about the operation of the arrangements should be directed to Ofqual who can be challenged by way of judicial review if they fail lawfully to deal with those complaints. That should provide effective relief, and it is both unnecessary and wrong in principle for the AOs to be parties to these proceedings at all.
15. This delineates the contours of the principal areas of dispute. There is also a distinct argument that in determining not to follow the January marking scheme, and choosing instead to let the higher standards apply in June, each of the defendants failed to carry out the public sector equality duty imposed by section 149 of the Equality Act 2010. I will deal with that argument at the end of this judgment.

*The facts.*

16. In order properly to analyse the merits of the respective arguments, it is necessary to set out in some detail the factual background. The court has been provided with extensive documentary material and many detailed witness statements from all parties constituting some 16 lever arch files. I shall try and summarise only the essential material necessary properly to understand and assess the legal arguments. In the course of this factual analysis it will be convenient to deal with one of the hotly disputed factual issues which in part underpins the claimants' case, namely whether the AOs wrongly considered themselves to be bound by the guidance given by Ofqual with respect to the tolerance limit.

*Ofqual and the assessment system.*

17. Ofqual is a non-Ministerial Government department, directly accountable to Parliament, with responsibility for regulating AOs. The AOs award certain specified recognised academic or vocational qualifications (but not degrees) in England, including the GCSEs. Ofqual was established on 1 April 2010 as an independent public body by the Apprenticeships, Skills, Children and Learning Act 2009 ("ASCLA").
18. ASCLA confers upon Ofqual a number of functions. These must be performed by reference to the five statutory objectives set for Ofqual in section 128(1) ASCLA. These are (a) the qualifications standards objective; (b) the assessments standards objective; (c) the public confidence objective; (d) the awareness objective; and (e) the efficiency objective.
19. There is no order of priority between these objectives, and not all of the objectives will necessarily be engaged in any particular case. Moreover, Ofqual has a broad discretion about how to achieve these objectives; it must "so far as is reasonably

practicable” act in a manner which is compatible with its objectives, and “which it considers most appropriate for the purpose of meeting its objectives” (section 129(1)(b)).

20. The two objectives which figure significantly in this case are the qualifications standards objective, and the public confidence objective. The former, set out in section 128, is designed to ensure that standards are maintained at a consistent level year on year and that there is consistency between the standards applied by different providers. In short, the currency of the qualification must not be debased and must not vary depending upon which AO is responsible for awarding the qualification.
21. The “public confidence objective” is defined in section 128(4) as being “to promote public confidence in regulated qualifications and regulated assessment arrangements”. Of course, maintaining the currency of the standard is an important element in maintaining public confidence in the system.

*Ofqual and the AOs.*

22. By section 132(1) ASCLA, Ofqual must recognise an awarding body in respect of certain defined categories of qualifications provided that the awarding body has applied for recognition and meets the relevant criteria for recognition. Ofqual is required to publish these criteria: section 133.
23. Recognition is subject to “the general conditions” which Ofqual sets and publishes (pursuant to sections 132(3) and 134(1)). These are defined in subsection 132(8) as “the general conditions for the time being in force under section 134 which are applicable to the recognition and the body”. The current General Conditions of Recognition (“COR”) were published in 2012.
24. By virtue of Conditions B7 and D5 of the COR 2012, the AOs who award GCSE and GCE (A levels) must comply with the principles contained in the *GCSE, GCE Principal Learning and Project Code of Practice*, published by Ofqual in May 2011 (“the Code”).
25. The Code is designed to provide practical guidance to assist AOs to achieve Ofqual’s objectives, and in particular to promote quality, consistency and fairness in the assessment and awarding of qualifications, and to ensure the maintenance of consistent standards, both within and between AOs, and from year to year. The Code sets out principles and practices for achieving these objectives and confirms that each AO’s governing body is responsible for setting in place appropriate procedures to give effect to them.
26. Section 151(1) ASCLA provides Ofqual with a power to issue directions where an AO “has failed or is likely to fail to comply with a condition to which the recognition is subject”. Specifically, by subsection (2), Ofqual may “direct the recognised body to take or refrain from taking specified steps with a view to securing compliance with the condition”. Such directions are enforceable on application by Ofqual to the High Court for a mandatory order: see section 151(7). Since one of the conditions is to comply with the principles in the Code, it follows that Ofqual can impose a direction requiring compliance with those principles if it considers that an AO is departing from them.

27. The AOs themselves are not statutory bodies. Edexcel is a private corporation answerable to its shareholders in the usual way; AQA is a charitable body. An AO enters into contracts with centres (which are typically schools) but not with the students themselves. No school is compelled to contract with one AO rather than another and there is a degree of market competition between AOs as to the quality of the service provided. Schools may contract with more than one AO for different qualifications. The payment of the registration fee entitles the school to enter its candidates for that particular AO's qualifications. The AO then has contractual obligations to provide services to the school necessary to enable the pupils to seek the relevant qualification. These services include not only the awarding of the qualifications but also the creation of syllabuses, the setting and provision of examination papers, and the marking of scripts. The contract provides, amongst other things, for the availability of appeal and complaints procedures in individual cases.

*Carrying out the assessments.*

28. As I have said, at the heart of this case are complaints about the processes of assessment by both AQA and Edexcel and the failure of Ofqual to correct what are alleged to be obvious injustices. However, although the nature of the challenge is similar with respect to each of the AOs in that both are said to have unfairly assessed units comprised in the English GCSE in June 2012, the particular grievance is different with respect to each. It is, therefore, necessary to consider the circumstances of each case separately.

*The Ofqual requirements.*

29. The assessments must comply with the principles set out in the Code. Section 6, entitled "Awarding, maintaining an archive and issuing results", sets out in considerable detail how AOs are to determine grade boundaries for the different units making up the qualification. In summary form they are as follows.
30. An AO is required to appoint an Awarding Committee which is responsible for checking that the required standards are brought to bear in each unit and for the qualification as a whole. The Committee must be chaired by the chair of examiners and include the chief examiner (responsible to the chair for the examination as a whole), the principal examiner (responsible for setting the paper and standardising its marking) and the principal moderator (responsible for each internally assessed unit). The Committee will have available information about the marks which (after moderation) have been thought appropriate for the candidates. These are assessed by examiners by reference to published grade descriptions which specify the skills and qualities necessary to achieve a particular mark. The Committee is then given a raft of information designed to enable it properly to inform its grading decisions. This includes information about grades in previous equivalent examinations, including examination papers and scripts exemplifying grade boundaries for those earlier awards; the appropriate range of candidates' work to enable grade boundaries to be assessed properly; and information, based on a preliminary calculation of outcomes, about where problems of consistency and comparability may arise.
31. The information available to the examiners is therefore both qualitative evidence relating to the assessment itself, and quantitative information including a variety of technical and statistical data. The AO staff will in advance of the meeting have



identified the range of marks within which they anticipate that the grade boundary will lie. They must provide the Awarding Committee with scripts reflecting performance within that range. The Awarding Committee is not, however, bound by the preliminary assessment and can ask for more scripts within a different range if it chooses. The Code provides that the determination of the provisional boundary suggested to the Awarding Committee should be based on the available statistical and technical data. The most important quantitative information is statistical information predicting the performance of the particular cohort of candidates based on past performance. In addition, the Principal Examiner and the Principal Moderator may be asked for a preliminary recommendation of the proposed range of marks (although this is not compulsory under the Code) and where they have been requested to do so, the determination should be informed by their recommendations.

32. The Code then sets out how the Awarding Committee is to carry out the task of forming a judgment on the appropriate grade boundary with respect to any particular unit. Broadly it is as follows. Each member of the Committee first works independently assessing each of the provided scripts and fixes what he or she considers to be the appropriate mark to reflect the grade boundary; this is done for the A, C and F boundaries. The chair then identifies the boundary after considering the lowest mark where there is consensus for a particular grade, and the highest mark where there is consensus that the mark does not justify that grade, and then fixes a mark, in the light of all the evidence, which the chair judges to be the appropriate boundary. The chair's recommendations are then considered by the officer of the AO with overall responsibility for the quality and standard of qualifications to ensure consistency. That officer may accept or vary the chair's recommendations and will subsequently make a final recommendation to Ofqual. Ofqual may approve the recommendation or give reasons for being dissatisfied with it. In the latter cases the AO must then reconsider and provide a final report. Ultimately, Ofqual can issue a direction to prevent the AO fixing what Ofqual considers to be unjustifiably high (or low) grade boundaries.
33. The policy which Ofqual required AOs to adopt with respect to all GCSEs in 2012 is what it termed "comparable outcomes". It described that concept as follows:

"In general the principle we have applied in setting standards for new qualifications is that a student should get the same grade as they would have done had they entered the old version of the qualification. We call this approach 'comparable outcomes'. It aims to prevent what is sometimes called grade inflation – that is, increases in the numbers of students achieving higher grades where there is not sufficient evidence of real improvements in performance. It also enables us to allow for a dip in performance. It can arise when the new qualification is first taken."
34. This policy, which is designed to promote the statutorily defined qualification standard objective, is wholly at odds with the suggestion advanced in argument by the claimants that the new GCSEs were designed to lead to improved results. On the contrary, comparable outcomes is designed to ensure that standards remain consistent

and without grade inflation. Moreover, Ofqual had made it clear that in so far as there was any conflict between comparable outcomes and comparable performance, priority would be given to the former. It was plainly legitimate for Ofqual to define its objectives in that way.

35. In practice, comparable outcomes can only be achieved where the cohort for the subject is similar in terms of ability to previous years; where there is no reason to suppose that the previous grade standards were inappropriate; and where there is no substantial improvement or drop in the quality of teaching and/or of learning at a national level. For a core subject like English, where a large number of candidates each year take the examination and teaching methods vary little year on year, these conditions will in practice be met.
36. The information which Ofqual requires should be made available to the Awarding Committee is designed to facilitate achieving comparable outcomes. The particular cohort of candidates qualifying in June 2012 was compared with the 2010 cohort (the last year when qualifications were awarded under the old system). Broadly, one year's cohort might be expected to achieve similar results to any other, given the significant number of candidates. However, there may be some variation in quality year on year. In order to reflect that possibility, the likely outcome of the 2012 group is achieved not simply by comparing them with the 2010 candidates but also by analysing the prior attainment for each of those two cohorts in earlier exams. In the case of GCSEs, Ofqual requires a comparison between the performance at GCSE and that same cohort's attainment at Key Stage 2. Key Stage 2 gives the results for that particular group of candidates (or at least a substantial number of them) in tests taken in their last year at primary school five years earlier. If, say, the candidates doing the GCSE in 2010 performed better at Key Stage 2 than the cohort doing the examination in 2012 then the evidence suggests that it would be reasonable to infer that the 2012 cohort would be likely to perform less well than the 2010 cohort in their GCSEs. In that way the predicted outcomes are designed to cater for the fact that some years may be of superior quality to others.
37. Mr Sheldon was critical of the use of KS2 predictions and he referred to observations from experts in the field who have considerable reservations about the legitimacy of relying on performance in KS2 as a guide to performance some five years later. No doubt they are not fool-proof. Nonetheless, they have been used in previous years and are widely thought to be the most reliable statistical evidence currently available for the purpose of comparing performance year on year. In those circumstances the court cannot in my view possibly say that it was an error of law for Ofqual to require all AOs to base predictions in part on the earlier KS2 results. Moreover, a common approach to prediction is important to ensure consistency as between the different AOs.
38. With the relevant statistical information, it is possible to identify the boundary which would place the appropriate proportion of candidates in grade C in accordance with the predicted outcome. That boundary is known as the 'Statistically Recommended Boundary' (SRB).
39. Ofqual is concerned if the actual grading departs in any significant way from the predicted outcomes, because that is likely to demonstrate that the standard has been set either too high or too low which would undermine the comparable outcomes

policy. Accordingly, in June 2012 Ofqual required AOs in respect of all GCSE subjects to report any case where the actual outcome differed from the predicted outcome (the SRB) by more than a margin referred to as the “reporting tolerance”. That tolerance varied depending on the number of entrants for the examination, but where it was a large number, as in the case of English, the reporting tolerance was fixed at 1%. The obligation is only to report results outside the tolerance limits; if grade boundaries lead to results departing from that tolerance they may still be permitted by Ofqual but the AOs will need to justify them. Ofqual must be satisfied that they genuinely reflect an improvement in standards, and not merely performance. The reporting obligation had not been in place with respect to assessments made before June 2012.

40. The claimants’ original case was that it was not legitimate to use statistical material at all when assessing grade boundaries, and they asserted that the Key Stage 2 information had been used to fix a quota for those who could obtain grade C. However, they subsequently resiled from those submissions and now accept that predicted results can properly be used as a guide; they can inform the decision as to the appropriate grade boundary. However, they now say that in practice the statistics dominated the analysis and diminished the importance of the qualitative assessment of the scripts. Academic judgment was no longer determining grade boundaries, and the reporting tolerance was in practice decisive in fixing them. This requires a consideration of precisely how the grade boundaries were fixed by the two defendant AOs.

*AQA and Edexcel: disputed papers and the marking differentials.*

41. As I have said, the complaint concerns certain papers from both AOs in which the marks required in June were higher - and the claimants say significantly higher - than the raw marks required in January.
42. AQA was responsible, in broad terms, for some 60% of the English candidates overall. Its GCSE English specification comprised three distinct, modular Units:
  - (1) Unit 1 is entitled “Understanding and producing non-fiction texts”: this carried 40% of the total marks available for the course and was assessed by unseen written external exam (marked out of 80 marks). It could be taken as part of a foundation or higher tier papers. The difference is that the foundation tier paper allows students to achieve only grades C-G or an unclassified (“U”), whereas the higher tier allows for grades A\*, A and B also. It is given the code ENG1F.
  - (2) Unit 2: “Speaking and listening”: this carries 20% of the total marks available for the course and is assessed by controlled assessments conducted in the classroom under supervised conditions. It is marked out of 45 marks (and given the Code ENG02).
  - (3) Unit 3 entitled “Understanding and producing creative texts”: this carries 40% of the total marks available for the course and is assessed by controlled assessments. It is marked out of 90 marks (and given the Code ENG03).

43. Before the decisions under challenge were taken, AQA had offered Units 1 and 2 in January 2011, June 2011 and January 2012. The Unit 3 controlled assessment was not available in January 2011 but had been previously submitted for assessment and award by some candidates in June 2011 and January 2012.
44. The changes to the raw marks in these different units between January and June were as follows:
  - (1) In June 2012 AQA students needed to obtain 53 raw marks - 10 more raw marks (out of a total of 80) on the Unit 1 Foundation Tier examination papers (ENG1F) than their peers who took the paper in January 2012 (43 marks). This was also significantly higher than their peers who took this paper in June 2011 (44 marks required) and January 2011 (46 required).
  - (2) Students needed to obtain 28 raw marks (out of a total of 45) in Unit 2; and 54 raw marks (out of 90) in Unit 3 to obtain grade C in those units. In each case they had to obtain 3 more marks than their peers who submitted Controlled Assessments in response to identical tasks in January 2012 and June 2011.
45. Edexcel was responsible for some 10% of the candidates overall. Its GCSE English specification also comprises three distinct, modular Units:
  - (1) Unit 1, entitled “English Today”: this carries 20% of the overall marks available for the course and is assessed by way of controlled assessment. The total number of marks available is 40. Unit 1 has the code “5EH01”. It is common to both the GCSE English and GCSE English Language qualifications.
  - (2) Unit 2, “The Writer’s Craft”: this carries 40% of the overall marks available for the course and is by way of a 2 hour written examination. Foundation and higher tier exams are available. The total number of marks available is 96. This unit has the code 5EH2F (Foundation) and 5EH2H (Higher)
  - (3) Unit 3, “Creative Responses”: this carries 40% of the overall marks available for the course and is assessed by controlled assessments conducted in the classroom under supervised conditions. It includes three speaking and listening tasks (which account for half the marks), one poetry reading task and one creative writing task. It is also marked out of 96 marks. This unit has the code 5EH03.
46. The first assessment opportunity for Unit 1 took place in January 2011, and in each following June and January. The first Unit 2 and Unit 3 assessments were not taken until June 2011, and thereafter in January and June 2012.
47. The complaint with respect to these assessments is again that the pass mark for grade C was significantly higher in June than it had been in January. There is no complaint about Unit 1 where the boundary remained the same on 24 marks. However, a June candidate needed to secure:
  - (1) 74 and 42 raw marks -- 8 more raw marks on each of the Unit 2 Foundation and Higher Tier examination papers (5EH2F and 5EH2H) respectively than

their peers who took the paper in January 2012 (where the boundary was 66 and 34 respectively);

- (2) 65 raw marks - 10 more raw marks on the Unit 3 Controlled Assessment (5EH03) than their peers who were assessed in January (55) or indeed, in June 2011.

48. The question is why these boundaries appear to have been fixed so much higher in June than in the previous January. The claimants focus in particular on the AQA Unit 1 paper where the increase was ten marks, and the Edexcel Unit 3 controlled assessment where again the grade boundary increased by ten marks. They contend that the latter increase in particular is simply inexplicable given that the tasks did not change save on the premise that a different standard is being applied. This requires a consideration of the way in which the grade boundaries were fixed by the two AOs in January and June.

*The fixing of grade boundaries by AQA.*

49. The court received written statements from two AQA officers. Ms Meadows was the Director of Education, Research and Policy in AQA and was the approval officer for the units sat in January and June 2011 and January 2012. She did not have that role in June 2012 but she did attend part of the Awarding Committee meeting. Mr Michael Jones was the Chair of the examiners in English in both January and June 2012. He gave information about precisely how the grade boundaries were set. In addition we were shown a range of documents concerning the approach of the Awarding Committee.
50. In January 2012 some 54,000 students took the Unit 1 ENG1F paper compared with some 140,000 the following June. The statistically determined SRB for grade C was 41 and the Principal Examiner suggested a range of 39-46. Scripts within that range were in fact considered. The Awarding Committee concluded that a mark of 44 was appropriate, and Ms Meadows, as the accountable officer, agreed.
51. For Unit 3, ENG03, the Principal Examiner recommended a range of 49-53 marks and in the event the C grade boundary adopted was the same as that chosen in June 2011, namely 51. The number of candidates taking this unit was only 1,231 compared with 97,000 who would take it in June. Ms Meadows explains in her statement that there was a concern that the SRB would be unreliable and should not be used as a guide to determine the grade boundary. There were two main reasons for this. First, the statistical relationship between Key Stage 2 and the outcome of particular units is weakened when applied to candidates taking the units at an early stage in the life of a new course. Second, the pool of candidates was small and not necessarily representative. Students taking the course early might be very bright, or they might be weak but take the assessment early knowing that they can re-sit if they do not do well. This was not a rejection of the statistical evidence in principle, but rather of its reliability when assessing that unit.
52. Similarly in ENG02, the speaking and listening teacher controlled assessment, the proportion of candidates taking the examination was small - some 15000 candidates compared with 380,000 in June – and the SRB in January was not considered to be

reliable. The starting point for considering the grade mark was fixed at 25, the same as in June 2011. That was considered to be a more reliable indicator than the SRB, and this was the mark in fact ultimately recommended to Ofqual and selected.

53. Mr Jones contended that there was in principle no difference in the way in which AQA approached its task of setting grade boundaries in June 2012 from the way it had done so in relation to earlier assessments in 2011 and 2012. The principal difference in practice was that the statistical information was much fuller and more reliable in June. The process can be described by reference to the fixing of the boundaries in June 2012.
54. So far as the boundary for ENG1F was concerned, the SRB had been determined by a reference to the Key Stage 2 results, as the Code requires. It had been calculated for grade C at 52 marks, 11 marks more than the SRB in January. Accordingly, examination scripts within a range of 50 to 54 marks were selected by staff in accordance with the Code of Practice and put before the Awarding Committee.
55. The members of the Committee worked independently as required and each reached a provisional mark as being appropriate for the boundary. The view of the Committee as a body was that 52 did not merit a C grade, albeit that this was the SRB. However, they concluded on balance that 53 did merit it and 54 was seen as a secure C grade. Accordingly, 53 was recommended by the Chairman. Although the mark was higher than that anticipated by the predictions, this was not unusual; it had also been the position with relation to the boundaries fixed in the January 2011 and June 2012 assessments.
56. In the controlled assessment, ENG03, where written folders of work were available, the Committee were given an SRB of 54. The consensus was that this merited the C grade, whereas work marked at 53 did not. So this was the boundary recommended by the Chairman.
57. As to the third Unit, ENG02, the speaking and listening requirement, there is no record of the candidates' performance of that assessment because no formal document is used. The teachers simply give marks in accordance with guidance from Ofqual, although the marking scheme is moderated.
58. In January 2012 the mark for a C grade had been fixed at 25. It was appreciated that if this same boundary were to be applied in June, there would be a very large discrepancy between the results achieved in the written examination and the written controlled assessment compared with the speaking and listening assessment. Over 86% candidates would have achieved a C grade in that unit, compared with 67% in English as a whole.
59. The Awarding Committee considered that there was no basis for assuming that the candidates would be relatively better at speaking and listening than the other skills, and concluded that teachers had been over-marking. Indeed, they considered that some had perhaps been marking strategically, assuming that the mark for a C grade in January would be replicated in June. The SRB for this Unit was 29 and this was the recommendation of the Chairman. Subsequently, however, he agreed to reduce that boundary to 28 and this was the boundary approved by the accountable officer.

60. Mr Jones claimed that in every case the boundary was fixed by a genuine exercise of academic judgment and the Awarding Committee was satisfied with its assessment. The claimants say that this statement should be viewed with considerable scepticism. They contend that it is not at one with the objective evidence. It is submitted that in practice AQA allowed the tolerance limits to dictate their approach to the assessment of the boundaries.
61. Mr Sheldon relied on a number of factors to support this submission. First, the principal examiner had given indicative marks in the range of 44 to 48. Yet the subsequent boundary was very significantly higher than that; a candidate with 44 would be 9 full marks off a C grade. No explanation has been given as to why the professional principal examiner should be so off the mark. Indeed, AQA's own procedures state that scripts in the Principal Examiner's recommended range should be considered by the Awarding Committee, but that was never done.
62. Second, Mr Sheldon pointed out that the Minutes of the Awarding Committee Meeting report that Mr Jones had said that all boundary grades were "subject to a very tight and limited flexibility". This, say the claimants, is not consistent with a recognition that the boundary could be fixed outside the tolerance limits. He also referred us to a number of other statements which suggested that the authors had interpreted the tolerance limit as binding.
63. Third, the Chairman stated that he had in the earlier January and June 2011 assessments compared the performance by reference to scripts in the 2010 final examination. Taking that same reference point for the June 2012 papers, it is difficult to see why there should be such a marked hike in the grade boundaries since the paper was not significantly easier. Mr Sheldon relies on the fact that the principal examiner for AQA had reported that overall the demands of this paper were similar to those of papers in earlier years. Mr Sheldon said that this evidence suggested that, contrary to the account given by Mr Jones, there was the imposition of a more rigorous standard than the Committee genuinely believed properly reflected student performance, the only purpose of which was to give effect to the preconceived notion of how the candidates were expected to perform.
64. I would reject this submission, and indeed it comes close to questioning the good faith of Mr Jones. It was entirely in line with the Code to use the SRB as the basis for fixing the range in which the grade boundary was likely to fall. The Code does indeed say that that preliminary determination should be informed by any recommendation of the Principal Examiner or moderator. However, where the recommendation is so far removed from the SRB, and given that consistency in standards will not justify any significant departure from the predicted outcome save where a powerful case can be made that standards have improved, it is perfectly understandable why no weight was given to the provisional recommendation of the examiner in this case. Similarly it would have been a fruitless exercise to consider scripts in the range proposed by the Principal Examiner in these circumstances. It is not in fact unusual for the Principal Examiner's indicative range to be so out of kilter with the SRB; we were shown evidence of many other cases where that was so.
65. That fact does not, in my view, begin to cast doubt on the genuineness of the assessment carried out by the Committee. It is also pertinent to note that the deputy principal examiner was a member of that Committee and approved the boundary

adopted by the Chairman. The evidence does suggest that at least at the final stage where reliable statistics are available, there is very little value in the provisional views of the examiner as to the appropriate range of marks to be considered. It is not difficult to understand why those putting material before the Awarding Committee will anchor the range of marks to be considered by reference to the SRB, since this best reflects past performance. A grade boundary which strays far outside that range is unlikely to be consistent with the principle of comparable outcomes and is unlikely to be acceptable to Ofqual.

66. I would not, therefore, be willing to infer that Mr Jones or the Committee in general were under the impression that they were bound by the tolerance limits specified by Ofqual or otherwise gave improper weight to the statistical material. Plainly the predicted outcome was always going to be highly material, and a departure from that pattern would require cogent justification. That, it seems to me, is consistent with Mr Jones' observation that there would be "tight and limited flexibility". It would be astonishing if Mr Jones was unaware of the fact that the reporting tolerance was precisely that; stepping outside the tolerance limits triggered an obligation to report. He may well have justifiably assumed that there was on the face of it no reason to suppose that the standard of this particular cohort of candidates had improved sufficiently to justify awarding marks outside the tolerance limits, in which case there would indeed have been limited flexibility.
67. I have no doubt that AQA's recommendation as to the appropriate boundary was one which, as a matter of academic judgment, it felt fairly reflected the achievements of that cohort of students, and that it was not premised on any belief that the reporting tolerance imposed an absolute barrier to fixing a grade boundary outside the tolerance limit. With hindsight it is difficult to draw any other conclusion than that the assessments made on earlier occasions, when the statistical evidence was much less reliable and in some cases ignored entirely, were too favourable to the students. That raises the question, which I consider below, whether fairness required that that more favourable assessment should be carried through from January to June so that all candidates were assessed in accordance with the same standard.

*The fixing of grade boundaries by Edexcel.*

68. The way in which Edexcel determined the C grade boundaries is addressed in considerable detail by Ms Karen Hughes, the officer responsible for recommending proposed grade boundaries to Ofqual. She too claims that the approach of Edexcel was consistent with the principles expressed in the Code.
69. There were 9,403 students who took the Unit 2 foundation paper in January 2012 compared with 16,539 in June. In January the C boundary was set at 66 marks. This was said to be "the best match of KS2 data and candidate response, archive material and grade descriptions".
70. Only 766 children submitted Unit 3 controlled assessments in January compared with 24,095 students in June. The SRB was referred to but seems largely to have been ignored. The grade boundary adopted was set at 55, which in fact was the same as that which had been adopted in June 2011.



71. For the June assessments, both the Principal Moderator and the Principal Examiner gave indicative grade boundary recommendations. The Principal Moderator was of the view that the C grade boundary for the Unit 3 controlled assessment should be the same as in January. His view was that the schools were marking the speaking and listening components accurately and consistently. The Principal Examiner considered that the Unit 2 foundation tier paper should also have the same mark as in January on the basis that the examination papers were of comparable difficulty. The staff initially planned to adopt these recommendations and use them to determine the range of scripts to be placed before the Awarding Committee. However, before the meeting of the Awarding Committee it became apparent that if the January boundaries were adopted, it would result in a significant departure from the KS2 predictions and so the meeting was postponed whilst various models were tried which might achieve compliance with the 1% tolerance limit.
72. Discussions were held with Ofqual who were told that there was a mismatch between the examiners' judgments and the predicted outcomes. As a result, and after intensive and thorough debate and analysis, both within Edexcel and between Edexcel and Ofqual and the other AOs, there were two modifications of the assessment criteria. First, there was a refinement of the KS2 model, suggested by AQA. Instead of measuring the candidates for GCSE English against the candidates who formerly did English language and literature, it was thought that the appropriate reference group might be those who only formerly did English language (and not English literature) in 2010. This was approved by Ofqual, not without some misgivings. Mr Sheldon was critical of this change. Mr Giffin claimed that its effect was to justify a lower grade C boundary mark than would otherwise have been the case and that the students benefited from this adjustment. It is not in fact clear to me that that is so, but the effect was marginal and the change was considered to be a principled one.
73. The second change was that Ofqual accepted that there could be a 3% tolerance for each of GCSE English and English Language, provided the tolerance did not exceed 1% for the combined pair of subjects. This also allowed greater flexibility to Edexcel.
74. However, even with these modifications it was still necessary significantly to shift the grade boundaries from those given in January. Edexcel was very reluctant to change them more than absolutely necessary, especially with respect to the controlled assessments where the tasks remained the same. It was recognised that changing the boundary would be very difficult to defend; the perception would be that there had been an unjustified hike in standards. Edexcel engaged in a balancing exercise seeking to fix the grade boundaries in a way which fairly reflected academic judgment as to the quality of the scripts and the comparable outcomes year on year, whilst at the same time minimising inconsistency between the January and June cohorts. As Ms Hughes admits, it had by then become apparent that the grade boundaries fixed in January had in fact been unrealistically generous, although at the time they had if anything been considered to be harsh. If those grade boundaries had been carried through, it would have led to a wholly unjustified increase in the number of candidates achieving a grade C.
75. Ms Hughes gives reasons why the January units were more leniently graded: it was a new and in some ways more rigorous examination than the former GCSE, and therefore it was much more difficult to compare standards with the earlier qualification; the cohort was small - indeed very small for Unit 3 - and not necessarily

typical or representative of the qualities of the candidates as a whole; and Key Stage 2 predicts the overall qualification and is a less reliable tool when setting boundaries for individual units.

76. When the Awarding Committee finally met, the views of the Principal Moderator and Principal Examiner no longer set the framework. Instead the indicative grade boundary was the SRB, and the range of marks around that figure was produced for consideration. The Chair had made it clear in advance of the meeting that he would not be willing to change the boundary for Unit 1 which he was satisfied had been fairly assessed, and the Awarding Committee agreed with his analysis. Accordingly, that grade boundary was in fact kept the same as in January and nobody later doubted that this was justified. The Committee, after analysing all the relevant evidence, recommended that the boundaries in the other two units should be as set out above i.e. an 8 mark increase on Unit 2 and a 10 mark increase on Unit 3. The principal rationale for increasing Unit 3 in that way was that it consisted in part of the speaking and listening component, which the experience of Ofqual and AQA suggested, from their wider perspective, had been overmarked (although Edexcel was not initially convinced that this was the case). Furthermore, this component was shared with one of the English language units where a 64 grade had been fixed for grade boundary C. The January mark would have been well out of line.
77. Ms Hughes, as the Responsible Officer (the officer ultimately responsible for recommending grade boundaries), remained concerned that an increase of 10 marks would still create major concerns of consistency as between January and June. It would be perceived as unfair by the schools. Accordingly, after further discussion with colleagues, she proposed reducing the increase in Unit 3 from 10 to 7 thus fixing the grade boundary at 62. She felt that overall this struck a fairer balance between the competing considerations than the marks proposed by the Awarding Committee. Her proposed grade boundary was not, however, acceptable to Ofqual, because it went too far outside the permitted tolerance limits and no justification had been provided for the dilution in standards. Ofqual made it clear that if necessary, in order to ensure fairness as between providers and year on year, it would issue a direction. There was the possibility that this would have increased the boundary even more than the 10 mark differential acceptable to the Awarding Committee since strictly a 12 mark increase was required to bring the boundary within tolerance. However, accommodation was reached and Ofqual agreed to accept a ten mark increase in Unit 3 even though this meant that the numbers qualifying with grade C was still outside the reported tolerance limit.
78. Mr Sheldon submits that this history again demonstrates, as with AQA, that when academic judgment came into conflict with the predicted outcomes, the AO allowed the statistics effectively to drive the outcome. His starker contention that the tolerance limits were slavishly followed is impossible to sustain given that there was in fact a departure from them. He is plainly right, however, to say that the predicted outcome played a very significant role in determining the grade boundaries. But for reasons I have already given when discussing similar arguments with respect to AQA, there is nothing improper in that. If the currency is not to be devalued, the starting point can properly be that predicted outcomes ought, within limits, to define grade boundaries. There may be a departure from that where there is a good explanation why the standard may have improved, but not otherwise. In so far as the academic judgment

based on qualitative assessment alone would lead to a disproportionate number of candidates acquiring a C grade, there is reason to question whether that judgment is correct.

79. Mr Sheldon did identify one possible reason why the performance might have improved as the qualitative material suggested. He referred to some thoughtful observations expressed in a paper produced by Mr Pritchard, Edexcel's head of technical support. Mr Pritchard suggested that because the proportion of controlled assessment work had increased to 60%, and given the modular nature of the examination and the opportunity for resits, it was inherently more likely to lead to an improvement in student performance relative to the old specifications. Students would not have been intrinsically cleverer than hitherto, but they may have found it easier to meet the criteria for grade C specified in the published grade descriptions, which were not changed from the legacy examinations. In order to counter this and to ensure no dilution of comparable standards, the grade descriptions themselves ought to have been altered so that more was expected of a student than had formerly been the case to attain a particular grade.
80. This is certainly a plausible explanation as to why the teachers' qualitative assessment of student performance was out of line with the predicted outcomes. It might have been advanced as an explanation why it was justified to step outside the tolerance limits, but it does not appear that anyone did suggest this to Ofqual. Mr Pritchard would personally have preferred not to depart from the January boundaries - which he described as an "indefensible "fix"" - and instead to correct problems of grade inflation in the longer term. That was not the view of Ms Hughes, nor did it represent a consensus within Edexcel, where it was accepted that it was appropriate to adjust the grade boundary to reflect more closely the statistical evidence. More importantly, Ofqual were entitled to take the view that priority should be given to ensuring that standards were consistent year on year and Edexcel could not ignore that.
81. Ms Hughes has emphasised that although in her judgment the grade boundary could have been lower than that which in fact prevailed, nonetheless the Awarding Committee was satisfied that the increase of 10 marks for unit 3 was academically justified. Fixing the grade boundary is an exercise of judgment and Edexcel considered that having regard to both qualitative and quantitative factors, the grade boundaries ultimately adopted resulted in candidates obtaining their appropriate grades when compared with other years. I see no reason for doubting that analysis. The initial qualitative assessments had to be modified to ensure comparable outcomes, and Edexcel took very great pains to give as much weight to the former as was compatible with the overall objective. Necessarily, however, Edexcel had to give considerable weight to the statistical data.
82. The claimants make a further complaint about the way in which Edexcel in particular chose to resolve the conflict between the original academic judgments and the predicted outcomes. They say that even assuming that Edexcel was entitled to give such weight to predicted outcomes as it did, the way in which Edexcel sought to bring their marks into line with the predicted outcomes was arbitrary and unfair. The reasons for imposing the increases on Unit 3, as opposed to the other units, were not entirely lawful and proper ones. That was not an argument advanced in the original grounds, but I will deal with it briefly later.

*The grounds of challenge.*

83. There are four principal grounds relied upon by the claimants. First, they contend that the conduct of the defendants was conspicuously unfair so as to amount to an abuse of power. The fundamental unfairness alleged is inconsistent treatment, although in support of this contention the claimants pray in aid a host of other factors, some of which are also relied upon as an abuse of power in their own right.
84. The second ground is that the AOs and Ofqual failed to give effect to the legitimate expectation engendered by statements to the effect that grading standards would be the same irrespective of when the assessment was completed. The necessary inference, it is said, was that grade boundaries would not change significantly for written examinations, and not at all for controlled assessments where the task was the same. The claimants also allege that for the most part this is in practice what happened.
85. Third, the claimants say that the defendants acted irrationally in failing to treat all candidates alike and subjecting them to different assessment standards. They deliberately adopted tougher standards in June.
86. Fourth, they say that the defendants acted unlawfully in treating the reported tolerance guidance as though it were a binding principle; alternatively, they gave too much weight to it. Moreover, it was a new factor only made public in May 2012 after many units had already been banked.
87. There is a fifth submission, of more limited significance, that the way in which Edexcel sought to bring their marks into line with the tolerance limits amounted to an arbitrary and unprincipled manipulation of unit grade boundaries.
88. It is instructive to note the remedy the claimants seek. It is to have the June papers assessed in accordance with the January boundaries. This could only be the appropriate remedy for breach of the public law duty if it was the only fair and lawful way in which the defendants could have assessed the units in June. The claimants submit that this is indeed their case but further contend that even if the court is unwilling to go that far, it should still declare the June results invalid and require the AOs to reconsider the assessments in the light of such guidance as is provided by this judgment.
89. Because the first ground, relying upon conspicuous unfairness, also incorporates the other grounds as elements of the unfairness relied upon, I will deal with those overlapping grounds first.

*Ground 4; tolerance and disproportionate weight to statistics.*

90. I can effectively dismiss ground four in the light of the analysis above of the way in which the grade boundaries were fixed. There was no improper fettering of the AOs' discretion by Ofqual, merely a proper concern that the AOs should only depart from the tolerance limits in circumstances where they could provide sufficient justification. This, in my view, is a perfectly legitimate principle to apply in order to ensure that there is broad consistency in standards year on year. Nor did the AOs treat the reported tolerance as an inflexible principle. Indeed, the fact that Edexcel fixed grade

boundaries outside the tolerance limits is evidence that the guidance was not treated as a binding rule, nor was it applied in a slavish fashion. Moreover, the guidance itself was modified as a result of discussions carried out in the course of the assessment process.

91. The statistical evidence identifying predicted outcomes was treated by Ofqual as a factor of considerable importance and there is no doubt that it significantly influenced the assessment exercises. The SRB, at least when it was thought to provide reliable predictive evidence, was in practice likely to dictate the range of marks which the Awarding Committees would consider when determining the appropriate grade. But that was important in order to achieve Ofqual's statutory objective to ensure consistency year on year. There was nothing wrong in giving such statistical data considerable weight; the statistics did not of themselves determine where the boundary was to be struck, as is manifest from the fact that the boundary frequently departed from the SRB, and even from the tolerance limit. Accordingly, I reject this ground of challenge.
92. Furthermore, the fact that the guidance was not made known until May 2012 involves no error of law. The purpose of this guidance was not to change the basic objective, which was always comparable outcomes; its effect was simply to alert Ofqual to the possibility that the objective may be at risk and thereby to facilitate its effective implementation. It was of no relevance to teachers' marking assessments and could not have affected how they approached that task.

*Legitimate expectation.*

93. The claimants contend that the doctrine of substantive legitimate expectation is engaged in this case and that the defendants have acted in breach of it.
94. The essential legal principles are not in doubt. A legitimate expectation may arise either out of an express promise given on behalf of a public body, or from the existence of a regular practice which a claimant can reasonably expect to continue. The public body should always at least have regard to the promise or practice before making a decision which is inconsistent with it, and in certain special situations it must honour the promise, if it would be an unfairness amounting to an abuse of power to do otherwise: see *R v North and East Devon Health Authority ex parte Coughlan* [2001] QB 213. The claimants submit that this is a case where the conduct in departing from January boundaries was so unfair as to constitute an abuse of power.
95. In order for the doctrine to be engaged so as to bind the decision-maker, the assurance must be clear and unequivocal and "pressing and focused", to use the language of Laws LJ in *R (Niazi) v Home Secretary* [2008] EWCA Civ 755, para 41. Similarly any practice must be unambiguous, widespread and well recognised.
96. I confess that even as originally formulated I doubt whether the assurance relied upon would satisfy those criteria. That assurance is said to be that grade boundaries would not change from one assessment date to the next; or at least it would not do so for controlled assessments and would only do so for written examinations where the difficulty of the examination justified it. The basis of this alleged assurance is said to be the following: first, the published documents did not identify the reported tolerance limits nor indicate that the statistical material would play such an important

role in the assessments; second, there had been assurances that standards would not vary depending upon when the unit was assessed, and implicit in that claim was that grade boundaries would stay the same unless there was a proper basis, based on the difficulty of the paper, for changing them; third, the consistent practice for GCSE English in the previous sessions had been consistent with an assurance of the kind now relied upon.

97. In my judgment, each basis is defective and this submission fails at the first fence. The claimants have not, in my view, been able to point to any clear and unequivocal assurance of the kind they seek to rely upon at all. The failure to make public a relevant criterion, such as the reported tolerance limit, cannot create any positive assurance so as to engage the doctrine of legitimate expectation. In any event, as I have said, the tolerance limit did not alter the assessment principles at all.
98. Moreover, as the defendants have to my mind clearly demonstrated, the public documents are in fact inconsistent with the assurance relied upon.
99. A fundamental difficulty is that the approach adumbrated by the claimants would be entirely at odds with the principle of comparable outcomes, which everyone knew was the overriding principle which all AOs had to respect. The application of the same, or broadly the same, grade boundary in both January and June would have significantly inflated the number of candidates obtaining a grade C. Moreover, the existence of detailed procedures laid down in the Code for determining grade boundaries is wholly inconsistent with a simplistic principle that grade boundaries will broadly stay the same. The fact that teachers acted on the assumption that they would is understandable, given that this was a new qualification and they had little else to assist them in determining what a C grade would look like. However, it was unfortunate, and I am satisfied that it was not fostered by any representation from Ofqual or the AOs.
100. AQA also referred the court to various passages in its communications with the schools in which it had been made clear that grade boundaries might change. The basic guide for standard setting stated that as well as changes in the difficulty of the examination paper, further factors may be taken into account when translating marks into grades. And on its website it is expressly stated that “it is sometimes ... necessary to move grade boundaries. This happens when a particular boundary from a previous exam does not represent the same standard as it did in previous years.” It goes on to give the example – which Ofqual and AQA strongly considered was the position here – where some schools are giving higher marks than in previous years for the same quality of work.
101. Mr Lewis QC, counsel for AQA, also pointed out that there had been a number of occasions when boundary changes between sittings in other subjects, such as Maths and French, had been marked just as in the case of English, sometimes even affecting candidates from some of the same schools as are claimants in this case. So practice in relation to GCSEs generally was not consistent with the case now being advanced. The claimants counter that these other examples were not in English. That is true, but in so far as the practice is said to rely on just the English results over the two year period, in my view that is far too short a period to constitute an established practice sufficient to engage the doctrine. In any event, even that practice is not consistent with the case now being advanced. There have been (admittedly small) variations in

the grade boundaries for certain units in English, including controlled assessment units, in that period.

102. Edexcel too pointed to documents which they say are simply incompatible with the argument now being advanced. For example, on its website it emphasised that grades are not confirmed until the assessment is actually awarded. There is also a document which states in terms that Edexcel was “required to review course work and controlled assessment grade boundaries in each series to ensure standards are maintained”.
103. In my judgment, this and other evidence to similar effect demonstrates that the assurance or practice necessary to engage the relatively narrowly-defined doctrine of substantive legitimate expectation does not exist. Of course, the claimants can reasonably expect candidates for the same examination to be treated the same way. That is an argument on consistency, which engages the first ground of complaint. But an expectation of that nature is not the kind of expectation protected by this doctrine, any more than a generalised expectation that the AO will determine grades lawfully.
104. Even if the doctrine were engaged, any assurance or practice creating the necessary expectation can be overridden where the public interest requires it, as *Coughlan* recognised. It will not then be unfair or an abuse of power to frustrate that expectation. I have no doubt that the public interest would justify any failure to give effect to the expectation here, for reasons I develop below when considering the claim based on conspicuous unfairness.

*Conspicuous unfairness.*

105. The claimants’ primary submission is that the defendants acted with conspicuous unfairness in the way they assessed some of the units in June. For reasons I explain below, in my view the irrationality ground adds nothing to this way of presenting the claim.
106. The origin of the concept of conspicuous unfairness is the judgment of Simon Brown LJ in *R v Inland Revenue Commissioners ex parte Unilever* [1996] STR 681. The Inland Revenue had over a period of some twenty five years and on some thirty occasions not sought to enforce time limits when Unilever applied for tax relief. They then chose to do just that in circumstances where the relief would have been granted had the application been made in time. The Court of Appeal held that this change in approach to late applications, without any warning, was so unfair as to amount to an abuse of power, notwithstanding that the court accepted that the practice was not such as to engage the legitimate expectation doctrine. Sir Thomas Bingham MR spelt out in some detail the reasons why, in what he described as the unique circumstances of that case, there was unfairness amounting to abuse of power. These included the fact that the practice was so well established, that the sums were very significant, and that the Revenue was not prejudiced by the late claim.
107. The Master of the Rolls went on to consider what had been advanced as a separate argument, namely that the decision not to exercise a discretion in Unilever’s favour was so unreasonable as to satisfy the public law test of irrationality. He concluded that it was. It is important to note, however, that he did not think that this was

conceptually a new point, but he treated it as such because the judge below had. In that context of irrationality he commented:

“The threshold of public law irrationality is notoriously high. ...And in all save exceptional cases the Revenue are the best judge of what is fair.”

He concluded that this was an exceptional case:

“I cannot conceive that any decision-maker fully and fairly applying his mind to this history .... , could have concluded that the legitimate interests of the public were advanced, or that the Revenue’s acknowledged duty to act fairly and in accordance with the highest public standards was vindicated, by a refusal to exercise discretion in favour of Unilever.”

108. Lord Justice Simon Brown’s analysis adopted the concept of conspicuous unfairness:

“Unfairness amounting to an abuse of power as envisaged in *Preston* and the other Revenue cases is unlawful not because it involves conduct such as would offend some equivalent private law principle, not principally indeed because it breaches a legitimate expectation that some different substantive decision will be taken, but rather because either it is illogical or immoral or both for a public authority to act with conspicuous unfairness and in that sense abuse its power.”

Later in his judgment he observed that there is a distinction between “on the one hand mere unfairness – conduct which may be characterised as “a bit rich” but nevertheless understandable, and on the other hand a decision so outrageously unfair that it should not be allowed to stand.”

109. The concept of conspicuous unfairness has been employed in a number of subsequent cases. The claimants relied in particular on the decision of Richards J in *R v National Lottery Commission ex p Camelot* [2001] E.M.L.R. 3. The Commission considered bids tendered in open competition to run The National Lottery. Neither of the two candidates who entered bids was considered to have satisfied all the criteria necessary to be given the relevant licence. The Commission resolved to abandon the competitive procedure and thereafter it gave one of the bidders an opportunity to allay its concerns about their suitability but not to Camelot. Richards J held that this was conspicuously unfair, relying upon the *Unilever* decision. There was a marked lack of even-handedness which required “the most compelling justification” and on the facts, no such justification was made out. This was a case, however, where the doctrine was used to impose procedural duties on the Commission. In my view, the traditional concept of fairness, without the epithet “conspicuous”, could have been deployed to achieve the same result.



110. The claimants contend that in the light of these authorities, it is for the court to decide whether there is any conspicuous unfairness. As part of that exercise the court would have to be satisfied that there was in fact justification for any prima facie unfair treatment, although the claimants accept that in so doing the court would have to give due weight to such justification as the defendants advanced. The defendants for their part contend that this would be contrary to the fundamental purpose of judicial review which is designed to allow supervisory control over specialist administrative bodies, not to usurp their function.
111. Logically, if there is a doctrine of conspicuous unfairness as a substantive head of judicial review which is to be treated as a distinct form of abuse of power, it must be for the court to decide whether in any particular case the decision-maker has infringed that principle since the court must decide whether power has been abused. It is no different from a court deciding that a decision has been exercised for an improper purpose or that an irrelevant consideration has been taken into account. But I do not believe that *Unilever* has formulated a fresh head of review conferring on the court a wide discretion to substitute its view of the substantive merits for the decision-maker. In order to constitute *conspicuous* unfairness, the decision must be immoral or illogical or attract similar opprobrium, and it necessarily follows that it will be irrational. I would treat this concept of conspicuous unfairness as a particular and distinct form of irrationality, which in essence is how it was viewed by Sir Thomas Bingham in *Unilever*. There are no doubt cases, of which *Unilever* is one, where the concept of fairness, and an allegation of conspicuous unfairness, better captures the particular nuance of the complaint being advanced than the concept of irrationality. Indeed, I think that is typically so in any case where the alleged unreasonable behaviour involves a sudden change of policy or inconsistent treatment. It is more natural and appropriate to describe such conduct as unfair rather than unreasonable. But in my view it is only if a reasonable body could not fairly have acted as the defendants have that their conduct trespasses into the area of conspicuous unfairness amounting to abuse of power. The court's role remains supervisory.

*Was there conspicuous unfairness?*

112. Mr Sheldon relied upon a whole range of factors which, whether taken cumulatively or individually, amounted to conspicuous unfairness within the *Unilever* principle. I have touched on most of the matters already but will briefly recapitulate what seem to me to be the essential points he encompassed under this heading.
113. First, he relied upon the observations of Mr Pritchard, referred to above, who had suggested that adherence to comparable outcomes with the imposition of the reporting tolerance principle failed to reflect the fact that standards of performance could be expected to rise as a result of the new examination structure. Second, he relied on the disconnect between compliance with the grade performance descriptors and the need to pay considerable respect to predicted outcomes. (I think that is in fact merely another way of putting the first point, namely that there may be improvements in performance even though the quality of the students has not improved.) Third, there was never any hint that the statistical material might require such a dramatic change in grade boundaries as were witnessed in this case. The natural assumption would be that grade boundaries would not change, particularly for controlled assessments where the tasks remained the same, and teachers reasonably acted on that assumption, in some cases to the detriment of their children. Fourth, the June assessments placed

disproportionate emphasis on the statistical information, and the reported tolerance limit effectively dictated the outcome. This was not a legitimate and genuine assessment of the academic quality of the students. Fifth, the effect of applying the tolerance limit was that there was in effect “hyper-correction”, with the June cohort being marked so as to correct what was perceived to be the earlier over-generous marking of candidates submitting units earlier in the series. Finally, there was in fact an inconsistency in the way in which the June candidates were assessed when compared with candidates submitting earlier papers, and a powerful justification was required for this difference in treatment. There was none in this case.

114. I have already explained why I am not persuaded that the first four grounds taken individually demonstrate any error of law. The first two fail to recognise that even if performance standards have improved, Ofqual was entitled to give priority to the comparable outcome approach. The third is really the legitimate expectation argument, and the fourth the criticism that the statistics improperly dominated the outcome, both of which I have rejected. Mr Sheldon submits that even if these do not establish legal errors in their own right, they are factors which can still carry some weight in an overall assessment of fairness. In principle, I would accept that they could in so far as they show unfairness, since fairness must be assessed in the round, taking into account all potentially relevant factors. However, they do not, in my judgment, demonstrate any unfairness here.
115. The underlying premise lying at the heart of the submission on conspicuous unfairness is that there was inconsistency in the approach in June and January. It is well established, at least since the judgment of Lord Justice Scarman in *HTV Ltd v Price Commission* [1976] ICR 170,192 that in certain contexts inconsistency is itself a head of unfairness which can be remedied in judicial review proceedings. The January cohort was treated more favourably than the June. This, if established, infringes the most elementary principle of fairness which requires that like cases are treated alike. This might be thought to be particularly important when the focus is on candidates who are all doing the same examination. To use a well-worn metaphor, it cannot be legitimate to change the goalposts once the game is in play to the detriment of some of the players.
116. The submission that the defendants acted irrationally in so doing, although advanced as a separate argument, is in reality just another way of putting this argument; no reasonable examiners would have treated the June candidates in the way in which Ofqual and these AOs did.
117. The defendants submit that the premise is unfounded. Essentially the same process was employed in both January and June to fix the grade boundaries, although the more sophisticated information available in June allowed for a more thorough and reliable and, as it happened to turn out, stricter assessment. In each case the comparable outcomes approach was the guiding principle but the ability to secure that objective was more difficult when the evidence required to achieve it was less reliable.
118. In my view, this fairly describes what they did; and in many circumstances I would accept that adopting a consistent approach would be a complete and sufficient answer to any inconsistency claim. But in my judgment the fact that we are concerned with candidates being examined for the same qualification demands equality of treatment,

and anything falling short of that requires justification. Candidates can reasonably anticipate that they will be examined and assessed in the same way and according to the same standards, not merely in accordance with common criteria whose content may change depending upon when the assessment is made. It was appreciated before the June grades were finalised that there was at least a risk that a tighter standard was being adopted. In those circumstances there remains the question whether the defendants were justified in applying established procedures notwithstanding that they led, even if only in a relatively minor way, to the application of different standards, or whether they ought, in fairness, to have adopted the January grades. In my judgment it is not a sufficient answer to say, as the AOs do, that on each occasion in January and June they exercised a judgment as to the proper grade boundary, and that was an academic judgment which, in accordance with cases such as *Clark v University of Lincoln and Humberside* [2000] 1 WLR 1988 is not suitable for adjudication by the courts.

119. However, whilst I accept that justification is required, I have no doubt that it is clearly established here. This case comes nowhere near establishing conspicuous unfairness against any of these defendants. If the AOs could have delayed grade assessments so that they were all carried out at the same time, irrespective of when the assessment in fact occurred, this would have enabled them to apply the same criteria to each unit. But under the rules as they stood, they could not do that. Were they obliged to apply the January boundaries even though this would have involved, in their judgment, giving C grades which they did not consider to be justified? In my view they were not. Even had they been minded to do so, Ofqual would if necessary have made a relevant direction, which is precisely what they indicated they would have done with respect to Edexcel Unit 3 had there been no accommodation of views. So in fact the AOs had no option other than to apply the standards they did in June, or something close to them, notwithstanding that they departed from the January standard. The substance of this particular complaint must therefore be directed against Ofqual.
120. Ofqual's position is different because they could have permitted a loosening of the tolerance limits. They were not obliged to issue a direction even if the AOs had chosen to depart from the tolerance limits to an extent undermining the comparable outcomes objective. The claimants submit that the only fair outcome was to follow the January grade boundaries, at least for the controlled assessments. Even if that had led to some grade inflation for those qualifying in 2012, that was an acceptable price to pay in order to secure greater even handedness between the January and June cohorts, and it could readily have been justified on the grounds that it was achieving greater public confidence in the process. In effect, this is giving priority to fairness within the year albeit at the expense of fairness as between years.
121. The question is not whether Ofqual could have adopted this approach; it is whether it was unfair for Ofqual not to have done so. In my judgment, Ofqual's decision to give priority to comparable outcomes cannot possibly be so characterised. One obvious reason why it would in fact have been wrong to apply precisely the same boundary mark is that Ofqual found strong evidence that some teachers had been marking more leniently in June, pulling students up to the mark which they anticipated, in the light of the January grade boundaries, would secure them a C grade. Plainly that does not account for the whole of the disparity between the January and June grade boundaries,

however, and it may be said that fairness required that the same boundary should be applied once some allowance had been made for that distortion.

122. Were this a single isolated examination taken in a time bubble, this would be a very powerful argument indeed. But the problem here is more complex because of the need to ensure comparability of outcomes year on year. Even accepting that there was more favourable treatment in January, Ofqual had to bear in mind the consequences of applying the more lenient standard in June. There were markedly more students being assessed in the critical papers at that stage than in January, and there would have been a significant dilution of standards, contrary to the vitally important objective of maintaining the currency of the qualification, if the AOs had simply applied the same standard.
123. The problem was that Ofqual could not remedy any unfairness between the January and June cohorts without creating further unfairness elsewhere. The 2012 students would have been assessed more leniently than students in earlier years. In addition, there would have been students being assessed for units in June 2012 who would have been qualifying in June 2013. They could not in fairness be assessed more strictly than others assessed in June 2012 but qualifying in that year. But if they were assessed in this more favourable manner it would mean that the unfairness now felt by the current June 2012 students would be similarly experienced by the cohort taking these units in June 2013, comparing themselves with those qualifying on the same date who had completed the relevant units in June 2012.
124. In my judgment, maintaining the currency of the qualification was a powerful and legitimate justification, in the public interest, even though that involved accepting as a necessary, albeit highly undesirable, consequence that the June students were to some extent subject to tougher assessments than their January (and indeed earlier) counterparts. This is particularly so given the fact that those benefiting from the over generous grading in January were a relatively small contingent when compared with the June cohort. That was especially true in relation to the boundary change most strongly challenged in this case, namely Unit 3 in the Edexcel qualification. If Ofqual had not brought a halt to the inconsistent standards at this stage, they would have had to have done so at some later stage unless they were prepared to forego their principal objective and debase the value of the currency of English GCSE. In all likelihood that would have compounded the unfairness because a greater number of students would have been more leniently treated. It was a cogent and rational decision for Ofqual to have grasped the nettle when they did.
125. In my view two authorities in particular lend some support to this conclusion. In *R (O'Brien) v The Independent Assessor* [2007] UKHL 10; [2007] 2 AC 312 three men were wrongly convicted of murder and were subsequently awarded compensation for miscarriage of justice. In each case the independent assessor deducted from the sums payable personal living expenses such as food, clothing and accommodation which the applicants had not had to incur because they were in prison. The House of Lords considered that this was in principle a justifiable deduction. However, one independent assessor made deductions of 25% and 20% in respect of two of the defendants, whereas another independent assessor had only reduced the figure by 10%. It was contended that this was unfair, not least because the defendant who had been subject to the lowest deduction had a more serious criminal record than the other two, and that was a relevant factor when assessing compensation. The House of Lords

rejected this ground of appeal. Lord Bingham of Cornhill, with whose judgment on this point Lord Rodger of Earlsferry, Lord Carswell, and Lord Brown of Eaton-Under-Heywood agreed, said this (para 30):

“It is generally desirable that decision-makers, whether administrative or judicial, should act in a broadly consistent manner. If they do, reasonable hopes will not be disappointed. But the assessor’s task in this case was to assess fair compensation for each of the appellants. He was not entitled to award more or less than, in his considered judgment, they deserved. He was not bound, and in my opinion was not entitled, to follow a previous decision which he considered erroneous and which would yield what he judged to be an excessive award.”

126. I recognise that this is not on all fours with the current application because these were different assessors and also because the imperative to treat like cases alike is arguably stronger where those affected are all candidates sitting the same examination. But the case does emphasise that there is nothing inherently unfair in putting right earlier errors rather than compounding them, even if this involves creating a disparity between similarly placed individuals.
127. Perhaps a closer analogy to this case is *R (Tate & Lyle Sugars Ltd) v Secretary of State for Energy and Climate Change* [2011] EWCA Civ 664. The facts were that the Secretary of State created a subsidy scheme to promote renewable energy generation. He set bands providing different levels of subsidy for different types of renewal generator. These bands were subject to review on a four-yearly basis. There was also, however, a power to carry out a specific review of specific bands within the four year period. The appellant complained that the type of generator which it operated had been placed in the wrong band due to a calculation error relating to costs. The consequence was that it received a smaller subsidy than it should have done, relative to other similarly placed operators. This complaint caused the Secretary of State to carry out a specific review into that band; he did not simply put in the correct cost figures but included other relevant and up to date information. The result was that the Secretary of State set an even lower band for the appellant’s generator than that which had been the subject of complaint. Tate & Lyle contended that this was unfair because of inconsistency. If the proper assessment had been made when the application for the subsidy had first been lodged they would have benefited from the higher subsidy. In the circumstances it was unfair for the Secretary of State to carry out a full review; she should simply have re-assessed the subsidy in the light of the proper costs figure.
128. The court rejected this submission. It recognised that other producers had received a windfall as a result of increase in electricity prices, but there was no obligation to extend what with hindsight might have been a generous subsidy to the appellant. In giving the judgment of the court, I said this (para 34):

“I recognise that fairness is an important principle of public law but in determining what is fair in any particular context it is necessary to have regard to the wider public interest. I am not persuaded that as a consequence of this review the Appellant is being unfairly treated. They are in fact receiving the

appropriate subsidy for someone incurring the costs involved in developing their particular technology. It is true that they were not obtaining the windfall resulting from the increase in electricity prices which they would have received had no error been made. Furthermore, it may be the case that other producers are receiving a windfall as a result of that price increase and will continue to do so until their technologies are reviewed (although as I have said there will be no windfall if costs have outstripped the electricity price). That is not, in my judgment, a sufficient reason to confer this benefit on the Appellant. It may be bad luck that but for the error the Appellant would have been treated more favourably than was necessary properly to subsidise their technology, particularly since some others will have received the more favourable treatment. It does not follow that it was unfair and an abuse of power to carry out a full review.”

129. In this case I am satisfied that the examiners in June made assessments which they thought fairly reflected the standard of the scripts. In the light of the fuller information then available to them, their judgments were more accurate and more reliable than the January assessments. Wider concerns about creating unfairness as between those qualifying in different years, and the need to retain the value of the qualification, strongly militated against applying the January grades to the June assessments (even with such modification as may have been necessary to account for more lenient marking) to the June assessments. There was no obligation to extend the generosity of January to June; on the contrary, there was every reason to correct the earlier erroneous standard. There was no unfairness, conspicuous or otherwise, in what they did.

*Was Edexcel arbitrary in its boundary fixing in June?*

130. Mr Sheldon advanced an argument during the course of the hearing which was not foreshadowed in his grounds. It really emerged as a result of disclosure in this case. It is directed at the way in which Edexcel sought to give effect to the comparable outcomes policy when making the assessments in June. The contention is that even if Edexcel could legitimately tailor their assessments to give effect to that policy, nevertheless they did it in an unfair way. They chose for improper purposes to adjust only one of the controlled assessments, namely Unit 3, and not the other. This does not assist the claimants in their broad challenge; it is a much more limited complaint and, if correct, its effect may have been that within Edexcel’s June cohort sitting these units, some candidates will have been treated more strictly than they ought to have been, and others less so.
131. As I have pointed out, in its attempt to stay within tolerance, Edexcel did not alter one of its controlled assessments, Unit 1, from the January grade boundary, but by contrast it increased Unit 3 by 10 marks. The contention is that the evidence demonstrates that a significant reason why this was done was because Unit 3 contains a speaking and listening component worth half the marks, that no record of performance is retained for that element of the qualification, and that accordingly it would be more difficult to show that Edexcel was acting inconsistently or unjustly in altering that grade boundary than if it altered Unit 1. It would minimise criticism

because there was nothing against which the assessment could be checked; the change could more readily be defended and this would save Edexcel considerable embarrassment. The effect of imposing all the change on Unit 3, speaking and listening, is that candidates who were particularly good at that element of the examination would be prejudiced as against those whose ability shone in Unit 1, not subject to any change at all.

132. I have already summarised in broad terms the way in which Edexcel went about fixing its grade boundaries. The chair of the examiners was Mr Farrell. He was unwilling to move the Unit 1 controlled assessment by even one mark. It was precisely the same as in January and the Awarding Committee, on considering the written papers, were satisfied that there was no justification in shifting it. The Awarding Committee agreed to increasing the Unit 2 written examination by 8 marks. There has been no real complaint about that grade boundary, save for the observation that even for a written examination, where it is recognised that grade boundaries may change, this was an unusually high increase.
133. The justification for requiring Unit 3 to bear the brunt of the increase, given in the post-awarding report produced by Edexcel, was that the speaking and listening element had been very lightly moderated, and had led to mark inflation at least across all the AOs as a whole. It was therefore likely that there had been over-marking. Moreover, 26% of the candidates had already banked Unit 1, whereas only 3% had banked Unit 3. Accordingly, varying the former would have a far less significant statistical impact.
134. It appears that after the Awarding Meeting, there was no reconsideration of the possibility of moving the grade boundary in Unit 1, although there was some discussion about Unit 2. Ms Hughes says that there were already significant changes there and that to move the boundary higher was unacceptable because it would no longer reflect the Awarding Committee's view as to the appropriate standard for the quality of the scripts. Thereafter, the discussion revolved around the extent to which the Unit 3 C grade boundary could properly be moved.
135. There is no doubt that Edexcel was greatly exercised by the difficulty of finding a fair solution to the problem of reconciling consistency of assessment with tolerance limits needed to achieve comparable outcomes. We were taken to voluminous correspondence which shows, not surprisingly, that they were acutely concerned about how they could convincingly explain a 10 point increase in the grade C boundary. Mrs Hughes herself thought that it would be "indefensible" in terms of the perception of schools. She did emphasise, however, that the problem was that the January marking was too generous; it was not that the June cohort were being awarded a lower grade than their performance merited.
136. There is no doubt that the claimants can point to certain observations, particularly from Mr Farrell, which lend support to their submission that he at least was not merely conscious of the embarrassment of having to explain significant grade increases, but adopted a route designed to minimise that embarrassment. For example, in his report he said that if the Unit 1 controlled assessment was moved at all, even by 1 mark, it would "create a negative PR effect out of proportion to any statistical advantage". In the Minutes of the Awarding Committee meetings he is reported to have said that "Unit 3 is going to be our means of manipulating the stats every time";

and he later commented that nobody could gainsay the change because in the speaking and learning unit there was no record against which the mark could be checked (which is in contrast with Unit 1 where a written record was produced.)

137. The claimants submit that these observations suggest that improper considerations have weighed with the Committee when making this recommendation. I recognise the force of this submission but ultimately, on a fair analysis of all the evidence, I am satisfied that whilst concerns about Edexcel's reputation were always present in the discussions, as they were bound to be, there was a genuine belief that it was legitimate for Unit 3 to bear the weight of the adjustment essentially for the reasons given. In reaching that conclusion I bear in mind the following considerations. First, the ultimate decision here was taken by Mrs Hughes, with the support of the senior officers in Edexcel, and not by Mr Farrell. I do not accept that they were simply cynically manipulating statistics in a wholly unprincipled way to protect their own position. That does not do justice to the detailed and professional way in which the discussions were held. Second, although Mr Farrell thought that it would be hard to defend increasing the Unit 1 boundary, it is not obvious why a small increase in that boundary, matched by a corresponding reduction in the Unit 3 boundary, would have been any more difficult to defend. Third, Ms Hughes obviously expected – rightly as it turns out – that the option Edexcel were adopting would create considerable difficulty for them. Indeed, she was particularly concerned that there would be a lot of criticism once the increase was in double figures, hence the efforts she made to try to avoid that result.
138. I would add that even if I am wrong about this aspect to the case, in my judgment, it would be wholly unrealistic to grant any relief in relation to this particular error. The court could not say what the proper approach should have been and so the matter would have to be remitted for further consideration by Edexcel. That would be wholly undesirable for a number of reasons. First, as I have said, there can be little doubt that speaking and listening would still bear the brunt of the grade boundary change even if Edexcel considered that Unit 1 should share some of the pain. Second, given that any change would be minimal, it would be very unlikely to have any significant impact on the GCSE grade of any specific candidate. Unit 3 is only part of the qualification as a whole and although a marginal regrading of that and Unit 1 could theoretically at least affect a small number of students, some for the better and some for the worse, with respect to the units actually changed, the impact on the qualification as a whole would be even smaller. Third, it would be invidious with respect to those (admittedly very few) candidates who might be worse off in any re-grading exercise. Fourth, this is peripheral to the principal challenge, which is concerned with what is alleged to be a very significant and unjust adjustment to boundaries which have caused a considerable number of candidates to fall short of the appropriate grade C qualification. Finally, it would be quite unrealistic to require the whole awarding process to be carried out again at this stage when some of the disappointed candidates will no doubt already have re-sat the examinations.

*Are the AOs judicially reviewable?*

139. Mr Giffin advanced an argument to the effect that the AOs are private law bodies exercising contractual powers and ought not to be subject to judicial review. Put in that bald way, I find the argument unconvincing and in fairness to Mr Giffin, he realistically did not pursue the point with any enthusiasm in his oral submissions. In



my view the decisions under challenge plainly have a “public element, flavour or character” to them, to use the language of Dyson LJ in *R (Beer) v Hampshire Farmers Market Ltd.* [2004] 1 WLR 233, para.16. The determination of GCSE grades, taken by students across the country, is a matter of very significant public importance potentially affecting the life chances of those who are candidates for the examination. This is a classic case of contracting out a public function. Mr Sheldon cited a variety of reasons, backed by a host of authorities, in support of the proposition that the private nature of these bodies was not decisive in this case, but it is not necessary to overburden an already lengthy judgment by referring to them. Moreover, it is plain that Ofqual is central to these applications, and it had to be challenged by way of judicial review. Since ultimately it is the decisions of the AOs which the claimants wish to change, they had to be parties to the proceedings.

140. Mr Giffin’s second and in my judgment much more powerful submission was that given the regulatory function of Ofqual, any complaint about the way in which the AOs have exercised their functions ought to be determined, in the first instance at least, by Ofqual. The claimants would have a remedy against Ofqual if Ofqual permitted an AO to act unlawfully because Ofqual has power to issue directions to prevent this. The scheme of the legislation should preclude any challenge to the AOs themselves.
141. It seems to me that in substance this is an argument about alternative remedies rather than the amenability of AOs to judicial review in principle. A court may refuse judicial review where the claimant has an alternative remedy which can be pursued, and will usually do so. But whether it is appropriate to take that step depends upon a number of factors, including the effectiveness of the alternative remedy and questions of delay. We did not hear argument about that, and were not taken to the investigatory powers of Ofqual. But in any event I do not think this was a realistic route in this case, given that by the time the claimants brought their case Ofqual had already produced its interim report in August 2012 in which it had concluded that there was nothing wrong with the way in which the assessments had been made. There was no realistic possibility that Ofqual would provide the remedy which the claimants seek, and moreover it was important to have a speedy resolution of the matter. To the extent that the AOs are saying that it is pointless to proceed against them because they were under the control of Ofqual and were effectively obliged to act as they did, that may be a defence to the claim but is not a justification for denying the court jurisdiction to consider it.

*The public sector equality duty.*

142. Section 149 of the Equality Act 2010 obliges public bodies or private bodies when exercising public functions, to have due regard when exercising their functions to the need to achieve certain equality objectives. These include the aim of advancing equality of opportunity between persons who share a relevant characteristic and those who do not. The protected characteristics include race, disability, age and sex.
143. This is a continuing and non-delegable duty. The function does not have to be exercised in a way which achieves the desired objectives, but they should be properly taken into account when the function is exercised.

144. The claimants submit that Ofqual introduced the 1% reporting tolerance limit, thereby radically increasing grade boundaries, without giving any consideration to this duty at all. Had they done so, it is very likely that they would have found that the implementation of stricter boundaries in June disproportionately adversely affected those for whom English is a second language. Mr Sheldon suggested that it is at least possible that statistics would demonstrate that such candidates would choose to submit papers for assessment at the latest possible time, taking all their units in June 2012. If this were the position, they would be disproportionately prejudiced by the application of stricter grade boundaries. Moreover, they would be more likely to be found around the C grade boundary than those for whom English was a first language. The defendants ought at least to have explored whether there was any adverse effect on racial groups before implementing the tolerance limit.
145. Ofqual and the AOs do not suggest that they did take account of their PSEDs in relation to the setting or implementation of the tolerance limits. Each emphasises that they take their PSEDs very seriously and have taken steps to ensure that the interests of these disadvantaged groups are fully taken into account when fixing curricula and drafting examination papers. However they contend that the duty has no relevance when it comes to assessing performance; all examinees must be assessed to the same standard so far as that can be achieved, irrespective of their personal characteristics, protected or otherwise. They relied upon certain observations of mine in *R(Greenwich Community Law Centre) v Greenwich London Borough Council* [2012] EWCA Civ 496 at [30]:

“Furthermore, as Pill LJ observed in *R (Bailey) v Brent London Borough Council* [2011] EWCA Civ 1586 para 83, it is only if a characteristic or combination of characteristics is likely to arise in the exercise of the public function that they need be taken into consideration. .... (Perhaps more accurately it may be said that whilst the Council has to have due regard to all aspects of the duty, some of them may immediately be rejected as plainly irrelevant to the exercise of the function under consideration — no doubt often subliminally and without being consciously addressed. As Davis LJ observed in *Bailey*, para 91, it is then a matter of semantics whether one says that the duty is not engaged or that it is engaged but the matter is ruled out as irrelevant or insignificant).”

The defendants submit that the duty was plainly wholly irrelevant in this context.

146. I agree that the argument must fail, for a number of reasons. First, it was not necessary to carry out an equality assessment in relation to the imposition of the tolerance limits themselves. As I have said, these did not alter the basic objective of ensuring comparable outcomes; they merely enabled Ofqual to be alerted to any risk that this might not be achieved and to be told why. The complaint has to be that the decision to apply the comparable standards principle itself should have engaged the duty.
147. As to that, Ofqual could properly take the view that it considered it critical that there should be no dilution of the C grade and that it would adopt necessary and objective procedures to ensure that outcome, irrespective of equality considerations. No-one

suggests that grade boundaries should vary depending on whether the candidate comes from a protected group or not. Grading involves making a judgment about the level of knowledge, skill and understanding of candidates against the background of the available qualitative and quantitative evidence. Ofqual and the AOs justifiably considered that it would be wrong to manipulate boundaries to favour groups with particular protected characteristics.

148. Accordingly, even assuming that there were equality consequences of the kind suggested by Mr Sheldon, it was not a breach of the PSED for the defendants to pay no heed to them. It would therefore have been a futile exercise to carry out the statistical exercise necessary to determine whether there were equality implications or not. They do not have to be explored where they can have no bearing on the decision under consideration.

*Conclusion.*

149. The claimants brought this case because they considered that students had been treated unfairly. There are two principal grievances: first, the actual performance of these students had not been fairly reflected in their grade because the results had been unjustly moulded to reflect predicted performance. The statistics had dominated the assessment process in a wholly unacceptable way. I have rejected that submission, essentially on the ground that it was legitimate for Ofqual to pursue a policy of comparable outcomes, ensuring a consistent standard year on year, and assessing marks against predicted outcomes was a rational way of achieving that objective. Moreover, the Awarding Committee in each of these AOs believed that the June grades fairly reflected the quality of the candidates.
150. The second grievance is a wholly understandable one, and relates to the inconsistent treatment meted out to the students taking assessments in January and June respectively. There is no doubt with hindsight that the former were treated more generously than the latter. Some teachers, again understandably, took the January grade boundaries as a strong guide to future assessment. They did not anticipate the boundary shifting as much as it did in certain units. The reason for the change was in part that some teachers had marked papers more leniently in June specifically in order to bring them just above the C grade; but that was far from the whole story. More significantly, there was fuller information available in June than in January and it became clear with hindsight that the January cohort had been treated too leniently.
151. Ofqual was in a difficult position. It considered and rejected the possibility of re-assessing the January grade assessments. Nobody seriously suggests that it should have retrospectively reduced a candidate's grade in that way when the result had been made public. Yet if it were to have applied the grade boundaries in June, it would have led to a significant dilution of standards, with an unrealistically high proportion of students obtaining a C grade. That would have created an injustice as between those qualifying in June 2012 when compared with students in earlier and subsequent years. Indeed, the problem is compounded when it is appreciated that some candidates for particular units in June 2012 were qualifying in June 2013. If they were to be assessed according to the January 2012 boundary marks, that would be unfair to candidates taking the same unit in January and June 2013. It would manifest precisely the same unfairness that the claimants now allege, but shifted to different victims.

152. The problem lies in the modular nature of the examination, coupled with the fact that grade boundaries were assessed and made public at each stage of the process. Mr Sheldon was highly critical of this structure. He rightly points out that a number of experts had predicted precisely the kind of difficulties which have, in fact, arisen. He says that the problem is of Ofqual's own making (or at least, Ofqual's predecessor). That may be so, but the judicial review challenge is not to the modular nature of the assessment process, or to the practice of assessments being made at different points in the two year qualification period. It is a challenge to the way in which Ofqual and the AOs sought to deal with the problems once they had materialised.
153. Initially it was assumed that since the same procedures were being adopted in January as in June, there should be no change in standards. In fact, this was not so and the January cohort were assessed more leniently. Once that became clear, Ofqual was engaged in an exercise of damage limitation. Whichever way it chose to resolve the problem, there was going to be an element of unfairness. If it imposed the same standard in June as it had in January, this would be unjust to subsequent cohorts of students taking the units in subsequent years. If it did not, that would favour the January cohort over the June cohort in 2012. Unless standards were to be lowered into the future and the currency of GCSE English debased, at some stage a decision would have had to be taken to depart from the less rigorous January grade boundaries and at that point, whenever it was, there would be winners and losers.
154. The claimants submit that even if the January cohort was treated unduly favourably, it was wrong to draw a distinction between groups of candidates qualifying in the same year. This was more important than equality as between years.
155. However, there is no obvious or right answer to the question where the balance of unfairness should lie. Ofqual's solution was in my judgment plainly open to them. Their priority was to protect the comparable outcomes objective, although it meant that January candidates were treated more generously. However, the adverse consequences were relatively contained by acting at that point since far fewer students took the relevant units in January than in June.
156. For these reasons, which briefly recapitulate those spelt out in some detail in this judgment, I do not think it can be said that Ofqual or the AOs erred in law.
157. I therefore dismiss these applications. As I have said, however, this is a rolled up hearing, and although nothing turns on the point, I would grant permission for the applicants to bring these proceedings. This was a matter of widespread and genuine concern; there was on the face of it an unfairness which needed to be explained. There is no question, in my view, that the matter was properly brought to court. Indeed, following the outcry when the results were published in August, Ofqual itself carried out an investigation into the concerns which were being expressed and produced two reports, an interim report and a final one produced after consulting widely with interested parties. Ofqual was not persuaded that it should require the grade boundaries to be changed, but it appreciated that there were features of the process which had operated unfairly and it proposed numerous changes for the future which are designed to ensure that the problems which arose in this case will not be repeated. It also took the unusual step of allowing students to take resits in November instead of having to wait until the following January. We are not directly concerned with those reports which simply reflect Ofqual's own views. However, having now reviewed the

evidence in detail, I am satisfied that it was indeed the structure of the qualification itself which is the source of such unfairness as has been demonstrated in this case, and not any unlawful action by either Ofqual or the AOs.

158. For these reasons, I therefore grant permission to bring judicial review proceedings but dismiss the applications.

**Mrs. Justice Sharp:**

159. I agree.